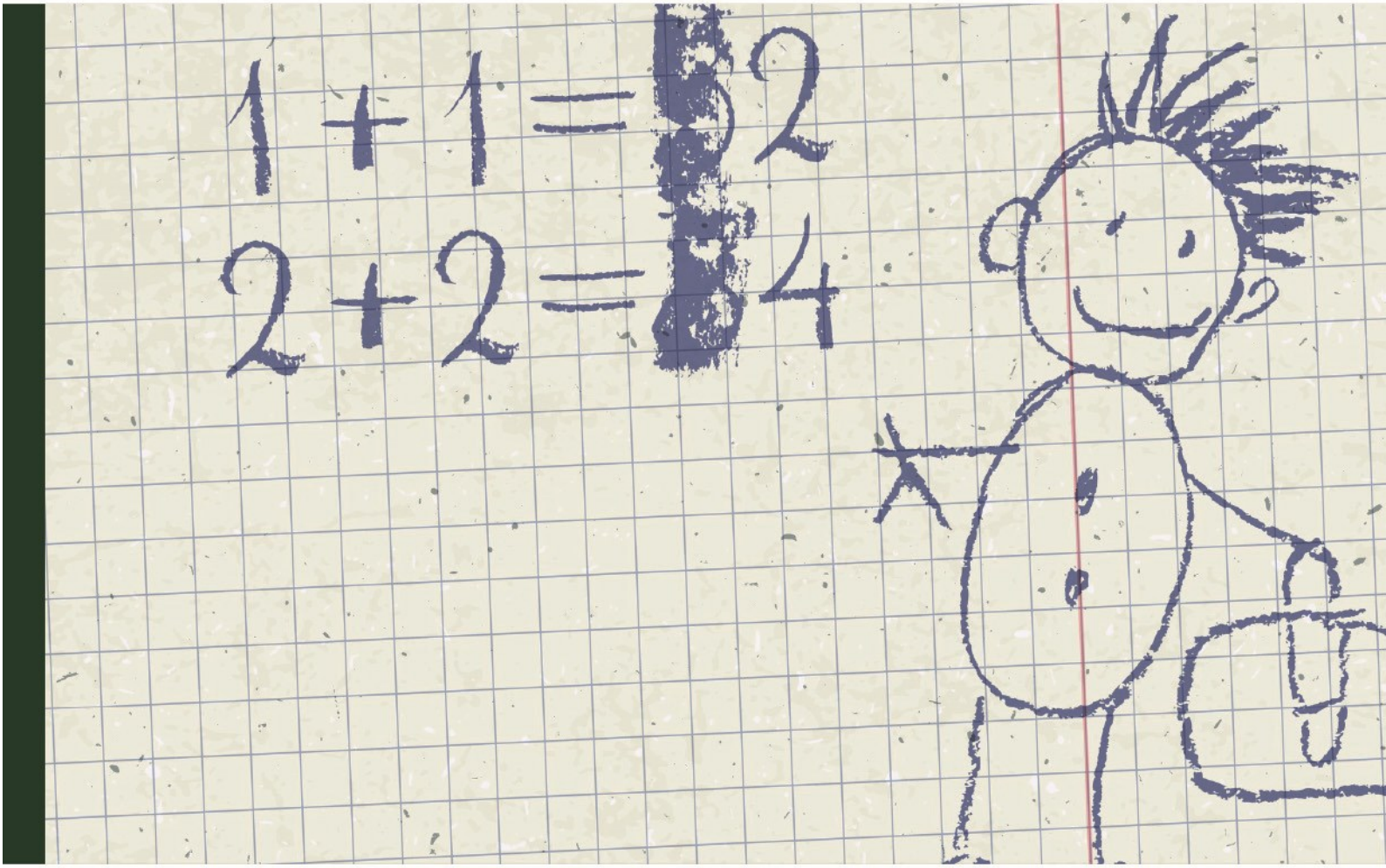


Stefan Voß

Curriculumbasierte Messverfahren im mathematischen Erstunterricht

Zur Güte und Anwendbarkeit einer Adaption US-amerikanischer Verfahren im deutschen Schulsystem

SVH Südwestdeutscher Verlag für Hochschulschriften

Stefan Voß

Curriculumbasierte Messverfahren im mathematischen Erstunterricht

Dissertationsschrift
zur Erlangung des akademischen Grades
Doctor philosophiae (Dr. phil.)
am Institut für Sonderpädagogische Entwicklungsförderung und Rehabilitation
der Philosophischen Fakultät
der Universität Rostock

Gutachter:
1. Gutachter:
Prof. Dr. Bodo Hartke,
Institut für Sonderpädagogische Entwicklungsförderung und Rehabilitation, Universität Rostock

2. Gutachter:
Prof. Dr. Katja Koch
Institut für Sonderpädagogische Entwicklungsförderung und Rehabilitation, Universität Rostock

3. Gutachter:
Prof. Dr. Alfons Strathmann
Department Heilpädagogik, Universität zu Köln

Datum der Einreichung: 02. August 2013
Datum der Verteidigung: 17. Dezember 2013

urn:nbn:de:gbv:28-diss2014-0087-1

Stefan Voß

Curriculumbasierte Messverfahren im mathematischen Erstunterricht

Zur Güte und Anwendbarkeit einer Adaption US-amerikanischer Verfahren im deutschen Schulsystem

Südwestdeutscher Verlag für Hochschulschriften

Impressum / Imprint
Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.
Alle in diesem Buch genannten Marken und Produktnamen unterliegen warenzeichen-, marken- oder patentrechtlichem Schutz bzw. sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Inhaber. Die Wiedergabe von Marken, Produktnamen, Gebrauchsnamen, Handelsnamen, Warenbezeichnungen u.s.w. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutzgesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.
Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this works is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Coverbild / Cover image: www.ingimage.com

Verlag / Publisher:
Südwestdeutscher Verlag für Hochschulschriften
ist ein Imprint der / is a trademark of
OmniScriptum GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Deutschland / Germany
Email: info@svh-verlag.de

Herstellung: siehe letzte Seite /
Printed at: see last page
ISBN: 978-3-8381-3810-7

Zugl. / Approved by: Rostock, Universität, Dissertation, 2013

Copyright © 2014 OmniScriptum GmbH & Co. KG
Alle Rechte vorbehalten. / All rights reserved. Saarbrücken 2014

Danksagung

„Geschafft!“ Welch ein schöner Moment. Ich befinde mich am Ende eines langwierigen Prozesses, das durch das Zutun vieler Menschen nun erreicht ist. Diesen Menschen möchte ich an dieser Stelle danken.

Mein ganz besonderer Dank gebührt Prof. Dr. Bodo Hartke und Prof. Dr. Katja Koch. Sie haben durch die Anstellung im Forschungsprojekt um das Rügener Inklusionsmodell dazu beigetragen, dass meine Promotionszeit finanziell tragbar wurde. Prof. Dr. Bodo Hartke hat zudem meine Aufmerksamkeit auf das Thema der Lernverlaufsdokumentation anhand von curriculumbasierten Messverfahren gelenkt. Sehr unterstützend waren dabei unsere häufigen Gespräche und seine Hinweise, sowohl fachlicher als auch persönlicher Natur. Nicht zuletzt hat er durch Verständnis und Rücksichtnahme das „Promovieren mit Kind“ ermöglicht. Lieber Bodo, danke für deine Unterstützung, die Herausforderungen und dein Vertrauen.

Weiterhin möchte ich allen Lehrkräften und Kindern danken, welche an der Umsetzung der vorliegenden Studie teilgenommen und somit einen ganz wesentlichen Beitrag zur Forschungsarbeit hinsichtlich curriculumbasierter Messverfahren in Deutschland geleistet haben.

Danken möchte ich auch Dr. Yvonne Blumenthal. Liebe Yvonne, du sagtest mal: „Hinter jedem starken Mann steht eine starke Frau“. Ich muss zugeben, du hast Recht. Danke, dass du meine starke Frau warst und bist. Nicht nur fachlich hast du mir Beistand geleistet, sondern auch durch ausgezeichnetes Familienmanagement.

Dank gebührt ebenso meiner Familie. Liebe Gudrun und lieber Ralf, euer elterlicher Glauben an mich und euer Interesse an meiner Arbeit haben mich ermutigt und gestärkt. Liebe Claudi, liebe Sylvia, lieber Peter, ich danke euch, dass ihr oft dazu beigetragen habt, den Spagat zwischen Privat- und Berufsleben zu halten. Ich weiß nicht, was wir manchmal ohne euch gemacht hätten.

Danke sagen möchte ich auch meinen tatkräftigen Kolleginnen und Freundinnen Katharina Ehlers, Katharina Marten und Lisa Kleinschmidt. Ihr habt Euch durch meine Berge von Seiten gelesen. Liebe Renate, du hast immer die richtige Antwort auf alle Fragen. Danke für die Geduld und Zeit.

Und schließlich, lieber Mio, hoffe ich, du verzeihst, dass ich einen beachtlichen Teil unserer gemeinsamen Zeit in deinem ersten Lebensjahr damit verbracht habe, an dieser Arbeit zu schreiben. Die für uns beste Entschädigung war, dass du dabei so oft es ging vor meinem Bauch geschlafen hast.

Inhaltsverzeichnis

1 EINFÜHRUNG 9

1.1 Erkenntnisinteresse und Ziel der Arbeit 9
1.2 Gegenstandsbestimmung und Fragestellungen 16
1.3 Aufbau der Arbeit 24

2 THEORETISCHE ASPEKTE DER ENTWICKLUNG FRÜHER MATHEMATISCHER KOMPETENZEN 27

2.1 Erstes Mengenwissen im frühen Säuglings- und Kleinkindalter 29
2.2 Entwicklung ersten Mengen- und Zahlenwissens im Vorschulalter 34
2.3 Voraussetzungen und Entwicklung der Zählfähigkeiten 37
2.4 Entwicklungsmodell früher mathematischer Kompetenzen nach Krajewski und Schneider (2006) ... 43
2.5 Modell der mathematischen Kompetenzentwicklung nach Fritz, Ricken und Gerlach (2007) 48
2.6 Das Konzept des number sense 56
2.7 Kognitionspsychologische und neurowissenschaftliche Aspekte der Entwicklung von Rechenkompetenzen 58
2.8 Weitere allgemeine kognitive Faktoren und deren Zusammenhang mit der Entwicklung von Rechenkompetenzen 66
2.9 Zentrale Elemente des mathematischen Erstunterrichts in deutschen Lehrplänen 72
2.10 Schwierigkeiten beim Erwerb früher mathematischer Kompetenzen 74
2.11 Schlussfolgerungen 86

3 DIAGNOSTIK MATHEMATISCHER KOMPETENZEN ALS ZENTRALES ELEMENT SCHULISCHER PRÄVENTION 93

3.1 Das Konzept des formative assessment im Rahmen schulischer Prävention 102
3.1.1 Eigenschaften und Anforderungen des formative assessments 106
3.1.2 Beispiele des formative assessment 109
3.1.3 Effekte des formative assessment 111
3.1.4 Ein Konzept zur strukturierten Umsetzung des formative assessment – Response-to-Instruction/Response-to-Intervention 115
3.2 Curriculumbasierte Messverfahren als Spezialform des formative assessment 138
3.2.1 Definition Curriculumbasierter Messverfahren (CBM) 138
3.2.2 CBM, CBA, DIBS, MM, GOM, SBM – Systematisierung und Abgrenzung zentraler Begriffe 142
3.2.3 Praktischer Einsatz von CBM im Bereich Mathematik 149
3.2.4 Verfügbare CBM-Verfahren für den Bereich Mathematik im US-amerikanischen Raum 158
3.2.5 Verfügbare CBM-Verfahren im deutschsprachigen Raum 165
3.2.6 Bedeutung von CBM-Verfahren 171
3.2.7 Grenzen von CBM-Verfahren 172
3.2.8 Forschungslage und -ergebnisse hinsichtlich der Konzeption und des Einsatzes von CBM-Verfahren im Bereich Mathematik 178
3.3 Schlussfolgerungen für die Konzeption und Evaluation von CBM für den mathematischen Erstunterricht 194
3.3.1 Notwendige Analysen bei der CBM-Entwicklung nach Fuchs (2004) 196
3.3.2 Notwendige Analysen hinsichtlich der von L.S. Fuchs (2004) vorgeschlagenen Stufe 1 198
3.3.3 Notwendige Analysen hinsichtlich der von L.S. Fuchs (2004) vorgeschlagenen Stufe 2 208

3.3.4 Notwendige Analysen hinsichtlich der von L.S. Fuchs (2004) vorgeschlagenen Stufe 3 211

4 CBM FÜR DEN MATHEMATISCHEN ERSTUNTERRICHT 215

4.1 Allgemeine Aspekte zur Konzeption und Evaluation von CBM für den mathematischen Erstunterricht 215
4.2 CBM „Zahlenlesen im Zwanzigerraum“ (Niveaustufe 1) 223
4.3 CBM „Zahlenreihe im Zwanzigerraum“ (Niveaustufe 1) 225
4.4 CBM „Mengenvergleich im Zwanzigerraum“ (Niveaustufe 1) 227
4.5 CBM „Arithmetische Basiskompetenzen“ (Niveaustufe 1 & 2) 228
4.6 CBM „Zahlzerlegung im Zwanzigerraum“ (Niveaustufe 2) 234
4.7 CBM „Addition im Zwanzigerraum“ (Niveaustufe 2) ... 235
4.8 CBM „Subtraktion im Zwanzigerraum“ (Niveaustufe 2) 240

5 FORSCHUNGSFRAGEN UND HYPOTHESEN DER UNTERSUCHUNG 245

5.1 Überlegungen zu Forschungsfragen der Stufe 1 (L.S. Fuchs, 2004) 245
5.2 Überlegungen zu Forschungsfragen der Stufe 2 (L.S. Fuchs, 2004) 249
5.3 Überlegungen zu Forschungsfragen der Stufe 3 (L.S. Fuchs, 2004) 250

6 METHODE DER UNTERSUCHUNG 253

6.1 Beschreibung der Stichproben 254
6.2 Untersuchungsdurchführung und -verlauf 259
6.2.1 Teilstudie 1 und 2 259

6.2.2 Teilstudie 3 267
6.2.3 Teilstudie 4 268
6.3 Erhebungsinstrumente 269
6.3.1 Kalkulie (Fritz et al., 2007) 271
6.3.2 Münsteraner Screening (MÜSC, Mannhaupt, 2005) .. 272
6.3.3 Rechenfische (Wagner & Hartke, 2006) 273
6.3.4 Würzburger Leise Leseprobe – WLLP (Küspert & Schneider, 1998) 274
6.3.5 Deutscher Mathematiktest für erste Klassen – DEMAT 1+ (Krajewski et al., 2002) 275
6.3.6 Deutscher Mathematiktest für zweite Klassen – DEMAT 2+ (Krajewski et al., 2004) 277
6.3.7 Aussagenkatalog zur Einschätzung der CBM sowie deren Einsatz 278
6.4 Statistische Methoden zur Datenauswertung 281
6.4.1 Allgemeine Angaben 281
6.4.2 Klassifikatorische Analysen 293
6.4.3 Hierarchisch-lineare Modellierung (HLM) 299
6.4.4 Angaben zur Datenauswertung 308

7 ERGEBNISSE ZUR GÜTE UND ANWENDBARKEIT DER ENTWICKELTEN CBM 309

7.1 Deskriptive statistische Angaben 309
7.1.1 CBM „Zahlenlesen im Zwanzigerraum“ 310
7.1.2 CBM „Zahlenreihe im Zwanzigerraum“ 316
7.1.3 CBM „Mengenvergleich im Zwanzigerraum“ 321
7.1.4 CBM „Arithmetische Basiskompetenzen“ 327
7.1.5 CBM „Zahlzerlegung im Zwanzigerraum“ 334
7.1.6 CBM „Addition im Zwanzigerraum“ 340
7.1.7 CBM „Subtraktion im Zwanzigerraum“ 345
7.1.8 Zusammenfassung der deskriptiven statistischen Angaben 350

7.2 Darstellung der Ergebnisse hinsichtlich der Reliabilität 351
7.2.1 Splithalf-Reliabilität 352
7.2.2 Retest-Reliabilität 356
7.2.3 Paralleltest-Reliabilität 360
7.3 Darstellung der Ergebnisse hinsichtlich der Validität .. 369
7.3.1 Augenscheinvalidität 370
7.3.2 Konstruktvalidität 371
7.3.3 Kriteriumsvaliditat 387
7.4 Darstellung der Ergebnisse hinsichtlich der Änderungssensibilität 412
7.4.1 CBM „Zahlenlesen im Zwanzigerraum“ 413
7.4.2 CBM „Zahlenreihe im Zwanzigerraum“ 420
7.4.3 CBM „Mengenvergleich im Zwanzigerraum“ 425
7.4.4 CBM „Arithmetische Basiskompetenzen“ 430
7.4.5 CBM „Zahlzerlegung im Zwanzigerraum“ 435
7.4.6 CBM „Addition im Zwanzigerraum“ 440
7.4.7 CBM „Subtraktion im Zwanzigerraum“ 444
7.5 Darstellung der Ergebnisse hinsichtlich der Ökonomie, Nutzbarkeit und Zumutbarkeit 450

8 ZUSAMMENFASSUNG DER ERGEBNISSE 463

8.1 Zusammenfassung der Ergebnisse zur Reliabilität 463
8.2 Zusammenfassung der Ergebnisse zur Validität 467
8.3 Zusammenfassung der Ergebnisse zur Änderungssensibilität 475
8.4 Zusammenfassung der Ergebnisse zur Ökonomie, Nutzbarkeit und Zumutbarkeit 478

9 DISKUSSION DER ERGEBNISSE 481

9.1 Diskussion der Ergebnisse hinsichtlich der Objektivität 481
9.2 Diskussion der Ergebnisse hinsichtlich der Reliabilität 483
9.2.1 Splithalf-Reliabilität 483

9.2.2 Retest-Reliabilität ... 485
9.2.3 Paralleltest-Reliabilität ... 487
9.3 Diskussion der Ergebnisse hinsichtlich der Validität ... 490
9.3.1 Inhaltsvalidität ... 490
9.3.2 Konstruktvalidität ... 495
9.3.3 Kriteriumsvalidität ... 498
9.4 Diskussion der Ergebnisse hinsichtlich der Änderungssensibilität ... 503
9.5 Diskussion der Ergebnisse hinsichtlich der Ökonomie, Nutzbarkeit und Zumutbarkeit ... 509
9.5.1 Ökonomie ... 509
9.5.2 Nutzbarkeit ... 511
9.5.3 Zumutbarkeit ... 515
10 FAZIT UND AUSBLICK ... 517
10.1 Kurzzusammenfassung zentraler Befunde ... 517
10.2 Aspekte zur Einschätzung der internen Validität der Befunde ... 520
10.3 Aspekte zur Einschätzung der externen Validität der Befunde ... 526
10.4 Abschließende Einordnung der Ergebnisse ... 528
10.5 Ausblick ... 533
ABSTRACT ... 541
LITERATUR ... 547
ABKÜRZUNGSVERZEICHNIS ... 619
TABELLENVERZEICHNIS ... 621
ABBILDUNGSVERZEICHNIS ... 635

1 Einführung

1.1 Erkenntnisinteresse und Ziel der Arbeit

6 Uhr, der Wecker klingelt. Aufstehen, waschen, frühstücken. Der Schulbus fährt 7 Uhr 15. –„Wie lange habe ich noch Zeit?“ Beim Frühstück stellt Mutter 2 Schalen Müsli auf den Tisch und sagt: „Such dir eine aus.“ –„Wo ist wohl am meisten drin?“ Geschafft, der Bus kommt. Ich steige als sechster ein. –„Wie viele Personen sind dann schon vor mir eingestiegen?“ Der Weg zur Schule beträgt 5 Kilometer und dauert etwa 10 Minuten. –„Fahren wir vielleicht zu schnell?“ Die Fahrkarte kostet 1,10€, die Wochenkarte 5€. –„Kann ich mit der Wochenkarte Geld sparen?“ In der Pause Fußballspielen auf dem Schulhof. Es steht 7 zu 8. –„Wer hat gewonnen?“ –Überall sind sie Begleiter – die Zahlen. Zahlen bilden ein komplexes System aus Bedeutungs- und Beziehungsebenen, das fast unablässig sehr stark den Alltag von Kindern und Erwachsenen beeinflusst. Nicht umsonst stellen mathematische Kompetenzen zusammen mit den Lese- und Rechtschreibfähigkeiten zentrale Grundpfeiler menschlicher Entwicklung und erfolgreicher Schul-, Ausbildungs- und Berufskarrieren dar (Havighurst, 1981; Tupaika, 2003).

Doch wie steht es um die Rechenfähigkeiten deutscher Schüler[1]? Die OECD hat mit dem Programme for International Student Assessment (PISA) eine Studie entwickelt, deren Ziel es ist, einen internationalen Vergleich der Kompetenzen von 15-jährigen Schülern beteiligter Länder in den Domänen der mathematischen, der naturwissenschaftlichen und der Lesekompetenz zu ermöglichen. In der Auswertung der Ergebnisse von PISA 2009 steht Deutschland mit dem Platz 13 in den Naturwissenschaften, dem Platz 16 in Mathematik und dem Platz 20 im Lesen insgesamt gesehen

[1] Aufgrund der besseren Lesbarkeit wird in der Arbeit durchgehend auf die weibliche Form verzichtet, obwohl diese selbstverständlich implizit ebenso gemeint ist.

vergleichsweise gut dar (OECD, 2010). Bezüglich der Leistungen im mathematischen und naturwissenschaftlichen Bereich sind diese Ergebnisse sogar statistisch signifikant über dem OECD-Durchschnitt einzuordnen. Betrachtet man die Ergebnisse differenzierter, wird deutlich, „dass Deutschland trotz der positiven Entwicklung in Richtung einer stärkeren Homogenisierung der Kompetenzverteilung im Vergleich zu den übrigen OECD-Staaten noch immer über die drittgrößte Streuung der Kompetenzwerte verfügt“ (Frey, Assemann, Carstensen, Ehmke & Blum, 2007, S. 266). In PISA 2009 ist Deutschland bezüglich der Streuung der Ergebnisse im Bereich Mathematik immerhin noch auf Platz 5 (vgl. OECD, 2010, Anhang B1, Tabelle I.3.3).

Überdies wurden im Rahmen der PISA-Studie Kompetenzstufen ermittelt, die dazu dienen, „Gruppen (Anteile) von Schülern zu ermitteln, die aufgrund ihrer Kompetenz sehr schlechte Chancen für nachfolgendes Lernen innerhalb und außerhalb der Schule haben. Für Schülerinnen und Schüler, die auf oder unterhalb der ersten Kompetenzstufe anzusiedeln sind, ist die Prognose für die weitere Bildungskarriere (auch bezogen auf eine berufliche Ausbildung) ungünstig“ (Prenzel, Drechsel, Carstensen & Ramm, 2004, S. 38). Die PISA- Ergebnisse der Jahre 2000, 2003, 2006 und 2009 belegen einen relativ hohen prozentualen Anteil von Schülern, deren Leistungsniveau nicht über die niedrigste Kompetenzstufe (Stufe I) in dem untersuchten Bereich der mathematischen Kompetenzen hinaus geht. In Tabelle 1 werden die veröffentlichten PISA-Ergebnisse der letzten vier Erhebungszeiträume für den Bereich der mathematischen Kompetenz zusammengefasst.

Tabelle 1: Zusammenfassung der prozentualen Anteile der Schüler unterhalb Kompetenzstufe I und auf Kompetenzstufe I in dem Bereich mathematischer Kompetenzen aus PISA 2000, 2003, 2006 und 2009

(Klieme, Neubrand & Lüdtke, 2001; Prenzel et al., 2004; Frey et al., 2007; OECD, 2010)

	Unterhalb der Kompetenzstufe I	Kompetenzstufe I	Gesamtanteil unter und auf Kompetenzstufe I
PISA 2000	7.0 %	17.0 %	24.0 %
PISA 2003	9.2 %	12.4 %	21.6 %
PISA 2006	7.3 %	12.5 %	19.8 %
PISA 2009	6.4 %	12.2 %	18.6 %

Zur Risikogruppe im Bereich Mathematik zählte im Jahr 2000 fast ein Viertel der deutschen Schüler. Das bedeutet, dass die mathematischen Fähigkeiten dieser Kinder nicht über das Grundschulniveau hinausgehen und somit weit von den Lernzielen der Sekundarstufe I entfernt sind (Blum, Neubrand, Ehmke, Senkbeil, Jordan, Ulfig et al., 2004). In den folgenden Untersuchungsjahren verringerte sich dieser Anteil zwar kontinuierlich, 2009 sind es jedoch immer noch knapp 19 % der 15-Jährigen, die lediglich über elementarste mathematische Kenntnisse verfügen.

Die Ergebnisse der international angelegten TIMS-Studie (Trends in International Mathematics and Science Study) von 2007, welche die mathematischen und naturwissenschaftlichen Bereiche in der Grundschule sowie Sekundarstufe I und II fokussiert, zeigen ein ähnliches Bild (Bos, Bonsen, Baumert, Prenzel, Selter & Walther, 2008). Ca. 20 % der untersuchten Viertklässler erreichen Leistungen entsprechend der Kompetenzstufe II, darin enthalten sind etwa 4 % von Schülern, die Leistungen erbringen, die nicht über die Kompetenzstufe I hinausgehen. Diesen Ergebnissen folgend, verfügt ein Fünftel der deutschen Viertklässler „allenfalls über elementares mathematisches Wissen sowie über elementare mathematische Fertigkeiten und Fähigkeiten. Besorgniserregend sind die mangelhaften mathematischen Leistungen der 4 % von Schülern auf der untersten Kompetenzstufe. Diese Viertklässler verfügen nur über rudimentäres mathematisches Anfangswissen […] was zu einer ausgesprochenen Belastung für die weitere Schullaufbahn führen

kann.“ (Bos et al., 2008, S. 4). Die tendenziell größten Schwierigkeiten zeigten die untersuchten Kinder dabei im Bereich „Arithmetik“, wohingegen die Ergebnisse in den Bereichen „Geometrie/Messen“ und „Daten“ besser ausfielen (Bos et al., 2008).

Studien, die nicht nur den Leistungsstand zu einem bestimmten Zeitpunkt, sondern den Entwicklungsverlauf im Längsschnitt untersucht haben, konnten zudem einen engen Zusammenhang zwischen schulischen Mathematikleistungen und dem im Vorschul- bzw. Grundschulalter erworbenen Mengen-Zahlenwissen nachweisen (Aunola, Leskinen, Lerkkanen & Nurmi, 2004; Gaupp, Zoelch & Schumann–Hengsteler, 2004; Geary Hamson & Hoard, 2000; Helmke & Weinert, 1997; Krajewski, 2003; Krajewski & Schneider, 2006; Kurdek & Sinclair, 2001; Mazzocco, Feigenson & Halberda, 2011; Mazzocco & Thompson, 2005; Stern, 2003; Weißhaupt, Peuker & Wirtz, 2006).

So zeigt bspw. die SCHOLASTIK-Studie („Schulorganisierte Lernangebote und die Sozialisation von Talenten, Interessen und Kompetenzen“) von Weinert und Helmke (1997), welche individuelle Entwicklungsverläufe von Grundschulkindern untersucht, dass mathematische Leistungen von Viertklässlern sich auf die im zweiten Schuljahr gemessenen Leistungen zurückführen lassen. Im Rahmen der LOGIK-Studie (Longitudinalstudie zur Genese individueller Kompetenzen), die die Entwicklung von Persönlichkeitsmerkmalen und schulischen Fertigkeiten vom Vorschul- bis ins frühe Erwachsenenalter untersucht, konnte Stern (2003) nachweisen, dass sich die mathematischen Schülerleistungen in Klasse zwei durch das zuvor in Klasse eins erhobene mathematische Wissen vorhersagen lassen. Überdies belegen Ergebnisse der Längsschnittstudie von Krajewski und Schneider (2006) einen mittleren Zusammenhang zwischen dem kurz vor Schuleintritt erfassten Mengen- und

Zahlenwissen, und den zukünftigen Mathematikleistungen zum Ende der ersten und der vierten Klasse.

Diese Ergebnisse zeigen auf, dass sich Entwicklungslinien zum Verlauf mathematischer Kompetenzen mit hoher Wahrscheinlichkeit schon anhand des ersten, vorschulischen und frühen schulischen Wissens über Mengen und Zahlen zeichnen lassen. Kinder mit einem gering ausgeprägten arithmetischen Vorwissen entwickeln diesen Befunden zufolge ihre mathematischen Kompetenzen langsamer als leistungsstarke Gleichaltrige (Aunola et al., 2004; Gaupp et al., 2004). Dies bedeutet für schwächere Kinder, dass es ihnen kaum gelingt, ihre Leistungen derart zu steigern, so dass ihre Leistungsposition zur Bezugsgruppe, wie z. B. der Klasse, verbessert wird (Helmke, 1997). Die Leistungsunterschiede zwischen starken und schwachen Schülern nehmen im Verlauf der Grundschulzeit und darüber hinaus weiter zu (Aunola et al., 2004; Gaupp et al., 2004; Becker, Lüdtke, Trautwein & Baumert, 2006). Dieses Phänomen wird in der einschlägigen Fachliteratur auch als Schereneffekt bezeichnet (Becker et al., 2006).

Darauf basierend wird angenommen, dass sich die Entwicklung des mathematischen Wissens kumulativ vollzieht, d h. dass mathematisches Wissen aufeinander aufbaut. Ein breit ausgeprägtes Vorwissen wirkt sich demnach positiv auf die Aneignung weiterführender mathematischer Kompetenzen aus und stellt somit eine wesentliche Voraussetzung für den schulischen Erfolg im Bereich der Mathematik dar. Schüler mit einem höher ausgeprägten mathematischen Vorwissen haben demnach bessere Chancen auf einen unbeeinträchtigten Lernverlauf. Hingegen wirken sich früh entstehende und nicht identifizierte Wissenslücken demnach negativ auf darauf aufbauende Lernprozesse aus und stellen eine Gefährdung für weiteren schulischen Lernerfolg dar (Aunola et al., 2004; Helmke & Weinert, 1997). So weist auch Lorenz (2003) darauf

hin, dass frühe Wissenslücken sich ohne entsprechende Förderung verfestigen und somit der weitere Lernerfolg in Mathematik erschwert wird. Dies hat zur Folge, dass es in höheren Klassenstufen Schüler gibt, die die grundlegenden Lerninhalte der ersten Schuljahre nicht beherrschen. Darüber hinaus geht aus den einschlägigen Längsschnittstudien hervor, dass der Entwicklungsverlauf mathematischer Leistungen stabil ist (Aunola et al., 2004; Fritz, Ricken & Gerlach, 2007; Krajewski, 2003; Krajewsi & Schneider, 2006; Peard, 2004; Weißhaupt et al., 2006). Für Kinder mit ungenügend ausgeprägten spezifischen Vorläuferfähigkeiten besteht demnach ein hohes Risiko, Minderleistungen im Rechnen zu zeigen.

Abhängig vom festgelegten Kriterium liegt die Prävalenz für Rechenschwierigkeiten zwischen 3 % und 15 % (Dilling, Mombour & Schmidt, 2011; Fritz et al., 2007; Jacobs & Petermann, 2007; Kaufmann, 2003; Schipper, 2003). Ausgehend von den Prävalenzzahlen und dem eher stabilen Entwicklungsverlauf wird deutlich, wie wichtig eine frühzeitige Identifikation von sog. Risikokindern ist, also Kindern, die bereits in der Schuleingangsstufe von Minderleistungen im Rechnen bedroht sind. Aktuell gibt es diagnostische Verfahren (Marko-D, Ricken, Fritz & Balzer, im Druck; Kalkulie, Fritz et al., 2007; Rechenfische, Knopp, 2010b), die helfen sollen, Kinder frühzeitig zu erkennen. Trotz guter Kennwerte zur Vorhersagegüte derartiger Verfahren (hier: Kognitiver Fähigkeitstest für 1. und 2. Klassen von Kawthar & Perleth [2005]; Inventar zur Erfassung der Lesekompetenz im 1. Schuljahr von Diehl & Hartke [2012], Würzburger Leise Leseprobe von Küspert & Schneider [1998]; Rechenfische von Wagner & Hartke [2006]; Deutscher Mathematiktest für erste Klassen von Krajewski, Küspert, Schneider & Visé [2002]) schlussfolgert Blumenthal (2012), dass

> *„die Fehlerquoten relativ hoch liegen und die Verfahren auch in kombinierter Betrachtung zu keiner validen Identifizierung der*

> *in Zukunft gefährdeten Schüler führen. So ist der Anteil der anhand dieser Verfahren potentiell gefährdeten Kinder ungefähr doppelt so hoch, wie die tatsächlich ermittelte Anzahl der Kinder, die zum Ende des darauf folgenden Schuljahres zu den untersten 15 Prozent in den durchgeführten Schulleistungstests gehören. Auf der anderen Seite werden Kinder als ungefährdet eingestuft, deren Entwicklungen im weiteren Verlauf jedoch von der gewünschten Norm abweichen. Das bedeutet, dass sich bereits der erste Schritt der Risikoprognose, nämlich die Identifizierung von Risikokindern, nach der bisher in der Schulpraxis vorfindbaren Vorgehensweise, als problematisch erweist"* (S. 306).

Es zeigt sich somit, dass derzeit der Einsatz gebräuchlicher Verfahren zur Frühidentifikation von Rechenschwäche nicht in ausreichendem Maß dazu führt, dass mittel- bis langfristige Prognosen über den Entwicklungsverlauf mathematischer Leistungen getroffen werden können. Um die Ergebnisse solcher Verfahren abzusichern, sollten diagnostische Informationen zum Leistungstand eines Schülers demnach ergänzend in kürzeren Abständen wiederholt und curriculumsnah erhoben werden. Anhand einer derartig hochfrequentierten Dokumentation des Lernverlaufs ist es möglich, den Lernprozess zu überwachen bzw. laufende Lehr- und Lernprozesse zu evaluieren und kurzfristig Unterrichts- bzw. Förderadaptionen vorzunehmen. Eine derartige Vorgehensweise hat sich seit den 1980er Jahren in den USA unter der Bezeichnung curriculum-based measurements bzw. curriculumbasierte Messungen (CBM) etabliert (Deno, 1985, 2003a, 2003b). In Deutschland befindet sich die Entwicklung und der Einsatz von CBM im Bereich Mathematik derzeit noch in den Anfängen (Diehl & Hartke, 2007; Klauer, 2006; Strathmann & Klauer, 2008, 2010; Walter, 2008b, 2009a, 2009b). Zentrales Ziel der vorliegenden Dissertation ist es, die Güte und Anwendbarkeit von in Anlehnung an US-amerikanische Verfahren entwickelte CBM im Bereich der Arithmetik zu prüfen, um die Identifikation von Kindern mit Risiken in ihrer mathematischen

Leistungsentwicklung für Grundschullehrkräfte zu optimieren. Die vorliegende Untersuchung trägt somit zur Entwicklung von Verfahren zur Früherkennung von Schülern mit mathematischen Minderleistungen bei, d. h. sie dient einer der zentralen Aufgaben der Präventionsforschung (Hartke & Koch, 2008).

Integriert ist dieses Forschungsanliegen in das aktuell laufende Drittmittelprojekt „Prävention und Integration im RTI-Paradigma“ des Instituts für Sonderpädagogische Entwicklungsförderung und Rehabilitation der Universität Rostock unter der Leitung von Prof. Hartke (Diehl, Mahlau, Voß & Hartke, 2010, 2012; Mahlau, Diehl, Voß & Hartke, 2011; Voß, Blumenthal, Diehl, Ehlers, Mahlau & Hartke, 2012).

1.2 Gegenstandsbestimmung und Fragestellungen

Wie im Abschnitt zuvor herausgearbeitet, haben Kinder, deren Schwächen frühzeitig erkannt sowie spezifisch und kontinuierlich gefördert werden, eine gesteigerte Entwicklungschance in Hinblick auf eine normgerechte Leistungsentwicklung. Bezogen auf den Schuleingangsbereich sollten demnach präventive Maßnahmen im Fokus der pädagogischen Arbeit stehen, die darauf abzielen, Lernbedingungen zu schaffen, die allen Kindern bestmögliche Entwicklungschancen bieten und dazu beitragen, Wissenslücken rechtzeitig zu schließen, um manifesten Lernproblemen vorzubeugen (Hartke, 2005; Kretschmann, 2003; Lorenz, 2003). Präventive Maßnahmen müssen dabei differenziert hinsichtlich ihrer zeitlichen Verortung sowie ihres Ziels betrachtet werden.

Die vom amerikanischen Psychiater Caplan (1964) vorgeschlagene Unterteilung präventiver Maßnahmen in *primäre*, *sekundäre* und *tertiäre Prävention* hat sich auch in der pädagogisch-psychologischen

und sonderpädagogischen Fachliteratur durchgesetzt. Dabei umfasst die primäre Prävention Maßnahmen zur allgemeinen pädagogischen Vorsorge, welche vor dem tatsächlichen Auftreten und der Manifestation einer Störung eingesetzt werden. Die Maßnahmen dienen einer Reduktion von Risikofaktoren (Caplan, 1964). So ist bspw. denkbar, dass eine Sensibilisierung von Schülern hinsichtlich der Diversität in der Klasse die Gefahr der sozialen Ausgrenzung einzelner Schüler senken kann. Sekundäre Prävention „setzt bei bereits manifesten Störungen an, mit dem Ziel, durch Verkürzung der Störungs- bzw. Erkrankungsdauer die Prävalenzrate von psychischen Störungen in der Population zu senken“ (Brandtstädter, 1982b). Als Maßnahmen sekundärer Prävention gelten Vorbeugungsprogramme für bestimmte Risikogruppen im Sinne der Früherkennung und Hilfe (Caplan, 1964), wie z. B. der Einsatz von Diagnoseverfahren zur Ermittlung von Rechenschwierigkeiten und der anschließenden gezielten Förderung. Tertiäre Prävention setzt an, wenn sich eine Beeinträchtigung bereits manifestiert hat. Es handelt sich um prophylaktische Maßnahmen zur Minderung von Folgeschäden (Caplan, 1964). Ein Beispiel für tertiäre Prävention ist die sonderpädagogische Arbeit mit lese-rechtschreibschwachen Kindern zur Verhinderung kombinierter schulischer Störungen.

Werden präventive Handlungen hinsichtlich ihres Ziels unterschieden, wird von spezifischen oder unspezifischen Präventionsmaßnahmen gesprochen (Perrez, 1991). So fokussiert spezifische Prävention konkrete Probleme bzw. eine bestimmte Risikogruppe mit dem Ziel spezielle Störungen zu verhindern (Perrez, 1991). Dies setzt Kenntnisse darüber voraus, „durch welche Bedingungen der zu verhindernde Zustand begünstigt bzw. herbeigeführt wird (Ätiologie) und ferner, durch welche Handlungen (Intervention) diese Bedingungen praktisch unterbunden werden können“ (S. 93). Unspezifische Präventionshandlungen hingegen richten sich nicht auf ein bestimmtes Störungsbild, sondern dienen

eher der Verbesserung allgemeiner Rahmenbedingungen zur Verringerung eines Risikos auftretender Beeinträchtigungen (Perrez, 1991). Da der Terminus unspezifische primäre Prävention meist so weit gefasst wird, dass darin jegliche Bemühungen zur Verbesserung der Lebensumstände impliziert sind, schlägt Hartke (2005) stattdessen den Begriff *Grundprävention* vor. Der Begriff der primären Prävention erhält dadurch einen spezifischeren Fokus (Hartke, 2005).

In Anlehnung an Hartke (2005) kann resümierend festgehalten werden, dass sich Prävention im schulischen Setting, bezeichnet als *schulische Prävention*, als ein Ensemble aus unspezifischen Maßnahmen zur Prophylaxe schulischer Minderleistungen für alle Kinder einer Klasse oder Schule (primäre Prävention) sowie spezifischen Maßnahmen für Schüler einer bestimmten Risikogruppe (sekundäre Prävention) zusammensetzt, um manifesten Lern- bzw. Verhaltensstörungen vorzubeugen (S. 13). Weiterhin kann auf Basis zahlreicher Studien zur Schulleistungsentwicklung (Gaupp et al., 2004; Helmke & Weinert, 1997; Weißhaupt et al., 2006) der Schluss gezogen werden, dass solche präventiven Bemühungen möglichst früh, also im vorschulischen Bereich und in der Schuleingangsphase einsetzen müssen.

Im Sinne spezifischer Prävention nach Perrez (1991), ist es darüber hinaus notwendig zu verstehen, wie sich erste mathematische Fähigkeiten entwickeln, um auftretende Lücken im Bereich der mathematischen Leistungsentwicklung durch gezielte spezifische Förderung zu schließen. In der Fachliteratur gibt es bis heute keinen gültigen Konsens darüber, wodurch Rechenstörungen verursacht werden. Aktuell diskutierte Ansätze entstammen der Entwicklungs-, der Neuro-, und der Kognitionspsychologie (vgl. u. a. Dehaene, 1992; von Aster, 2003; Krajewski, 2003; Grube & Barth 2004; Krajewski & Schneider, 2006; von Aster, Schweiter & Weinhold Zulauf, 2007; Fritz et al., 2007). Insgesamt kann davon ausgegangen werden, dass es

sich bei den Ursachen einer Rechenstörung um ein multifaktorielles Bedingungsgeflecht handelt (Fritz & Ricken, 2009).

Folgt man den aktuellen Forschungen der Entwicklungspsychologie ergeben sich Schwierigkeiten im Rechnen aufgrund fehlender bzw. nicht genügend ausgeprägter Einsichten und Erkenntnisse grundlegender mathematischer Konzepte, wodurch ein sicherer Umgang mit Mengen und Zahlen und das Entwickeln effektiver Rechenstrategien erschwert bzw. unmöglich wird (Krajewski & Schneider, 2006; Fritz et al., 2007). Krajewski und Schneider (2006) sowie auch Fritz et al. (2007) haben die Systematik dieser Einsichten in Entwicklungsmodelle zur Entstehung mathematischer (arithmetischer) Kompetenzen zusammengefasst, welche in Abschnitt 2.4 bzw. 2.5 detaillierter thematisiert werden.

Ausgehend von diesen entwicklungspsychologischen Modellen wird deutlich, dass die Förderarbeit im Bereich der Mathematik rechenschwache Kinder beim Erreichen zentraler Entwicklungsstufen unterstützen soll, also beim Erwerb wesentlicher mathematischer Einsichten und Fähigkeiten (Ricken, 2009). Um spezifische Unterrichts- bzw. Förderangebote für die Schüler planen zu können, ist eine Verortung der Schülerleistungen auf diesen Entwicklungsstufen notwendig. Dies soll auf Grundlage entsprechender diagnostischer Informationen geschehen. Eine enge Verbindung diagnostischer Maßnahmen zur Identifikation von Entwicklungsrisiken und zur Überwachung des Lernprozesses sowie sich anschließender, angepasster spezifischer Förderbemühungen bildet einen zentralen Aspekt schulischer Prävention. So schlussfolgert auch Hartke (2005): „Entgegen häufig vertretenen Auffassungen in der deutschsprachigen Literatur sind insbesondere diejenigen Förderprogramme erfolgreich, die strukturiert und systematisch fördern [Herv. d. Verf.], Lerndefiziten und Verhaltensproblemen konzeptionell spezifisch [Herv. d. Verf.]

begegnen und in ihrem Prozessverlauf diagnostische Kontroll- und Evaluationsschleifen [Herv. d. Verf.] integrieren“ (S. 34).

Prozessbegleitende diagnostische Kontroll- und Evaluationsschleifen sind in diesem Kontext als wiederholt eingesetzte Maßnahmen zu verstehen, die der Überprüfung der Wirksamkeit angebotener Unterrichts- bzw. Fördermaßnahmen dienen. Eine Unterscheidung von Evaluationsmaßnahmen im Kontext von Lernprozessen in *formativ* bzw. *summativ* wurde 1967 durch Scriven vorgenommen. Dabei bezieht sich formative Evaluation auf eine fortlaufende, prozessbegleitende Dokumentation, ganz im Sinne des Postulats Hartkes. Sie erlaubt frühzeitige Eingriffe in den Entwicklungsgang während des Lernverlaufs, falls Schwierigkeiten auftreten. Dazu können regelmäßig wiederholt standardisierte Messungen zur Erfassung des Lernfortschritts verwendet werden, um Entwicklungen wie Zunahme, Stagnation oder auch Rückschritte aufzudecken. Ausgehend von den Ergebnissen der formativen Evaluation wird demnach Einfluss auf den Lernprozess selbst genommen, um ungünstige Verläufe zu vermeiden (Scriven, 1967). Hingegen umfasst summative Evaluation Maßnahmen, die eine abschließende Wertung der Ergebnisse eines Prozesses ermöglichen sollen. Es handelt sich dabei um Verfahren, welche einmalig, zum Ende eines Prozesses eingesetzt werden, um zu prüfen, ob ein Ziel in welchem Umfang erreicht wurde (ebd.). Schulleistungstests oder lehrzielorientierte Tests am Ende eines Schuljahres, auch Statusdiagnostik genannt, zählen zur summativen Evaluation (Klauer, 2006). Aufgrund der Tragweite der Auswirkungen ihrer Ergebnisse müssen statusdiagnostische Verfahren standardisiert sein, d. h. testtheoretischen Gütekriterien der Objektivität, Reliabilität und Validität genügen.

Die Ergebnisse der Diagnostik sollen nicht nur dazu dienen, den Risikostatus eines Kindes zu identifizieren, sondern zudem dabei

helfen, die zukünftige Leistungsentwicklung vorherzusagen. Aus dieser Prognose leitet sich ab, ob ein Kind spezifischer Hilfsangebote bedarf und im Idealfall auch wie diese inhaltlich und methodisch aussehen sollen. Prognosen stellen daher die Grundlage für Prävention dar (Brandtstädter, 1982a, b) und sind ein zentrales Element zwischen Diagnose und Förderung (Krapp, 1993). Langfristige Prognosen werden in ihrer Gültigkeit von zwei wesentlichen Aspekten limitiert. Zum einen durch die multivariate Bedingtheit von Schulleistungen, zum anderen durch relativ hohe Irrtumswahrscheinlichkeiten derzeit vorliegender diagnostischer Verfahren zur Früherkennung von Risikokindern (Blumenthal, 2012). Kinder, die durch Fehlklassifikationen zu Unrecht einen Risikostatus erwerben (falsch positive Klassifikation), erliegen oft den aktuell noch üblichen Separierungsprozessen des deutschen Schulsystems (z.B. der Zuweisung in eine Förderschule). Werden vorherrschende Entwicklungsrisiken bei Kindern hingegen nicht erkannt (falsch negative Klassifikation), greifen keine spezifischen Unterstützungsmaßnahmen, so dass Entwicklungsrückstände erst identifiziert werden, wenn sie sich zu einem deutlichen Schulversagen ausgeweitet haben. Dieses Vorgehen wird von Vaughn und L.S. Fuchs (2003) als *wait-to-fail* (sinngemäß: Abwarten bis zum Scheitern) bezeichnet. Es wird in der Fachliteratur als für die Betroffenen wenig hilfreich kritisiert (Fletcher, Coulter, Reschly & Vaughn, 2004; D. Fuchs, Mock, Morgan & Young, 2003; Gersten & Dimino, 2006; Vaughn & L.S. Fuchs, 2003).

Der Problematik der Fehlklassifikationen – gleich ob falsch positiv oder falsch negativ – müssen sich vor allem statusdiagnostische Maßnahmen stellen, welche den Anspruch erheben, aufgrund einmaliger Messungen langfristige valide Prognosen zu stellen. Und obwohl Statusdiagnostik nicht die Prozesskontrolle gewährleisten und somit nur bedingt Aussagen über Leistungsentwicklungen treffen kann, werden pädagogische Entscheidungen in der aktuellen

schulischen Praxis, wie z. B. die Notwendigkeit der Beschulung in einer Förderschule basierend auf statusdiagnostischen Urteilen getroffen. Speziell für Kinder mit Hinweisen auf einen ungünstigen Entwicklungsverlauf ihrer mathematischen Leistungen gibt Ricken (2009) diesbezüglich zu bedenken:

> *„Mit Blick auf die Kinder, die ein Risiko für die Entwicklung einer Rechenstörung* [Herv. d. Verf.] *tragen, reicht es aber nicht aus, summative Effekte zu betrachten […] Hier müssen sehr differenziert Wirkungen beobachtet werden, und im Sinne einer formativen Evaluation* [Herv. d. Verf.] *sind Förderziele und -maßnahmen zu verändern und anzupassen“* (S. 125).

Ein innovatives und aktuell viel diskutiertes Konzept zur Prävention, das versucht diese Probleme zu überwinden und Schülerleistungen zu verbessern ist ein in den USA verbreitetes Förderkonzept *Response-to-Instruction* bzw. *Response-to-Intervention* (RtI). Der Begriff Response-to-Instruction bzw. Response-to-Intervention spiegelt dabei den folgenden Gedanken wider: Abhängig von der Reaktion, der Antwort (Response) des Kindes auf den Unterricht (Instruction) bzw. auf die Förderung (Intervention) kann ermittelt werden, ob die pädagogischen Bemühungen der beteiligten Lehrpersonen Früchte tragen. D. h. Probleme beim Erlernen schulischer Inhalte werden nicht zuerst im Kind, sondern in der didaktischen und methodischen Gestaltung des Unterrichts bzw. der Förderung gesucht. Entsprechend muss bei ausbleibenden Lernerfolgen der Unterricht in der Weise modifiziert bzw. optimiert werden, dass er den individuellen Voraussetzungen eher entspricht.

Es gibt nicht *das* RTI-Konzept als solches, vielmehr muss es als Gerüst verstanden werden, welches folgende Kernelemente miteinander verbindet (vgl. Abbildung 1, National Center on Response to Intervention, 2010):

- eine Kombination diagnostischer Verfahren zur Feststellung von Entwicklungsrisiken sowie zur Dokumentation von Lernverläufen (Lernfortschrittsdokumentation), auf deren Grundlage Förderentscheidungen getroffen werden (datengeleitete Praxis),
- ein mehrere Ebenen übergreifendes Fördersystem, das flexibel auf aktuelle Förderbedürfnisse der Kinder reagieren kann (Mehrebenenprävention) und
- der Einsatz von Unterrichts- und Fördermaßnahmen sowie -materialien, welche nachweislich, zu Leistungssteigerungen der Kinder führen (Evidenzbasierte Praxis).

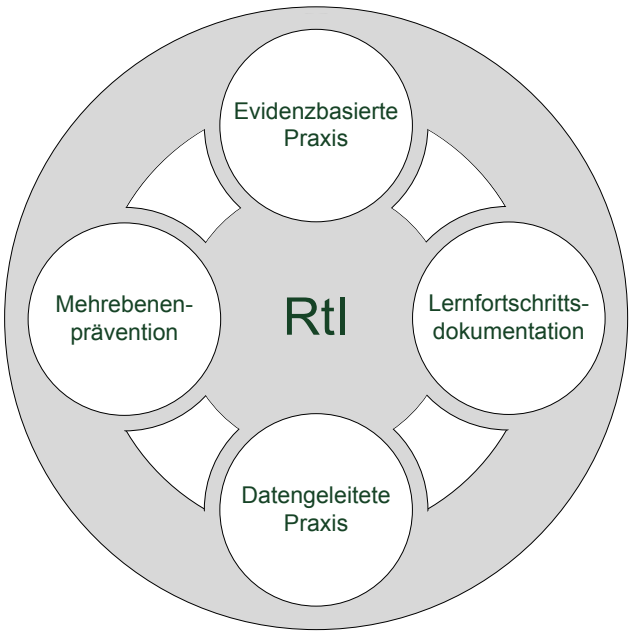


Abbildung 1: Kernkomponenten des Response-to-Intervention- Ansatzes (adaptiert und modifiziert nach National Center on Response to Intervention, 2010)

Besonders innovativ erscheinen die damit verbundenen, von Deno (1985) entwickelten, oben bereits erwähnten CBM. Darunter sind Student Monitoring Systeme (zu Deutsch: Schülerbeobachtungssysteme) zu verstehen, welche es ermöglichen sollen, Lernfortschritte von Schülern mittels regelmäßig wiederholter Messungen zu quantifizieren und den sich abzeichnenden Verlauf zu

dokumentieren, um daraus Schlüsse über die Effektivität des Unterrichts oder der Förderung bzw. auf das Ansprechen des Schülers darauf zu ziehen. Diese formellen Kurztests von wenigen Minuten Dauer sind nah am Curriculum, mit besonderem Fokus auf die testtheoretischen Gütekriterien konstruiert und lassen sich gut in den Schulalltag integrieren (Deno, 1985, 2003a, 2003b).

Offen ist bislang, ob ein solches RtI-Konzept bzw. einzelne Kernkomponenten auch auf das deutsche Bildungssystem übertragen werden können. In der hier vorliegenden Arbeit soll daher geprüft werden, inwieweit sich curriculumbasierte Messungen als Verfahren zur Dokumentation von Lernfortschritten als ein Kernelement von RtI auch im deutschen Schulsystem zur Prävention von Rechenschwäche anwenden lassen. Daher soll in der vorliegenden Studie, den folgenden Fragen nachgegangen werden:

- Genügen die Adaptionen US-amerikanischer CBM Testgütekriterien wie Objektivität, Reliabilität und Validität sowie den Kriterien der Parallelität und Veränderungssensitivität?
- Inwieweit sind Adaptionen US-amerikanischer CBM geeignet, den Lernfortschritt deutscher Erstklässler abzubilden und können sie frühzeitig Hinweise auf eine ungünstige Entwicklung mathematischer Leistungen geben?
- Ist eine erfolgreiche Implementation von Adaptionen US-amerikanischer CBM im deutschen Schulsystem möglich?
- Können Lehrkräfte auf Grundlage der erhobenen CBM-Daten der Schüler Hinweise für die weitere Unterrichts- bzw. Fördergestaltung ziehen?

1.3 Aufbau der Arbeit

Um diesen Fragen nachzugehen, werden im ersten Teil der vorliegenden Arbeit theoretische Aspekte der Entwicklung früher mathematischer Kompetenzen erläutert. Hierfür wird die aktuelle Fachliteratur zur mathematischen Kompetenzentwicklung vom Vorschul- bis zum frühen Grundschulalter analysiert. Wesentliche Ergebnisse und Auffassungen sowie Modellvorstellungen werden beschrieben und es wird darüber hinaus überblickartig dargestellt, wie diese Aspekte in den Lehrplänen der deutschen Bundesländer verankert sind. Anschließend werden mögliche Schwierigkeiten beim Erlernen zentraler mathematischer Einsichten und Kompetenzen aufgezeigt. Im Kapitel 1 wird die Bedeutung der Diagnostik zur objektiven, reliablen und validen Erfassung der Lernstände der Schüler erläutert. Diagnosekonzepte zur Früherkennung von Lernschwierigkeiten werden als zentrales Element der schulischen Prävention herausgearbeitet. Dabei wird vor allem das prozessbezogene Diagnosekonzept des formative assessment zur Dokumentation von Lernfortschritten und zur Evaluation und Modifikation von Unterrichts- und Fördermaßnahmen betrachtet. Anschließend wird differenziert auf Curriculumbasierte Messverfahren (CBM) als eine Spezialform des formative assessment und in den USA mittlerweile bewährte Methode der Lernverlaufsdiagnostik eingegangen. Neben einer Verortung von CBM in den Maßnahmen des formative assessment werden theoretische Aspekte bei der Konzeption, sowie Aspekte zur praktischen Umsetzung solcher CBM im schulischen Alltag beleuchtet. Zudem wird ein Überblick über verfügbare Verfahren gegeben. Anschließend werden die derzeitigen Forschungsstände zur wissenschaftlichen Güte und zur Effektivität verfügbarer CBM in den USA und im deutschsprachigen Raum referiert. Das Kapitel schließt mit der Vorstellung eines systematischen Evaluationsprogramms zur Erforschung von CBM-Verfahren nach L.S. Fuchs (2004) und den daraus folgenden Implikationen zur Methodik der vorliegenden Arbeit. Auf Basis der Ausführungen zum

Erlernen arithmetischer Kenntnisse und zur Bedeutung der Lernverlaufsdiagnostik als ein zentrales Element der schulischen Prävention von Rechenschwächen werden in Kapitel 4 die innerhalb der vorliegenden Studie erarbeiteten CBM vorgestellt und in einen theoretischen Kontext eingebettet. Die dieser Studie zugrundeliegenden forschungsleitenden Fragen und Hypothesen werden in Kapitel 5 dargestellt. Anschließend wird im Kapitel 6 die Methodik der vorliegenden Studie beschrieben. Es folgt die Darstellung der Ergebnisse in Kapitel 7, welche abschließend im Kapitel 8 hypothesenbezogen zusammengefasst werden. In Kapitel 9 werden die Forschungsergebnisse der vorliegenden Studie in Hinblick auf die Forschungshypothesen diskutiert. Inwieweit die hier erarbeiteten CBM den testtheoretischen Gütekriterien genügen, inwieweit es möglich ist, mit ihrer Hilfe die Lernverläufe von Kindern zu dokumentieren und inwieweit eine erfolgreiche Implementation in das deutsche Schulsystem gelingt, wird methodenkritisch und vor dem Hintergrund zuvor dargestellter theoretischer Ausführungen zur Entwicklung mathematischer Kompetenzen und zur formativen Leistungserfassung im Rahmen schulischer Prävention eingeschätzt. Den Abschluss der Arbeit stellen ein Fazit der gewonnenen Erkenntnisse, ein Ausblick sowie wesentliche Implikationen für die Forschung, die schulische Praxis und die Schulpolitik dar.

2 Theoretische Aspekte der Entwicklung früher mathematischer Kompetenzen

Zu den arithmetischen Fähigkeiten von Kindern bei Schuleintritt gibt es eine Reihe von Untersuchungen (Grassmann, Mirwald, Klunter & Veith, 1995; Grassmann, Klunter, Köhler, Mirwald, Raudies & Thiel, 2002; Hengartner & Röthlisberger, 1995; van Heuvel-Panhuizen, 1995; Kesting, 2005; Knapstein & Spiegel, 1995; Rinkens, 1997; Rinkens & Eilerts, 2001; Schipper, 1996, 1998; Schmidt, 1982; Selter, 1995). Anfang der 80er Jahre untersuchte Schmidt (1982) als erster im deutschsprachigen Raum das mathematische Wissen von 1138 Grundschulkindern zum Schuleintritt. Inhalte der geführten Einzelinterviews waren verbale Zählfähigkeiten, das Zuordnen von Zahlen zu vorgegebenen Mengen, das Zuordnen von Mengen zu vorgegebenen Zahlen sowie das Ziffernschreiben und -lesen. Schmidt stellte fest, dass knapp die Hälfte der in der Studie untersuchten Schulanfänger schon bis 29 zählen konnte. Knapp ein Drittel der Kinder konnte zumindest acht der Ziffern 0 bis 9 korrekt schreiben. Bis auf die Zahlen 6, 8 und 9 wurden alle Zahlen im Zahlenraum bis 10 von mindestens 90 % der untersuchten Schüler richtig vorgelesen. Die von Schmidt publizierten Ergebnisse können als Beginn für eine Reihe weiterer Untersuchungen zur mathematischen Lernausgangslage von Kindern zum Eintritt in die Grundschulzeit angesehen werden. Trotz unterschiedlicher Forschungsdesigns, kommen die oben genannten Studien im Wesentlichen zu vergleichbaren Resultaten. Sie belegen, dass der Erwerb erster arithmetischer Kompetenzen nicht erst bei Eintritt in die Schule beginnt, sondern dass viele Schulanfänger, entgegen der Einschätzung von Lehrkräften (Rinkens & Eilerts, 2001), über ein umfangreich ausgeprägtes mathematisches Vorwissen verfügen. Da nicht jedes Kind zur Einschulung ein ähnlich breites Vorwissensspektrum zeigt, können hier keine allgemeinen Schlüsse gezogen werden. Die aufgeführten Studien weisen vielmehr auf hohe

interindividuelle Differenzen im mathematischen Vorwissen hin. So schlussfolgert Schmidt (2009):

> *„Die in der fachdidaktischen Literatur zu findenden statistischen Ergebnisse sollten zwar als Beleg dafür genommen werden, dass es grundsätzlich nicht abgängig ist, von einer vermeintlichen Stunde null im Hinblick auf die arithmetischen Kompetenzen von Kindern am Schulanfang auszugehen; wichtiger freilich ist es, die in einer wirklich-konkreten Klasse versammelten Ausprägungen dieser arithmetischen Kompetenzen in ihrem gesamten Spektrum auszumachen, und zwar möglichst prozessorientiert (vs. bloß ergebnisorientiert)“* (S. 77).

Damit rücken unter anderem folgende Fragen in den Fokus: Wie entwickeln sich mathematische Kompetenzen? Welches sind zentrale Einsichten? In welcher Beziehung stehen sie zueinander? Wodurch sind die Unterschiede in den Entwicklungsständen gleichaltriger Kinder begründet?

Die Beantwortung dieser zentralen Fragen ist Ziel des nächsten Kapitels. Ausgehend von den Erkenntnissen zur Entwicklung erster mathematischer Konzepte im Säuglingsalter (Abschnitt 2.1) werden Annahmen zur Entwicklung erster Mengenvorstellungen in Anlehnung an Resnick (1989, 1992) (Kapitel 2.2) sowie Ausarbeitungen zu den Voraussetzungen und zum Prozess der Entwicklung von Zählfähigkeiten nach Gelman und Gallistel (1978) sowie Fuson (1988) (Kapitel 2.3) referiert. Im Anschluss daran werden in Abschnitt 2.4 und Abschnitt 2.5 die Entwicklungsmodelle mathematischer Kompetenzen nach Krajewski und Schneider (2006) sowie nach Fritz et al. (2007) erläutert. Abschnitt 2.7 gibt einen zusammenfassenden Überblick über in der kognitionspsychologischen Literatur berichtete Aspekte mit Relevanz für die Entwicklung mathematischer Kompetenzen. Im Mittelpunkt des Abschnitts 2.8 stehen mit der Entwicklung mathematischer

Kompetenzen verbundene kognitive Korrelate. In Abschnitt 2.10 wird ein Überblick über zentrale Lernplaninhalte der deutschen Bundesländer für den Bereich Arithmetik gegeben. Abschnitt 2.10 führt die Modellvorstellungen vorangegangener Kapitel zusammen, um Erklärungsansätze für Schwierigkeiten im Erwerb mathematischer Kompetenzen besser nachvollziehbar zu machen.

2.1 Erstes Mengenwissen im frühen Säuglings- und Kleinkindalter

Es gibt einige Hinweise darauf, dass selbst Säuglinge über die Kompetenz verfügen, Mengen zu unterscheiden. So berichten Starkey und Cooper (1980), dass sie auf Grundlage des Habituationsparadigmas nachweisen konnten, dass vier bis fünf Wochen alte Säuglinge auf Bildkarten präsentiere Punktmengen (zwei vs. drei Punkte) voneinander unterscheiden können. Dabei machten sich die Autoren zu Nutze, dass Säuglinge die Umwelt visuell erkunden, wobei die Zeitspanne des Fixierens eines Reizes bzw. Sachverhalts abhängig von dessen Bekanntheit ist. Kinder in diesem Alter richten ihren Blick länger auf unbekannte, als auf bereits bekannte (habituierte) Reize. Zu Beginn der Untersuchung wurde den Kindern wiederholt die Abbildung einer bestimmten Punktmenge gezeigt. Dabei wurde die Aufmerksamkeitsspanne der Kinder auf die Vorlage gemessen. Es ließ sich festhalten, dass die Säuglinge sich mit der Zeit an die Abbildung (die Anzahl der Punkte) gewöhnen, sie habituieren. Bei Vorlage einer anderen Abbildung (drei Punkte) erhöhte sich die Zeitspanne des Fixierens auf die Abbildung. Starkey und Cooper (1980) schlossen daraus, dass die Kinder in der Lage sind, die Unterschiede der auf den Bildkarten präsentierten Mengen zu erkennen. Da die Bildkarten nacheinander und nicht gleichzeitig präsentiert wurden, gingen Starkey und Cooper (1980) zusätzlich davon aus, dass die Kinder, um einen Vergleich zu realisieren, über eine innere Repräsentation der gezeigten Mengen verfügen müssen.

Zu ähnlichen Ergebnissen kamen auch weitere Forschungsgruppen (Antell & Keating, 1983; Starkey, Spelke & Gelman, 1990; Strauss & Curtis, 1981; van Loosbroek & Smitsman, 1990).

Obgleich die hier vorgestellten Ergebnisse von den Autoren als Nachweis dafür angesehen werden, dass schon Neugeborene über ein angeborenes Mengenwissen verfügen, sind die Studien dennoch kritisch zu betrachten. So führen Clearfield und Mix (2001) sowie Feigenson, Carey und Spelke (2002) in ihren ähnlich aufgebauten Studien die verlängerten Fixierzeiten der Kinder nicht auf die konkreten Mengenunterschiede zurück, sondern auf die Unterschiede der physikalischen Ausdehnung der präsentierten Mengen. Insgesamt lässt sich somit festhalten, dass die hier angeführten Studien keine eindeutigen Hinweise darauf geben, dass Neugeborene über frühe numerische Kompetenzen verfügen. Vielmehr zeigen die Ergebnisse von Clearfield und Mix (2001), Feigenson, Carey & Spelke (2002) sowie Feigenson, Carey & Hauser (2002), dass Mengendiskriminationen in diesem Alter anhand der räumlichen Ausdehnung der Mengen und nicht anhand ihrer Mächtigkeit, also nicht als Ergebnis eines ausgeprägten Mengenkonzeptes, vorgenommen werden.

In einer Studie von Wynn (1992) wurden fünf Monate alten Kindern mithilfe von Puppen einfache Additions- und Subtraktionsaufgaben vorgeführt. Wie in Abbildung 2 verdeutlicht, wurde zur Demonstration der Additionsaufgabe 1 + 1 hinter einer aufklappbaren Wand zu einer Puppe, die den Kindern vorher gezeigt wurde, eine weitere Puppe hinzugefügt. Zur Demonstration der Subtraktionsaufgabe 2 – 1 wurde hinter dem Sichtschutz von zwei Puppen, die den Kindern vorher gezeigt wurden, eine weggenommen. Dabei wurden mögliche (Addition: beim Aufklappen der Wand sind zwei Puppen sichtbar, Subtraktion: beim Aufklappen der Wand ist eine Puppe sichtbar) und unmögliche Ergebnisse (Addition: beim Aufklappen der Wand ist eine

Puppe sichtbar, Subtraktion: beim Aufklappen der Wand sind zwei Puppen sichtbar) präsentiert. Es zeigt sich, dass die untersuchten Säuglinge längere Zeit auf die unmöglichen Ergebnisse als auf die möglichen Ergebnisse blickten. Folglich, so schloss Wynn (1992), verfügen Neugeborene offenbar bereits über rudimentäres Wissen bezüglich additiver bzw. subtraktiver Situationen. Darüber hinaus dient dieses Experiment als Hinweis dafür, dass die Kinder über ein gewisses Repräsentationsvermögen der gezeigten Mengen verfügen, da die Puppen hinter einem Sichtschutz verdeckt wurden.

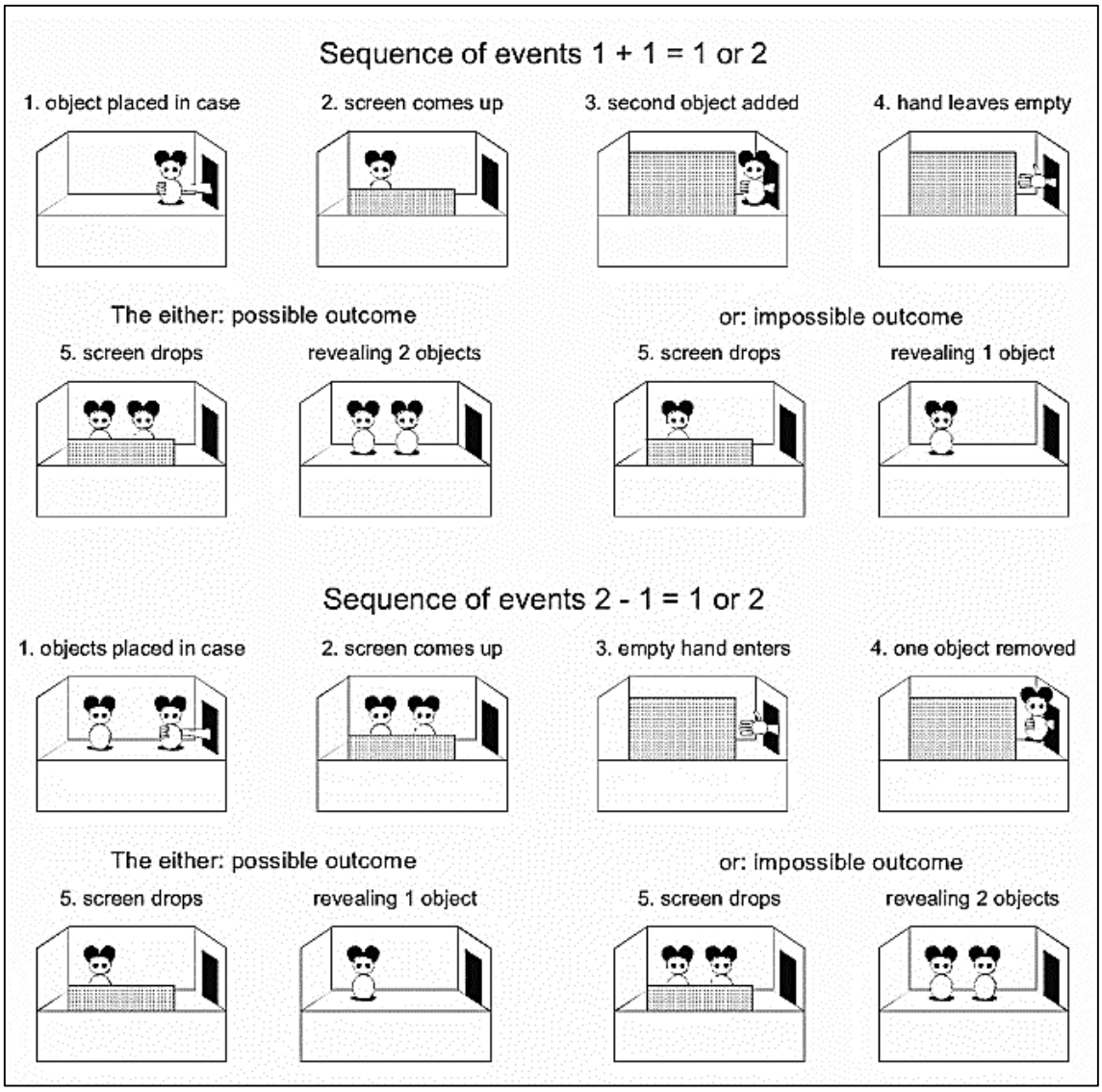


Abbildung 2: Versuchsaufbau zur Darstellung der Additionsaufgabe 1 + 1 und der Subtraktionsaufgabe 2 – 1 für Säuglinge nach Wynn (1992) (entnommen aus: Wynn, 1992, S. 749)

Feigenson, Carey und Hauser konnten 2002 in ihrer Untersuchung bei 10 bis 12 Monate alten Kindern eine Sensitivität für kleine Mengen nachweisen. In dieser Studie wurde nicht das Habituationsparadigma genutzt. Vor den Augen der Kinder wurden unterschiedlich viele Kekse, einzeln und nacheinander in zwei Dosen getan. Anschließend durften die Kinder sich eine Dose aussuchen und die darin enthaltenen Kekse behalten. In einem Verhältnis von ein zu zwei Keksen und von zwei zu drei Keksen, wählten die untersuchten Kinder sicher die Dose mit den meisten Keksen. Bei Verhältnissen

von drei zu vier, zwei zu vier oder drei zu sechs gelang dies nicht. Da die Mengen (hier Kekse) auch bei dieser Studie nicht sichtbar sind, liegt der Schluss nahe, dass die Kinder eine präzise mentale Repräsentation zumindest von kleinen Mengen aufbauen können.

2007 berichten Xu und Arriaga von Kindern im Alter von sechs Monaten, welche Bildkarten mit vier von solchen mit acht Punkten, Bildkarten mit acht von solchen mit 16 Punkten oder auch Karten mit 16 von solchen mit 32 Punkten unterscheiden können. Dies gelingt nicht bei Karten mit 4 vs. 6, 8 vs. 12 oder 16 vs. 24 Punkten. Offensichtlich sind größere Mengen in diesem Alter noch unpräzise repräsentiert, wodurch eine solche Anzahlunterscheidung abhängig vom Verhältnis der gezeigten Mengen zueinander ist. Anzahlen im Verhältnis 1-zu-2 können sicher unterschieden werden, Anzahlen in einem anderen Verhältnis wie bspw. 2-zu-3 jedoch noch nicht (Xu & Arriaga, 2007). Ab zehn Monaten können Anzahlen im Verhältnis 2-zu-3 diskriminiert werden, mit etwa fünf Jahren können Mengen im Verhältnis 1-zu-1.15 unterschieden werden (Huntley-Fenner, 2001).

In anderen Studien (Xu & Spelke, 2000; Xu, 2003) konnte gezeigt werden, dass das Unterscheiden von Mengen nicht nur mit dem Verhältnis der Mengen zueinander, sondern auch mit der Größe der Differenz (Distanz) der zu vergleichenden Anzahlen zusammenhängt. Je größer diese Distanz ist, desto präziser können Mengen bereits im frühen Kindesalter hinsichtlich ihrer Mächtigkeit diskriminiert werden. „Diesem Distanzeffekt zur Folge wird angenommen, dass Mengen ihrer Größe entsprechend auf einem »mentalen Zahlenstrahl« analog im Gehirn repräsentiert sind“ (van Eimeren & Ansari, 2009, S. 26).

In den 70er Jahren fanden Chi und Klahr (1975) heraus, dass zum Erkennen einer Menge von ein bis vier Elementen immer die gleiche Zeit benötigt wird und dass sich diese Reaktionszeit bei Fünfjährigen

nicht von der benötigten Zeit erwachsener Probanden unterscheidet. Wird die Anzahl allerdings weiter erhöht, nimmt mit steigender Anzahl auch die Reaktionszeit deutlich zu. Dies deutet darauf hin, dass die Erfassung kleiner Anzahlen – ohne Notwendigkeit des Abzählens – vom Menschen simultan erfasst werden kann (*subitizing*). In ihren Untersuchungen konnten Starkey und Cooper (1995) zeigen, dass Zweijährige zwei bis drei Elemente, 3;5-Jährige ein bis vier Elemente und Vier- bis Fünfjährige ein bis fünf Elemente simultan erfassen können. Hierbei bleibt jedoch anzumerken, dass nicht hinreichend geklärt ist, ob fünf Elemente noch simultan erfasst werden können, oder ob nicht eher eine Anzahl von fünf Objekten quasisimultan, d. h. als eine Menge aus zwei simultan erfassbaren Teilmengen zusammengesetzt, erfasst wird.

Insgesamt lässt sich festhalten, dass Kinder im frühen Kindesalter, bzw. bereits bis zur Vollendung des ersten Lebensjahres über verschiedene Repräsentationssysteme von Mengen verfügen. Zum einen werden hinreichend kleine Mengen sehr präzise mental repräsentiert, zum anderen werden größere Anzahlen unpräzise, jedoch analog repräsentiert, wodurch Mengenschätzungen ermöglicht werden.

2.2 Entwicklung ersten Mengen- und Zahlenwissens im Vorschulalter

Resnick (1989, 1992) geht ebenfalls davon aus, dass Menschen im frühen Kindesalter über eine angeborene Sensitivität für kleine Quantitäten verfügen, welche grundlegend für sog. protoquantitative Schemata, also eine Art intuitives Mengenwissen ist. Dabei handelt es sich um Einsichten bezüglich des Vergleichs verschiedener Mengen bzw. Veränderungen von Mengen, welche zunächst jedoch ohne jeglichen numerischen Bezug vorhanden sind und sich parallel zum Prozess der Zählentwicklung ausbilden. Kinder im Alter bis zu

zwei Jahren können demnach Mengenvergleiche bzw. -veränderungen grob begrifflich, jedoch nicht mit einem exakten quantitativen Bezug beschreiben. Resnick (1989, 1992) führt drei verschiedene protoquantitative Schemata auf: das Vergleichsschema, das Zunahme-Abnahme-Schema und das Teile-Ganzes-Schema.

Kinder unter zwei Jahren können demnach Mengenunterscheidungen mit den Begriffen „mehr“, „weniger“ oder „gleich viel“ vornehmen (Vergleichsschema), jedoch geschieht dies noch ohne die präzise Unterscheidung diskreter Mengen (Resnick, 1989). Dies spiegelt sich ebenso in der oben genannten Studie von Feigenson, Carey und Hauser (2002) wider. So wählten die Kinder (bei hinreichend kleinen Anzahlen) immer die Dose mit den meisten Keksen und verfügen demnach scheinbar über ein Konzept von „mehr“ und „weniger“ ohne die Anzahl der Elemente der Mengen ermitteln zu müssen. „Im Simultanerfassungsbereich (eins bis vier) können sie die Anzahl der Elemente auf einen Blick erfassen und erwerben vermutlich zuerst in diesem Zahlenraum das Schema des Mengenvergleichs, das sie dann auf größere Zahlbereiche übertragen können“ (Weißhaupt & Peucker, 2009, S. 56f.).

Im Alter zwischen drei und vier Jahren können Kinder Mengenveränderungen mit den Ausdrücken „mehr als zuvor“, „weniger als zuvor“ oder „genauso viel wie zuvor“ beschreiben (Zunahme-Abnahme-Schema, Resnick, 1989). Werden zu einer Menge weitere Elemente hinzugetan, sind es mehr Elemente als zuvor und werden von einer Menge Elemente weggenommen, sind es weniger Elemente als zuvor. Nach Piaget und Szeminska (1975) geht damit das Wissen einher, dass, dass sich Mengen nur dann verändern, wenn Elemente hinzukommen oder weggenommen werden, nicht aber bei Veränderung der räumlichen Ausdehnung (*Zahlinvarianz*).

Weiterhin wird den Kindern ab etwa vier Jahren bewusst, dass Mengen in unterschiedlicher Weise in Teilmengen zerlegt, sowie auch wieder zu einer Gesamtmenge zusammengesetzt werden können, ohne dass dies etwas an der Gesamtmenge ändert (Teile-Ganzes-Schema, Resnick, 1989). Damit verbunden ist die Einsicht, dass wenn eine Teilmenge vergrößert wird, so vergrößert sich auch die Gesamtmenge, wird ein Teil des Ganzen kleiner, so wird auch die Gesamtmenge kleiner. Und weiterhin: Wird von einer Teilmenge etwas weggenommen und zur anderen Teilmenge hinzugetan, wird die Veränderung kompensiert, die Gesamtmenge bleibt gleich groß. Erst mit der Verfügbarkeit des Teile-Ganzes-Schema sind Kinder in der Lage, die Komplementarität von Addition und Subtraktion als Strategie zur Lösung von Rechenaufgaben anzuwenden (Resnick, 2000).

Neben diesen protoquantitativen Schemata beschreibt Resnick (1989) ein weiteres kognitives Schema: den *mentalen Zahlenstrahl*. Es handelt sich um die internale Vorstellung der Zahlenfolge als eine Kette von Elementen. Jedes Element dieser Kette, also jede Zahl, steht dabei für eine konkrete Anzahl. Ein solches inneres Bild setzt demnach ein Verständnis für die Kardinalität, also die durch sie repräsentierte Quantität, der Zahlen voraus. Diese geistige Repräsentation beinhaltet neben der Reihenfolge der Zahlen ebenso basale Informationen zu den Beziehungen zwischen ihnen. Nachfolgende Zahlen entsprechen dabei größeren Mengen, vorangegangene Zahlen entsprechen kleineren Anzahlen. Neben den oben bereits erwähnten Studien von Xu und Spelke (2000) sowie Xu (2003) liefern auch Huntley-Fenner und Cannon (2000) einen Nachweis eines solchen inneren Konzeptes. Sie präsentierten Kindern im Alter von fünf bis sieben Jahren für einen kurzen Augenblick verschiedene Mengenabbildungen von 5, 7, 9 und 11 Elementen. Anschließend bekamen die Kinder die Aufgabe, die gezeigten Anzahlen auf einem Zahlenstrahl abzutragen. Die auf dem

Zahlenstrahl eingetragenen Markierungen glichen in ihren Proportionen den vorgegebenen Anzahlen, was als Hinweis für eine innere Repräsentation des Zahlenstrahls betrachtet werden kann.

Zusammenfassend lässt sich festhalten, dass Kinder schon vor Eintritt in die Schule auf kognitive Schemata zum Vergleichen und zur Veränderung von Mengen zurückgreifen können. Auch diese Schemata sind für kleine Mengen schon früh sehr präzise. Ebenso entwickelt sich recht früh ein Gespür für das Zerlegen von Mengen in Teilmengen sowie für das Zusammenlegen von Teilmengen zu einer Gesamtmenge. Dies sind wesentliche Einsichten für ein Verständnis der Rechenoperationen der Addition und Subtraktion. Durch eine systematische geistige Ordnung der Zahlen auf einem mentalen Zahlenstrahl wird ein flexibler Umgang mit den Zahlen möglich.

2.3 Voraussetzungen und Entwicklung der Zählfähigkeiten

Seit den 70er und 80er Jahren wurde neben der Entwicklung erster Kompetenzen im Umgang mit Mengen auch die Entwicklung des Zählprozesses erforscht (Gelman & Gallistel, 1978; Fuson, 1988). Als zentrale Frage dabei sollte ermittelt werden, ob Zählkompetenzen angeboren sind oder sie erst im Laufe der Entwicklung erlernt werden müssen.

Gelman und Gallistel (1978) vertreten die Annahme, dass das Zählwissen angeboren ist. Laut den Autoren sind folgende fünf Prinzipien als Voraussetzung des korrekten Zählens notwendig und von Geburt an verfügbar (Gelman, 2000; Gelman & Gallistel, 1978):

1. das Prinzip der Eins-zu-eins-Zuordnung (one-one principle),
2. das Prinzip der stabilen Reihenfolge (stable-order principle),

3. das Kardinalitätsprinzip (cardinal principle),
4. das Abstraktionsprinzip (abstraction principle) und
5. das Prinzip der irrelevanten Reihenfolge (order-irrelevance principle).

Dabei beziehen sich die ersten drei Prinzipien darauf *wie* gezählt wird (*how to*), die letzten beiden eher darauf *was* gezählt wird (*permissions*). Nach dem Prinzip der Eins-zu-eins-Zuordnung wird jedem zu zählenden Element genau eine Zahl zugeordnet. Verschiedene Objekte erhalten verschiedene Zahlzuordnungen. Dabei ändert sich die Ordnung der Zahlen nicht, ihre Reihenfolge bleibt erhalten (Prinzip der stabilen Reihenfolge). Die im Zählvorgang als letzte vergebene Zahl gibt die Mächtigkeit der gezählten Menge an (Kardinalitätsprinzip). Unter Berücksichtigung dieser ersten drei Prinzipien ist es unerheblich, was gezählt werden soll (Abstraktionsprinzip) sowie auch in welcher Reihenfolge die Elemente gezählt werden (Prinzip der irrelevanten Reihenfolge) (Gelman, 2000).

Diese Prinzipien werden von Gelman und Gallistel (1978) als angeboren angenommen, sind jedoch nicht sofort zugänglich. Vielmehr müssen sie in verschiedenen Kontexten ausgebaut werden. Gestützt sehen Gelman und Gallistel (1978) ihre postulierten Prinzipien durch Studien, welche zeigen, dass junge Kinder zwar ab und zu Zahlworte beim Aufsagen der Zahlwortreihe auslassen, jedoch keine Zahlworte doppelt nennen. Außerdem können Kinder in Experimenten Verstöße gegen die beschrieben Zählprinzipien bei anderen Personen aufdecken, selbst dann, wenn die eigenen Zählkompetenzen nicht vollends ausgeprägt sind (Gelman & Meck, 1983, 1986; Gelman, Meck & Merkin, 1986).

Die Annahme angeborener Zählprinzipien von Gelman und Gallistel (1978) ist in der Literatur kritisch reflektiert worden (Krajewski, 2003;

Moser Opitz, 2001). Vor allem der Auffassung eines angeborenen Kardinalkonzepts wurde an mehreren Stellen widersprochen (Baroody, 1992; Fuson, 1988; Sophian, 1998).

Entgegen der Ansicht von Gelman und Gallistel (1978) geht Fuson (1988) davon aus, dass Zählkompetenzen im Laufe der Entwicklung erlernt werden müssen. Animiert von Erwachsenen und älteren Kindern imitieren Kinder ab einem Alter von ca. zwei Jahren die ihnen vorgeführten Zählaktivitäten. Durch wiederholtes Anwenden in verschiedenen Situationen und durch Integration des vorhandenen Wissens über Mengen durchlaufen Kinder verschiedene Qualitätsstufen des Zählens. Fuson beschreibt fünf verschiedene Stufen der Entwicklung der Zählkompetenzen:

1. string level,
2. unbreakable list level,
3. breakable chain level,
4. numerable chain level,
5. bidirectional chain level.

Mit der Entwicklung der sprachlichen Fähigkeiten fangen Kinder, ungefähr ab einem Alter von zwei Jahren an, Zahlworte zu sprechen. Die Zählschritte werden zu diesem Zeitpunkt jedoch noch nicht mit diskreten Elementen verknüpft. Es erfolgt keine Eins-zu-Eins-Zuordnung von Zählzahl zu Objekten, sodass die Zahlen nicht als getrennte Elemente der Zahlwortreihe wahrgenommen werden. Vielmehr geschieht die Reihung der Zahlworte automatisiert als ein Auswendiglernen, stets beginnend von Eins. Fuson (1988) bezeichnet dies als "undifferenziertes Wortganzes" (*string level*). Das bedeutet, dass Kinder auf dieser Stufe das Zählen imitieren, ohne ein Verständnis für die Struktur der Zahlenfolge verinnerlicht zu haben. Diese Aneinanderreihung von Zahlwörtern kann demnach nicht zum Nach- bzw. Abzählen eingesetzt werden. Laut Fuson (1988) handelt

es sich hier um eine recht kurze Phase, welche nicht bei jedem Kind zu beobachten ist.

In der Zeit zwischen zwei und vier Jahren wird den Kindern bewusst, dass die Zahlenfolge eine Reihung diskreter Elemente darstellt. Die Zahlen sind die Glieder einer Kette, welche jedoch zu diesem Zeitpunkt als „unzerbrechlich“ (*unbreakable list*) wahrgenommen wird und entsprechend nur vollständig, beginnend ab der Startzahl eins wiedergegeben werden kann. Durch eine eindeutige Zuordnung der Zahlworte zu Objekten wird ein erstes Abzählen möglich. Dies geschieht in den meisten Fällen noch durch ein Antippen der zu zählenden Elemente. Mittels Zusammenzählen zweier Mengen beginnend mit eins („Counting-all-Strategy“, Fuson, 1988, S. 406) sind einfache Additionsaufgaben lösbar, wobei die letzte genannte Zahl die Mächtigkeit der gezählten Mengen angibt. Das heißt, dass das Ende der Zählkette als Kardinalwert bewusst wird.

Ab einem Alter von etwa vier Jahren verstehen Kinder, dass die Zahlen in ihrer Folge auch unterbrochen, von jeder beliebigen Startzahl aus aufgesagt werden können. Dies ist eng verbunden mit dem ersten Verständnis für die Kardinalität von Zahlen. Dies ist Grundlage des Verständnisses, dass mit der jeweils gewählten Startzahl des Zählens bereits eine konkrete Teilmenge, nämlich die Mächtigkeit dieser Startzahl, inkludiert ist. Durch ein solches Verständnis der Zahlenfolge sind Kinder auf dieser Stufe befähigt, Vorgänger und Nachfolger von Zahlen zu bestimmen sowie die Zahlen zu benennen, die zwischen zwei Zahlen liegen. Im Laufe der Zeit lernen Kinder, das Wissen über den Zählprozess auf das Rückwärtszählen zu übertragen. Fuson (1988) bezeichnet diese Stufe des Verständnisses der Zahlen als „aufgebrochene Kette“ als *breakable chain level*. Additionsaufgaben können auf dieser Stufe mittels Weiterzählen („Counting-on-Strategy“, Fuson, 1988, S. 407) gelöst werden, d. h. die Mächtigkeit der ersten Menge wird als

Startzahl aufgefasst, von welcher beginnend die Zahlenfolge weitergesagt wird.

Zwischen etwa vier und sieben Jahren wird die Zahlenreihe auf der folgenden Stufe als numerische Kette (*numerable chain*) verstanden, d. h. eine Zahl repräsentiert nicht nur die Mächtigkeit einer Menge von Elementen, sondern spiegelt ebenso eine konkrete Anzahl an Zählschritten wider, das bedeutet, dass Kinder auf dieser Stufe der Zählentwicklung auch von einer beliebigen Startzahl aus um eine vorgegebene Anzahl an Schritten weiterzählen können. Dieses Wissen kann auch auf ganze Abschnitte der Zahlenfolge verallgemeinert werden. Die Drei beschreibt demnach nicht nur den Abschnitt eins-zwei-drei in der Zahlwortreihe, sondern genauso auch den Abschnitt zwei-drei-vier oder fünf-sechs-sieben. Auf dieser Stufe ist der Zählprozess demnach nicht mehr nur noch an konkret abzählbare Objekte gebunden, sondern die Zahlen selbst werden als einzelne numerische Einheiten begriffen (Fuson, 1988).

Die Zählkompetenzen entsprechend der fünften Stufe, die Vorwärts-Rückwärts-Kette bzw. *bidirectional chain* bezeichnet Fuson als „truly numerical counting“ (1988, S. 412), denn auf dieser Entwicklungsstufe ist die Zählkompetenz vollständig entwickelt. In der Regel erreichen Kinder diese Kompetenzen erst nach dem Eintritt in die Schule. Sie haben dann ihre Zählkompetenzen soweit ausgebaut, dass ihnen das Vorwärts- und Rückwärtszählen leicht und flexibel gelingt. Richtungswechsel erfolgen dabei ohne Probleme. Den Kindern wird in diesem Zusammenhang die Beziehung zwischen Vorwärts- und Rückwärtszählen und die damit verbundene Umkehrbarkeit von Addition und Subtraktion deutlich. Zählen wird von den Kindern auf dieser Stufe als eine, aber nicht einzige Rechenstrategie erkannt. Dies ist mit der Einsicht verbunden, dass drei Zahlen wie bspw. 3, 5 und 8 als numerische Situation mit konkreter Beziehung zueinander (3 + 5 = 8, 5 + 3 = 8, 8 – 5 = 3 sowie

8 – 3 = 5) aufgefasst werden können. Dadurch werden flexible Operationen mit den Zahlen möglich, so können z. B. durch Ausnutzung der Umkehrbarkeit der Addition und Subtraktion und Anwendung der entsprechenden Zählprozeduren Ergänzungs- bzw. Platzhalteraufgaben gelöst werden (Fuson, 1988).

In den von Fuson (1988) beschriebenen Stufen der Entwicklung des Zählens zeigen sich deutliche qualitative Unterschiede. Während auf den ersten drei Stufen der Zählentwicklung nur das unmittelbar Wahrnehmbare, also konkrete Objekte, gezählt werden können, gelingt es den Kindern in den darauf folgenden Stufen, sich auf eine abstrakte Ebene zu beziehen. Sie verstehen dann, dass die Zählschritte an sich zählbare Einheiten darstellen können. Weiterhin werden bei Zählprozessen der unteren Stufen die Zahlen noch nicht mit den zugehörigen Mengenentsprechungen in Verbindung gebracht, so wird den Kindern erst ab der dritten Stufe die Kardinalität von Zahlen bewusst. Mithilfe dieser Einsicht werden Grundlagen für ein tiefes Verständnis der Struktur der Zahlen gelegt sowie erste Ideen für effektive Rechenstrategien entwickelt. Das von Fuson beschriebene Modell der Zählentwicklung findet auch in aktuell veröffentlichten theoretischen Modellvorstellungen, wie z. B. bei Moser Opitz (2001), Krajewski (2003) sowie Fritz et al. (2007) Berücksichtigung. Zwei in der Fachliteratur derzeit etablierte Entwicklungsmodelle erster mathematischer Kompetenzen, die auf den Erkenntnissen Fusons basieren und für die vorliegende Arbeit grundlegend sind, – die Modelle von Krajewski und Schneider (2006) sowie Fritz et al. (2007) – sollen in den nächsten Abschnitten differenziert erläutert werden.

2.4 Entwicklungsmodell früher mathematischer Kompetenzen nach Krajewski und Schneider (2006)

Folgt man den aktuellen Ansätzen der Entwicklungspsychologie in Deutschland entstehen mathematische Kompetenzen in hierarchisch aufeinander aufbauenden Entwicklungsstufen, die wiederum mit zentralen mathematischen Einsichten und Konzepten verbunden sind. Zeitlich parallel haben 2006 bzw. 2007 voneinander unabhängige Forschergruppen diese Stufen in Modellen zur Entwicklung mathematischer Kompetenzen bis ins frühe Grundschulalter zusammengefasst: das „Entwicklungsmodell früher mathematischer Kompetenzen“ nach Krajewski und Schneider (2006) sowie das „Modell der mathematischen Kompetenzentwicklung“ nach Fritz et al. (2007). Zentralerer Aspekt beider Entwicklungsmodelle ist die Integration der sich parallel zueinander entwickelnden protoquantitativen Vorstellungen (Resnick, 1989) und Zählfähigkeiten (Fuson, 1988). Dennoch gibt es einige Unterschiede, welche in den folgenden Abschnitten herausgearbeitet werden sollen.

Wie aus Abbildung 3 ersichtlich, bildet das „Entwicklungsmodell früher mathematischer Kompetenzen“ (Krajewski & Schneider, 2006) den hierarchischen Aufbau erster mathematischer Fähigkeiten von Kindern auf drei Ebenen ab.

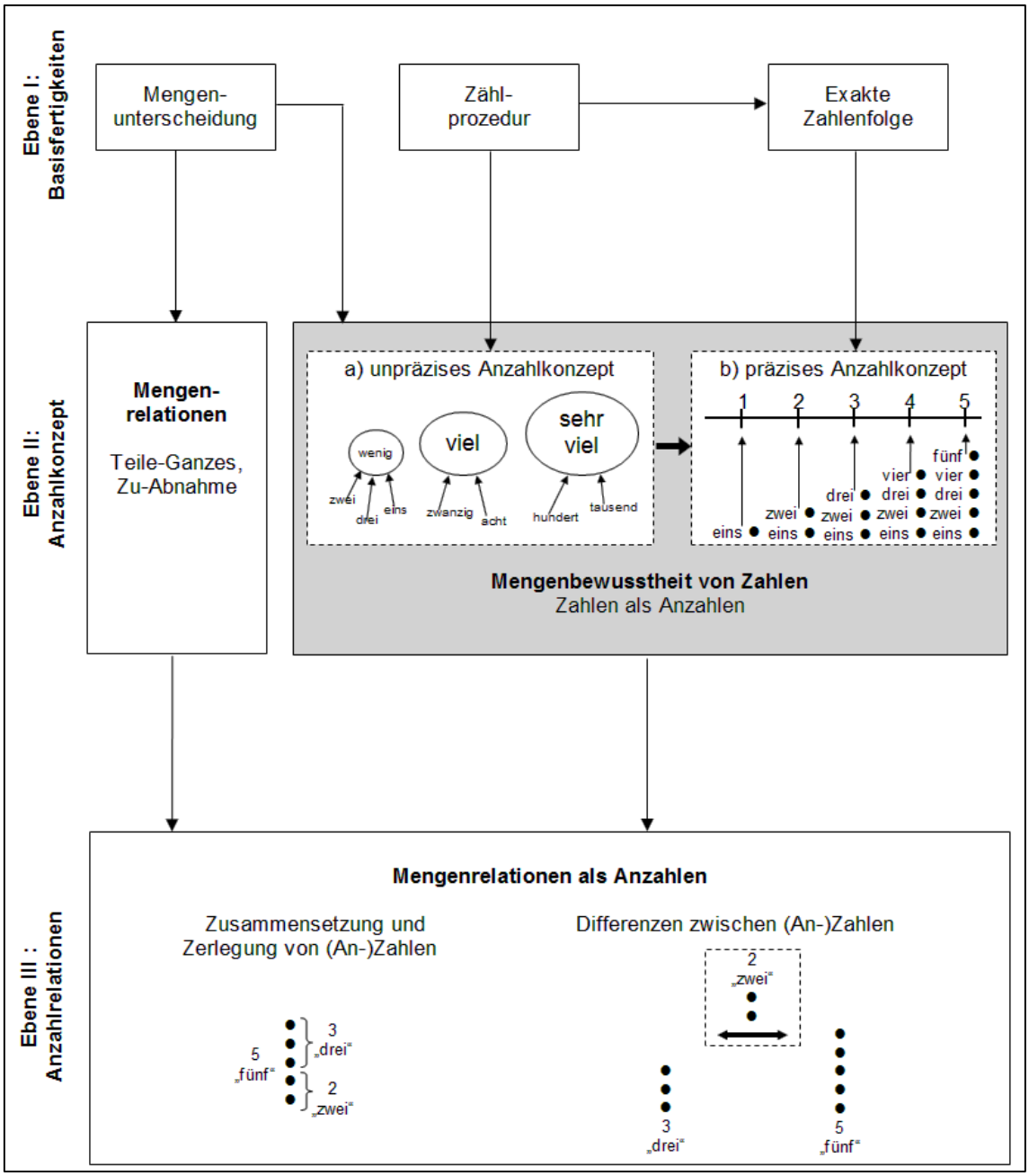


Abbildung 3: „Entwicklungsmodell früher mathematischer Kompetenzen“ nach Krajewski und Schneider (2006) (entnommen aus: Krajewski & Schneider, 2006, S. 250)

Auf Ebene I beschreiben Krajewski und Schneider (2006) wie sich mathematische *Basisfertigkeiten* entwickeln. Ausgangspunkt für Entwicklungen auf dieser Stufe stellen die in Abschnitt 2.1 aufgeführten Studien zur Bewusstheit von Mengen im frühen Kindesalter dar (Feigenson, Carey & Hauser, 2002; Starkey & Cooper, 1980; Wynn, 1992). Schon von Geburt an besitzen Kinder

die Fähigkeit, kleine Mengen zu diskriminieren. Diese Unterscheidung wird jedoch nicht auf Grundlage der unterschiedlichen Mächtigkeit der zu vergleichenden Mengen getroffen (diskret), sondern anhand der von ihnen eingenommenen Flächen (kontinuierlich) (Clearfield & Mix, 2001; Feigenson, Carey & Spelke, 2002). Zunächst gelingt es den Kindern dabei, Unterscheidungen von Mengen vorzunehmen, wenn deren Differenz hinreichend groß ist. Da sich auf dieser Ebene das Verständnis von Mengen und das Verständnis von Zahlen noch unabhängig voneinander entwickeln, werden diese Unterschiede jedoch nicht numerisch belegt, sondern lediglich mit den Begriffen „mehr“ oder „weniger“ ausgedrückt (Vergleichsschema, Resnick, 1989). Ab etwa zwei Jahren fangen Kinder an, die Zahlwortreihe aufzusagen, ohne jedoch einen Zusammenhang zu den dahinter verborgenen Mengen zu sehen. Die Zahlwortreihe wird zunächst ähnlich einem Gedicht aufgesagt, ohne dass die Kinder bewusst realisieren, dass die Reihe aus einzelnen Elementen mit fester Reihenfolge besteht (string level, Fuson, 1988). Die stabile Reihenfolge der Zahlen wird durch spielerisches Üben im Laufe der Zeit bewusst (exakte Zahlenfolge) (Krajewski & Schneider, 2006).

Auf der Ebene II, im Alter von drei bis fünf Jahren, bilden Kinder ein Verständnis für unbestimmte Mengen aus. Sie gelangen zu der Einsicht, dass Mengen in Teilmengen zerlegt werden können. Im Umkehrschluss können Teilmengen auch zu einer Gesamtmenge zusammengeschlossen werden. Diese Einsicht bezeichnen Krajewski und Schneider (2006) analog zu Resnick (1989) als Teile-Ganzes-Prinzip. Ebenso wird bewusst, dass Mengen kleiner werden, sobald man ihnen Elemente entnimmt bzw. größer werden, sobald man ihnen Elemente hinzufügt (Zunahme-Abnahme-Schema, Resnick, 1989). Die Kinder begreifen die Relationen zwischen den Mengen, können diese jedoch zunächst nicht im numerischen Bezug sehen. Parallel zur Entwicklung dieser Mengeneinsichten erwerben

Kinder ab dem Kindergartenalter eine Mengenbewusstheit von Zahlen (Anzahlkonzept). Dieses Anzahlkonzept ist das Resultat der Integration der beiden Teilfertigkeiten des Zählens und des Mengenunterscheidens. Kinder verstehen demnach, dass Zahlen als Repräsentanten zur Beschreibung der Mächtigkeit von Mengen dienen, also Anzahlen darstellen (Kardinalitätsprinzip, Gelman & Gallistel, 1978). Krajewski und Schneider (2006) gehen davon aus, dass sich dieser Prozess in zwei Phasen gliedert. Zunächst, ab ca. drei Jahren, besitzen Kinder unpräzise Vorstellungen von Mengen, die sie zuerst nur in grobe Kategorien wie „wenig“ (z. B. eins, zwei, drei), „viel“ (z. B. zehn, zwanzig) und „sehr viel“ (z. B. hundert, tausend) einordnen können (Ebene IIa: *unpräzises Anzahlkonzept* nach Krajewski & Scheider, 2006). Mit dieser Einsicht gelingen den Kindern nun erste Zahlenvergleiche. Dies ist allerdings nur möglich bei Zahlen, die verschiedenen Mengenkategorien, wie bspw. „wenig“, „viel“ oder „sehr viel“ angehören. Vergleiche von Zahlen innerhalb einer Kategorie sind für Kinder auf der Ebene II noch nicht lösbar. Im Laufe der Zeit präzisiert sich dieses Anzahlwissen der Kinder jedoch. Mit etwa vier Jahren können die Zahlen dann exakt der entsprechend repräsentierten Quantität zugeordnet werden. Die Zahlenfolge wird als Reihung aufsteigender Anzahlen begriffen, welche mit jedem Schritt um eins erhöht werden und somit ihren festen Platz in dieser Reihe besitzen (Ebene IIb: *präzises Anzahlkonzept* nach Krajewski & Schneider, 2006). Mit dieser Einsicht ist es den Kindern möglich, nun alle Zahlen, auch innerhalb dieser groben Kategorien, exakt miteinander zu vergleichen. Ebenso können Vorgänger- und Nachfolger-Zahlen bestimmt werden (Krajewski & Schneider, 2006).

Der größte Meilenstein der Entwicklung mathematischer Kompetenzen liegt schließlich auf Ebene III. Hier können Kinder ihr Mengen- und Zahlwissen vollständig ineinander integrieren. Sowohl das Zu-Abnahme- als auch das Teile-Ganzes-Prinzip präzisieren sich und Mengenänderungen bzw. -unterschiede können nun numerisch

belegt werden. D. h. nicht nur Mengen können zerlegt und zusammengefügt werden, sondern auch Zahlen. Weiterhin gelingt es den Kindern, nicht nur zu verstehen, dass 5 (fünf Elemente) sich aus 2 (zwei Elemente) und 3 (drei Elemente) zusammensetzt, sondern auch, dass die Differenz bzw. der Unterschied zwischen zwei Zahlen durch eine Zahl bestimmt ist: „fünf ist zwei mehr als drei“. Dieses relationale Verständnis von Zahlen bezeichnen Krajewski und Schneider (2006) als *Anzahlrelationen* bzw. *Relationszahlkonzept*. Stern (1998) beschreibt diesen relationalen Aspekt von Zahlen als „eine der unmittelbaren Wahrnehmung nicht zugängliche Beziehung zwischen zwei Mengen“ (S. 77). Im Gegensatz zum kardinalen Verständnis einer Zahl kann eine Zahl im relationalen Sinn nicht einfach durch Abzählen ermittelt werden, sondern es verlangt ein tiefes konzeptuelles Verständnis der Beziehungen zwischen Zahlen (Krajewski & Schneider, 2006). Mit den Erkenntnissen der Ebene III des Entwicklungsmodells früher mathematischer Kompetenzen sind das Hoch- und Herunterzählen von einem Summanden möglich, d. h., dass Kinder nun nicht nur in der Lage sind, erste Rechenoperationen durchzuführen, sie können diesen Zusammenhang darüber hinaus auch numerisch belegen. Erkenntnisse der Ebenen I und II können demnach als mathematische Vorläuferfähigkeiten angesehen werden, Einsichten der Ebene III hingegen spiegeln erstes arithmetisches Wissen wider (Krajewski & Schneider, 2006).

In zahlreichen Studien konnte ein Zusammenhang zwischen den von Krajewski und Schneider (2006) konstatierten ersten mathematischen Kompetenzen mit späteren schulischen Mathematikleistungen von Kindern nachgewiesen werden. Dadurch sind erste empirische Hinweise für die Bedeutung des Stufenmodells gegeben (Aunola et al., 2004; Gaupp et al., 2004; Geary et al., 2000; Krajewski & Schneider, 2006; Landerl, Bevan & Butterworth, 2004; Moser Opitz, 2007; Ostad, 1998; Stern, 1997). So führte die Studie

von Aunola et al. (2004) zu dem Ergebnis, dass die mathematischen Leistungen finnischer Grundschüler durch die im Vorschulalter erhobenen Zählfähigkeiten vorhergesagt werden können. Weiterhin konnte ein Zusammenhang von der Fähigkeit zum Vergleichen von Zahlen, der Positionierung von Zahlen auf dem Zahlenstrahl und einem Verständnis von Teile-Ganzes-Beziehungen mit den Rechenkompetenzen der Schüler belegt werden (Gaupp et al., 2004; Moser Opitz, 2007). Studien zufolge (Gaupp et al., 2004; Geary et al., 2000; Landerl et al., 2004; Moser Opitz, 2007; Ostad, 1998) zeigen insbesondere rechenschwache Kinder große Rückstände in ersten mathematischen Einsichten und Konzeptvorstellungen entsprechend der Ebenen I bis III des Entwicklungsmodells nach Krajewski und Schneider (2006).

2.5 Modell der mathematischen Kompetenzentwicklung nach Fritz, Ricken und Gerlach (2007)

Das Modell der mathematischen Kompetenzentwicklung nach Fritz et al. (2007) beschreibt die Entwicklung erster mathematischer Kompetenzen, nicht in drei wie Krajewski und Schneider (2006), sondern in fünf hierarchisch aufeinander aufbauenden Stufen (Abbildung 4). Es wird versucht, im Entwicklungsfluss zentrale Meilensteine auszumachen, welche wesentliche Einsichten für den Erwerb des Verständnisses von Zahlen, Mengen und ihren Beziehungen zueinander beschreiben. Diese Meilensteine bzw. Grundkompetenzen erlauben wiederum das Erkennen verschiedener Rechenstrategien, die das Lösen von Aufgaben im Bereich der Arithmetik begünstigen bzw. überhaupt erst ermöglichen. Nachfolgend sollen die einzelnen Stufen des Modells detaillierter beschrieben werden.

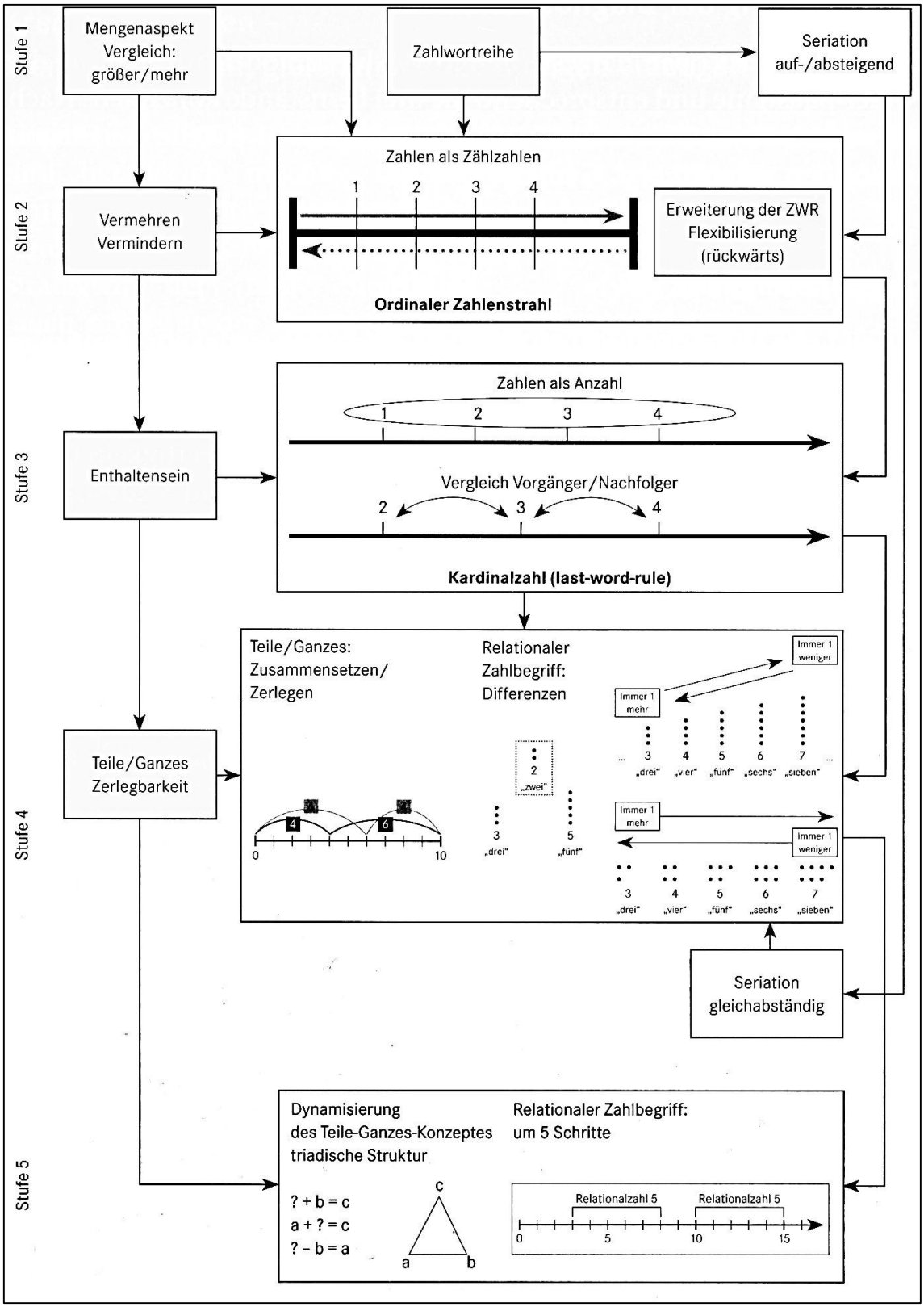


Abbildung 4: Modell der mathematischen Kompetenzentwicklung nach Fritz et al. (2007) (entnommen aus: Fritz et al., 2007, S. 15)

Das Modell von Fritz et al. (2007) ist ähnlich wie das Modell von Krajewski und Schneider (2006) ausgehend von theoretischen Konzeptvorstellungen nach Resnick (1989) und Fuson (1988) erarbeitet worden und wird durch die Forschungen hinsichtlich des Mengenverständnisses im Säuglingsalter (Feigenson, Carey & Hauser, 2002; Feigenson, Carey & Spelke, 2002; Starkey & Cooper, 1980; Wynn, 1992) gestützt.

Auf Stufe 1 entwickeln sich das Wissen um Mengen und das Wissen um Zahlen parallel und unabhängig voneinander. Durch die angeborene Sensibilität für kleine Mengen (Resnick, 1989) und den frühen Kontakt mit Mengen und Zahlen im Spiel, durch bspw. Abzählreime und -lieder, beginnen Kinder sich eine immer weiter ausdifferenzierende Wissensbasis hinsichtlich Mengen und Zahlen anzueignen. Mit dem Beginn des Sprechens erlernen Kinder die Zahlwortreihe, zunächst jedoch ohne kardinalen Bezug der Zahlen. Die Zahlen werden hier als bloße Reihung von Wörtern verstanden, noch nicht als einzelne Elemente in der Zahlenfolge mit fester, geregelter Ordnung. Entsprechend wird die Zahlwortreihe zu diesem Zeitpunkt noch nicht zum Zählen von Objekten genutzt, sondern eher wie ein zusammenhängendes Wort aufgesagt. Dies entspricht dem von Fuson (1988) geschilderten *string-level* der Entwicklung des Zählprozesses und wird auch in dem bereits beschriebenen Modell (Ebene I) von Krajewski und Schneider (2006) berücksichtigt. Darüber hinaus können Kinder auf Stufe 1 Mengen zwar grob vergleichen, allerdings nur im Sinne von „größer“ bzw. „kleiner“ oder „mehr“ bzw. „weniger“, d. h. ohne numerischen Bezug (Fritz et al., 2007). Die Unterscheidung hängt nicht von der Anzahl der Elemente, sondern von der Größe der Ausdehnung bzw. des Umfangs der Mengen ab. Voraussetzung für eine solche Diskrimination ist, dass der Unterschied zwischen den Mengen hinreichend groß ist (vgl. Vergleichsschema, Resnick, 1989), andernfalls kann es passieren, dass Mengen mit unterschiedlichen Anzahlen an Elementen als

gleich groß identifiziert werden (bspw. 6 und 8). Eine weitere Möglichkeit der Mengendiskrimination erfolgt über eine Eins-zu-Eins-Zuordnung, d. h. eine Menge wird dann als größer als die Vergleichsmenge eingeschätzt, wenn sie nach einem Eins-zu-Eins-Vergleich der jeweils enthaltenen Elemente noch Objekte aufweist (Fritz et al., 2007).

Auf Stufe 2, im Alter von ca. drei bis vier Jahren, verfügen Kinder über die Einsicht, dass Zahlworte bestimmte Elemente in der Zahlenfolge repräsentieren (unbreakable list level, Fuson, 1988). Sie verstehen den Vorgang des Zählens, bei dem jedem Objekt genau ein Zahlwort zugeordnet wird (Prinzip der Eins-zu-eins-Zuordnung, Gelman & Gallistel, 1978). Zusammen mit der Fähigkeit zur Mengenunterscheidung entsteht eine mentale Vorstellung vom Aufbau und der Abfolge der Zahlen (mentaler Zahlenstrahl, Resnick, 1989). Zwar werden die Zahlen auf dieser Stufe grundlegend nach ihrer Position auf dem Zahlenstrahl, also rein ordinal, betrachtet, die Kinder verstehen jedoch bereits, dass jede Zahl einen Nachfolger hat, der größer ist, d. h. der Zahlenstrahl wird als Abfolge immer größer werdender Zahlen verstanden. Auf Grundlage dieser Einsichten werden Vergleiche zwischen Zahlen nach dem Prinzip „Je länger ich zählen muss, umso größer ist die Zahl.“ möglich. Wird in dieses Zahlenwissen das Prinzip der Zu- bzw. Abnahme von Mengen (Resnick, 1989) integriert, werden erste Additionsaufgaben durch zählendes Rechnen lösbar. Dazu werden alle Elemente der ersten Menge von eins an, meist mit den Fingern als Hilfsmittel, ausgezählt und analog dazu die Elemente der zweiten Menge. Anschließend werden die Elemente beider Mengen, wieder von eins an, zusammengezählt. Dieses scheinbare sichere Rechnen ist trügerisch, da es ein Verständnis für Kardinalität von Zahlen widerspiegelt, über welches Kinder auf dieser Stufe jedoch noch nicht verfügen (Fritz et al., 2007).

Zahlen nicht nur ordinal, sondern auch kardinal zu erfassen, gelingt Kindern erst ab etwa vier Jahren auf Stufe 3 des Entwicklungsmodells von Fritz et al. (2007). Ob Kinder diese Einsicht besitzen, zeigt sich, wenn sie, gefragt nach der Anzahl der von ihnen zuvor gezählten Objekte, das zuletzt genannte Zahlwort als Mächtigkeit der abgezählten Menge verstanden haben (last-word-rule, Fuson, 1988). Mit dem Wissen über die Kardinalität von Zahlen müssen Kinder beim Zusammenzählen zweier bzw. mehrerer Mengen nicht mehr immer bei eins anfangen, sondern können vom ersten Summanden an weiterzählen (counting-on-strategy, Fuson, 1988). Nach Fritz et al. (2007) können Kinder das Verständnis für die Kardinalität von Zahlen nur erlangen, wenn sie das Konzept des Enthaltenseins verinnerlicht haben. Ihnen ist demnach bewusst, dass Zahlen andere Zahlen enthalten. So enthält die Zahl vier alle ihr vorangehenden Zahlen eins, zwei und drei. Diese Einsicht entwickelt sich auf Grundlage des Prinzips der Klasseninklusion nach Piaget (1964). Ricken, Fritz und Balzer (2011) beschreiben darüber hinaus:

> *„Die Zahlwortreihe wird als Sequenz größer werdender kardinaler Einheiten verstanden. Das bedeutet, Größenvergleiche zwischen zwei Zahlen finden nun nicht mehr auf der Ebene des Rangplatzes in der Zahlwortreihe statt (die 5 ist größer als die 4, da sie später in der Zahlwortreihe auftaucht), sondern auf der Ebene des Vergleichs der Mächtigkeit zweier Mengen (die 5 ist größer als die 4, da sie mehr Elemente enthält)“* (S. 260).

Auf Stufe 4 wird das Verständnis um das Teile-Ganzes-Konzept bezüglich Mengen (Resnick, 1989) erkannt, ausgebaut und um einen numerischen Bezug ergänzt (Fritz et al., 2007). Das bedeutet, dass aus zwei Teilmengen eine Gesamtmenge gebildet werden kann und das Zusammenzählen somit als additive Handlung bewusst wird. Andersherum können Zahlen auch in Teile zerlegt werden. „Folgende Aufgabe wird nun lösbar: ‚Gib mir 5 Bauklötze, 3 *davon* [Herv. d. Verf.] sollen rot sein‘“ (Fritz et al., 2007, S. 11). Durch die Integration der

Vorstellungen zum Zahlenstrahl und dem Verständnis von Quantitäten, wird Kindern die Zahlenfolge als Reihung von Elementen bewusst, welche schrittweise, zunehmend jeweils um eine Einheit, größer werden. Darüber hinaus können Kinder auf Grundlage dieser Einsicht den Unterschied zwischen zwei Mengen beziffern. Das bedeutet, dass sich ein relationales Verständnis der Zahlen zueinander anbahnt. Eine Zahl gibt demnach nicht nur eine Position innerhalb der Zahlenfolge oder die Mächtigkeit der mit ihr verbundenen Menge an, sondern kann darüber hinaus einen beliebigen Abschnitt des Zahlenstrahls beschreiben (Fritz et al., 2007). So steht die Drei für den Abschnitt eins-zwei-drei, kann jedoch auch den Abschnitt fünf-sechs-sieben bzw. zwölf-dreizehn-vierzehn usw. repräsentieren (numerable chain level, Resnick, 1989).

Auf der Stufe 5 des Entwicklungsmodells nach Fritz et al. (2007) wird das zuvor erworbene Wissen zur Relation zwischen den Zahlen konkretisiert und flexibilisiert. Kinder haben verstanden, dass Zahlen in Teilmengen zerlegt werden können und sind in der Lage, dies flexibel ausführen. So kann etwa die Sieben in die Vier und die Drei zerlegt werden, aber auch in fünf und zwei oder sechs und eins. Das Wissen über die Beziehungen von Zahlen zueinander wird ebenso weiter ausdifferenziert und somit flexibel nutzbar. Auf dieser Stufe erkennen Kinder die Zusammensetzung von Aufgaben zu Zahlentripeln. Bei dem Tripel 2 - 4 - 6 ist ihnen klar, dass 6 das Ganze repräsentiert und die Zahlen 2 und 4 für die Teile des Ganzen stehen. Sie erkennen, dass die Aufgaben

2 + 4 = _ 4 + 2 = _ _ + 4 = 6 _ + 2 = 6 4 + _ = 6 2 + _ = 6
sowie
6 – 4 = _ 6 – 2 = _ 6 – _ = 2 6 – _ = 4 _ – 4 = 2 _ – 2 = 4

zusammenhängen. Das Teile-Ganzes-Schema wird demnach in das relationale Zahlverständnis integriert. Dadurch sind Aufgaben lösbar

wie z. B.: „Auf der Rutsche sind 5 Kinder. Auf dem Klettergerüst sind 4 Kinder mehr als auf der Rutsche. Wie viele Kinder sind auf dem Klettergerüst?“ (Fritz et al., 2007, S. 13). Aber ebenso Aufgaben wie: „Auf der Rutsche sind einige Kinder. Es kommen noch 4 Kinder dazu. Nun sind 7 Kinder auf der Rutsche. Wie viele Kinder waren zu Beginn da?“ (Fritz et al., 2007, S. 13) oder „Auf dem Klettergerüst sind 7 Kinder weniger als auf der Rutsche Auf beiden Spielgeräten sind zusammen 19 Kinder. Wie viele Kinder sind auf dem Klettergerüst“ (Fritz et al., 2007, S. 14) werden für die Kinder lösbar. Ebenso werden mit den Einsichten der Stufe 5 verschiedene effektive Rechenstrategien erkannt und nutzbar. Bspw. können Aufgaben zerlegt und in einfacheren Teilschritten gelöst werden: Die Aufgabe 5 + 7 kann umstrukturiert werden zu 5 + (5 + 2) = 10 + 2 = 12. Die Kommutation der Addition kann ausgenutzt werden, um schwere Aufgaben wie 5 + 9 umzukehren zu 9 + 5 (Fritz et al., 2007).

Im Jahr 2009 publizierten Fritz und Ricken eine modifizierte Version des Modells der mathematischen Kompetenzentwicklung. In dieser Fassung wird die erste Stufe des Modells in zwei Stufen (Entwicklungsstufe 0 und 1) ausdifferenziert. Die Stufe 0 gilt als Vorstufe und umfasst konzeptuelles Wissen hinsichtlich der Kenntnis erster Zahlen sowie hinsichtlich des unpräzisen Unterscheidens von Mengen. Erst wenn die Zahlen als Zählzahlen aufgefasst werden, befindet sich ein Kind auf Stufe 1 des Entwicklungsmodells. Zudem wird die Reihenfolge der Konzeptentwicklung etwas abgeändert. So schreiben Fritz und Ricken (2009), entgegen der Ausführungen bei Fritz et al. (2007), dass das kognitive Konzept des Enthaltenseins erst auf der Niveaustufe 4, somit nach Einsicht in das Teile-Ganzes-Konzept (dort als Bestandteil der Entwicklungsstufe 3), verinnerlicht wird. Offen bleibt in der 2009 veröffentlichten, modifizierten Version, warum Änderungen vorgenommen wurden.

Das ursprüngliche (Fritz et al., 2007) sowie das modifizierte Modell (Fritz & Ricken, 2009) berücksichtigen, dass die mathematischen Kompetenzen von Kindern etwa hinsichtlich verschiedener Zahlenräume gleichzeitig auf verschiedenen Niveaus eingeordnet werden können. So kann ein Kind sich bspw. bezüglich einstelliger Zahlen schon auf dem Niveau der Stufe 4 bewegen, bezüglich zweistelliger Zahlen jedoch erst über Einsichten unterer Niveaus verfügen.

Nicht eindeutig geklärt ist die Altersspanne, in der sich die Fähigkeiten entsprechend Stufe 5 entwickeln. Untersuchungen von Resnick (1989) und Stern (1994, 1998) zeigen, dass dieses Entwicklungsniveau bei zahlreichen Kindern im Alter von acht oder neun Jahren noch nicht voll ausgebildet ist und die Entwicklung der einzelnen Fähigkeiten sich somit weit in den Grundschulbereich hineinzieht.

Die im Kontext mit dem Entwicklungsmodell von Krajewski und Schneider (2006) aufgeführten Studien zum Zusammenhang zwischen ersten mathematischen Kompetenzen und späteren Rechenleistungen in der Schule (Aunola et al., 2004; Gaupp et al., 2004; Geary et al., 2000; Krajewski & Schneider, 2006; Landerl et al., 2004; Moser Opitz, 2007; Ostad, 1998; Stern, 1997), können auch als Hinweise für die Gültigkeit des Modells von Fritz et al. (2007) gewertet werden. Ebenso bestätigen die Ergebnisse der Studien zu Mathematikleistungen rechenschwacher Schüler die Wichtigkeit der an die fünf Entwicklungsniveaus geknüpften Konzeptvorstellungen (Gaupp et al., 2004; Geary et al., 2000; Landerl et al., 2004; Moser Opitz, 2007; Ostad, 1998). Darüber hinaus konnten, im Zusammenhang mit der Entwicklung eines Diagnoseverfahrens zur Erfassung von mathematischen Konzepten, welches sich an den Niveaustufen des Modells von Fritz, Ricken und Gerlach (2007) orientiert (MARKO-D, Ricken et al., im Druck), Operationalisierungen

dieser fünf Niveaustufen an einer Stichprobe mit ca. 3000 Kindern erprobt werden. Grundlegend, so Ricken, Fritz und Balzer (2011) „konnte das Modell in seinem Aufbau repliziert werden“ (S. 261).

2.6 Das Konzept des number sense

Analog zum Konzept der phonologischen Bewusstheit als zentrale Voraussetzung zum erfolgreichen Leseerwerb wird angenommen, dass es eine Bewusstheit für Zahlen und Mengen gibt, welches ebenso eine tragende Bedeutung bei der Entwicklung mathematischer Kompetenzen spielt: das Konzept des number sense (Berch, 2005; Gersten & Chard, 1999; Reys, 1991). Obwohl das Konzept des number sense schon vor mehreren Jahrzehnten vor allem in der US-amerikanischen Literatur auftaucht, gibt es bis dato noch keine einheitliche Konzeptualisierung (Clarke & Shinn, 2004). Aufgrund dessen soll in dem vorliegenden Abschnitt auf das Thema number sense nur überblickartig eingegangen werden. Vieles deckt sich zudem mit den in den Abschnitten zuvor erläuterten Konzept- und Modellvorstellungen.

Berch (2005) stellt die mit dem number sense beschriebene Bewusstheit für Zahlen bzw. den Zahlensinn im Wesentlichen als Konstrukt dar, welches sich auf die Geschwindigkeit und Flexibilität eines Kindes bezieht, Zahlen zu verstehen und mit ihnen umzugehen. Trafton und Thiessen (1999) definieren number sense etwas konkreter als “a cluster of ideas, such as the meaning of a number, ways of representing numbers, relationships among numbers, the relative magnitude of numbers, and skill in working with them" (S. 50). Forschungsergebnisse führen folgende Aspekte an, die auf einen breit ausgeprägten Zahlensinn hinweisen:

- Ein ausgeprägtes Verständnis für Zahlen sowie deren Beziehungen zueinander (z. B. Clarke & Shinn, 2004; National Council of Teachers of Mathematics, 1989),
- die Fähigkeit zur Zahlzerlegung und -zusammensetzung (z. B. Fischer, 1990; Ross, 1989),
- die Fähigkeit zur Unterscheidung von Zahlen aufgrund ihrer Menge (z. B. National Council of Teachers of Mathematics, 1989; Sowder, 1988),
- das Verständnis für Operationen und deren Effekt auf Zahlen (z. B. National Council of Teachers of Mathematics, 1989; Sowder, 1992),
- die Entwicklung von Rechenstrategien (z. B. Sowder, 1992),
- die Fähigkeit zum Kopfrechnen (z. B. Hope & Sherrill, 1987; Sowder, 1991; Trafton, 1992),
- die Fähigkeit zum Schätzen (z. B. Bobis, 1991; Case & Sowder, 1990),
- das Verständnis des Stellenwertsystems (z. B. Fischer, 1990; Ross, 1989) sowie
- die Fähigkeit zur Lösung mathematischer Probleme (z. B. Cobb, Wood, Yackel, Nicholls, Wheatley, Trigatti & Perlwitz, 1991).

Griffin, Case und Siegler (1994) vermuten, dass es sich bei der Aneignung eines Zahlensinns eher um einen informellen Prozess handelt, dessen Beginn vor dem Eintritt in die Schule liegt. Vorstellungen dazu, wie sich erste Einsichten zu Zahlen und Mengen bei einem Kind entwickeln, wurden bereits in den Abschnitten 2.1, 2.2 sowie 2.3 beschrieben.

Eine ausgeprägte Bewusstheit für Zahlen im Sinne des number sense gilt nach Gersten und Chard (1999) als zentrale Voraussetzung zur Lösung arithmetischer Problemstellungen und somit auch zur Automatisierung von mathematischen Faktenwissens. Diese

Annahme wird durch verschiedene empirisch angelegte Studien bestärkt, welche Hinweise auf einen engen Zusammenhang zwischen einem eher gering ausgeprägten Zahlensinn und Rechenschwierigkeiten geben (McCloskey & Macaruso, 1995; McIntosh, B.J. Reys & R.E. Reys, 1992).

2.7 Kognitionspsychologische und neurowissenschaftliche Aspekte der Entwicklung von Rechenkompetenzen

In den bisherigen Ausführungen wurde der Erwerb mathematischer Kompetenzen in erster Linie aus entwicklungspsychologischer Sicht betrachtet. Eine weitere Domäne, die sich mit den Rechenfähigkeiten von Menschen befasst, ist die Kognitionspsychologie. Innerhalb dieses Forschungsgebietes wird der zentralen Frage nachgegangen, wie Zahlen geistig repräsentiert und verarbeitet werden (u. a. Ashcraft, 1992; von Aster, 2005; Dehaene, 1992). Besondere Beachtung finden dabei sog. Netzwerkmodelle (Ashcraft, 1992). Darin werden Wissensnetzwerke beschrieben, in denen grundlegende arithmetische Fakten organisiert sind. Rechenaufgaben können nach diesen Modellen durch Abruf aus der Wissensbasis des Kindes gelöst werden (Ashcraft, 1992). Neben den Modellen zum Wissensabruf arithmetischer Fakten aus mentalen Netzwerken sind eine Reihe neuropsychologischer Ansätze beschrieben worden, welche die geistige Repräsentation von Zahlen und deren Verarbeitung erklären sollen (McCloskey, Caramazza & Basili, 1985; Campbell, 1992; Dehaene, 1992). Das nachfolgende Kapitel stellt das *network retrieval model* von Ashcraft (1992) vor. Anschließend wird das in der neuropsychologischen Forschung der kognitiven Zahlenverarbeitung bedeutsame *triple-code-model* nach Dehaene (1992) beschrieben.

Untersuchungen der 70er Jahre belegten, dass sich die Zeit zum Lösen von Rechenaufgaben proportional zu der Größe der Faktoren bzw. Summanden erhöht (Groen & Parkman, 1972; Ashcraft & Battaglia, 1978). Während anfangs vermutet wurde, dass sich die Lösung von Additionsaufgaben durch Weiterzählen vom größeren Summanden aus vollzieht (*min-counting-strategy*) und die Lösungsgeschwindigkeit somit von der Größe des kleineren Summanden abhängt (Parkman & Groen, 1971), zeigten spätere Studien, dass die ansteigende Lösungszeit mit größer werdenden Summanden durch Faktenabruf aus dem Gedächtnis zu erklären ist (Ashcraft, 1992). Demnach sind Ergebnisse einfacherer Rechenaufgaben leichter aus dem Wissensnetz arithmetischer Fakten, also aus dem Gedächtnis, abrufbar, da sie durch Übungssituationen häufiger aktiviert wurden und somit besser vernetzt sind. Aufgaben größerer Zahlenräume hingegen, werden seltener geübt, der Wissensabruf erfolgt dementsprechend langsamer (*network retrieval model* nach Ashcraft, 1992).

Studien von Carpenter und Moser (1984) zeigen, dass sich der Faktenabruf aus dem Wissensnetz zum Lösen von Rechenaufgaben bei Grundschülern ausgehend von verschiedenen zählbasierten Strategien (counting-all-strategy, counting-on-strategy, min-strategy) entwickelt. Dieser Entwicklungsverlauf folgt nicht zwangsläufig als eindeutige Abfolge von Stufen (Siegler & Jenkins, 1989), wie Siegler (1987) in seiner Studie zur Ermittlung der Rechenstrategien an Kindern im Kindergartenalter (N = 22), der ersten Klasse (N = 28) und der zweiten Klasse (N = 18) zeigte. In der Untersuchung wurden allen teilnehmenden Schülern 45 mathematische Aufgaben gestellt (dieselben in jeder Altersstufe) mit dem Hinweis ihre Lösungswege zu verbalisieren. Siegler fand heraus, dass die untersuchten Schüler oftmals mehrere Strategien zeitgleich nutzten. So überlappt der Einsatz früher genutzter Strategien mit dem Anwenden effizienterer, neu erworbener Strategien (*overlapping waves*). Der Autor fand

zudem heraus, dass durchschnittlich 35 % der Kinder (Kindergarten: 16 %; Klasse eins: 44 %; Klasse zwei: 45 %) die Strategie des Faktenabrufs aus dem Gedächtnis nutzten, durchschnittlich 36 % (Kindergarten: 30 %; Klasse eins: 38 %; Klasse zwei: 40 %) lösten die Rechenaufgaben durch Weiterzählen vom größeren Summanden an, 7 % (Kindergarten: 2 %; Klasse eins: 9 %; Klasse zwei: 11 %) nutzten Zahlzerlegungsstrategien und 8 % der Kinder (Kindergarten: 22 %; Klasse eins: 1 %; Klasse zwei: 0 %) lösten die Aufgaben mittels der ineffizienten count-all-Strategie (Siegler, 1987, S. 256). Dabei stellte sich der Faktenabruf als die schnellste und mit am wenigsten Fehlern belastete Strategie (neben dem Einsatz von Zerlegungsstrategien) heraus (Siegler, 1987, S. 256).

Dehaene (1992) betrachtet die Zahlenrepräsentation und -verarbeitung aus einer neuropsychologischen Sichtweise. Das von ihm entwickelte triple-code-model, dargestellt in Abbildung 5, beschreibt die Repräsentation numerischen Wissens im Gehirn. Er geht davon aus, dass Zahlen in drei eigenständigen kognitiven Modulen verarbeitet werden und dort jeweils in verschiedenen Codierungen repräsentiert sind: *analog magnitude representation*, *visual arabic number form* und *auditory verbal word frame* (Dehaene, 1992, S. 31).

Im Modul *analog magnitude representation* (Analoge Repräsentation von Größen, Übersetzung nach Jacobs & Petermann, 2005, S. 20) sind Zahlen als ungefährer Ort auf einem mentalen, räumlich strukturierten Zahlenstrahl visualisiert. Dabei können kleinere Zahlen präziser erfasst werden als größere. Diese Tatsache lässt sich durch das Weber-Fechner-Gesetz auf zweierlei Art erklären (Dehaene, 1992):

(1)„the number line is linear, and variability increases with numerical magnitude;

(2) the number line is a compressive logarithmic or power function, and variability is constant“ (S. 23).

Auf Grundlage dieser analogen Form der Repräsentation können kleine Mengen bis vier präzise simultan erfasst werden. Größere Mengen hingegen werden näherungsweise geschätzt. Darüber hinaus erlaubt diese Funktionseinheit approximative Rechenleistungen (Dehaene, 1992).

Im Modul *visual arabic number form* (Visuell-arabische Repräsentation, Übersetzung nach Jacobs & Petermann, 2005, S. 20) werden Zahlen symbolisch in entsprechenden arabischen Ziffern kodiert, welche durch Lesen und Schreiben aufgenommen bzw. wiedergegeben werden. Diese Kodierung erfolgt nach Regeln des dekadischen Zahlsystems auf Grundlage des Stellenwertprinzips. Daneben sind Zahleneigenschaften wie die Parität (also eine Zuordnung in gerade und ungerade Zahlen) und Lösungsalgorithmen zur Bearbeitung von Additions- und Subtraktionsaufgaben auf symbolischer (schriftlicher) Ebene gespeichert (Dehaene, 1992).

Das Modul *auditory verbal word frame* (Auditiv-sprachliche Repräsentation, Übersetzung nach Jacobs & Petermann, 2005, S. 20) bezieht sich auf die sprachliche Repräsentation der Zahlen in Zahlworte, welche durch Hören und Lesen aufgenommen und durch Sprechen und Schreiben wiedergegeben werden können. In dieser Einheit werden Zählprozesse und der Abruf von arithmetischem Faktenwissen gesteuert (Dehaene, 1992).

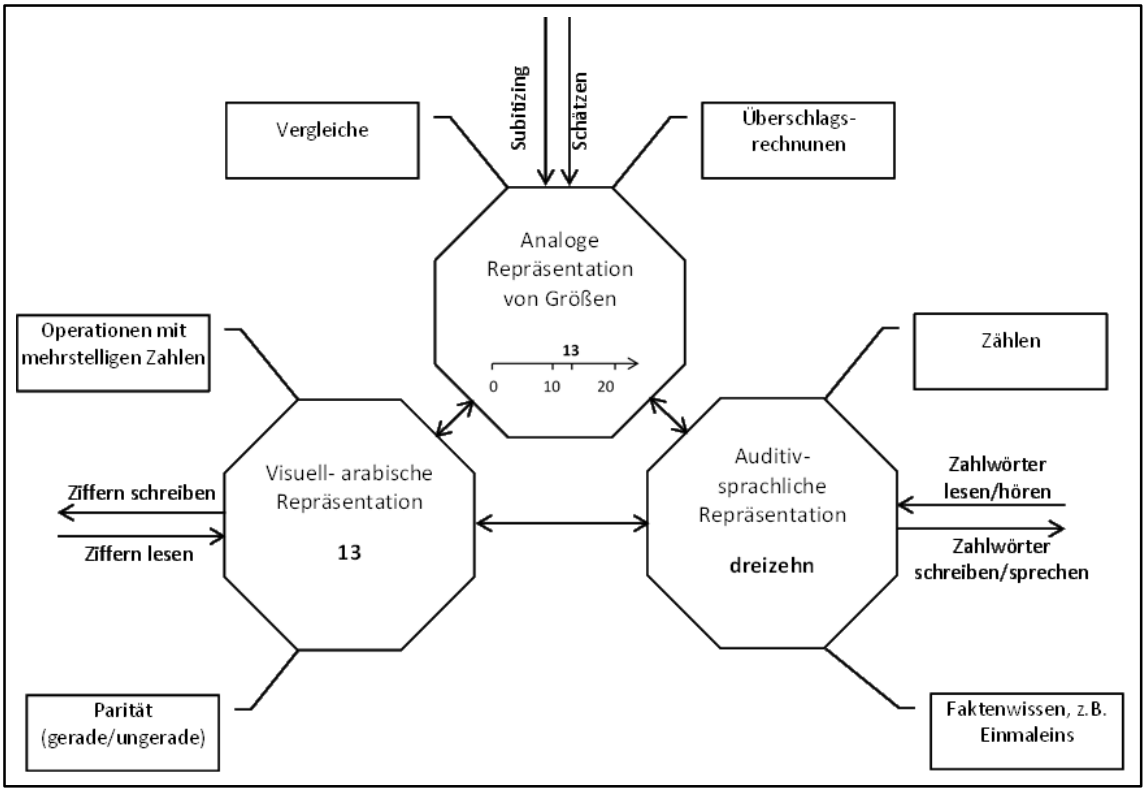


Abbildung 5: Triple-code-model nach Dehaene (übersetzt und vereinfacht nach Dehaene, 1992, S. 31)

Obgleich diese Module eigenständige Funktionseinheiten darstellen und in unterschiedlichen Hirnarealen lokalisiert sind, gibt es doch kognitive Verknüpfungen zwischen ihnen. Zahlen können, durch Transkodierprozesse im Gehirn, in die jeweils andere Repräsentationsformen übersetzt werden. Die Zahl 13 (visuell-arabische Repräsentation) wird dabei mit dem Zahlwort „dreizehn“ (auditiv-sprachliche Repräsentation) in Beziehung gesetzt, kann aber auch gleichzeitig als eine (unscharfe) Menge auf einem inneren Zahlenstrahl (analoge Größenrepräsentation) lokalisiert werden. D. h., dass beim Rechnen mehrere Übersetzungsprozesse durchlaufen werden. Die Rechenleistungen eines Menschen hängen demnach maßgeblich von der Transkodierleistung zwischen diesen drei Repräsentationsebenen ab. Der Zahlbegriff eines Schülers kann basierend auf dem Modell Dehaenes (1992) als das Ergebnis der Arbeit einzelner Funktionsmodule im Gehirn verstanden werden,

welche im Laufe der Entwicklung aufeinander abgestimmt werden müssen, um ein gesichertes Zahlenverständnis zu ermöglichen.

Dass die Zahlenverarbeitung in den unterschiedlichen Modulen bereichsspezifische Besonderheiten aufweist, schlussfolgert Dehaene (1992) auf Basis verschiedener Studien, in denen Personen mit aphasischer Akalkulie (durch erworbene Hirnschädigung verursachte Schwierigkeiten im Umgang mit mathematischen Kontexten) untersucht wurden (Dehaene & Cohen, 1991; Guttman, 1937; Warrington, 1982). Diese Personen zeigten Ausfälle in je einem der drei Module. So beschreibt Warrington (1982) einen Patienten, dem es nicht gelang, arithmetische Fakten zur Addition, Subtraktion oder Multiplikation aus dem Gedächtnis abzurufen (Auditiv-sprachliche Repräsentation), jedoch Rechenaufgaben schätzend lösen konnte (Analoge Repräsentation von Größen). Guttman (1937) berichtet von einem umgekehrten Fall, in welchem die Person Schwierigkeiten beim Schätzen zeigte (Analoge Repräsentation von Größen), jedoch durch Faktenabruf des Einmaleins einfache Multiplikationsaufgaben lösen konnte (Auditiv-sprachliche Repräsentation). Dehaene und Cohen (1991) beschreiben einen Patienten, welcher ausschließlich Zugriff auf seine analoge Repräsentation von Größen hatte und Störungen in den beiden anderen Modulen aufwies. Ob und inwieweit diese Einzelfallstudien an Erwachsenen mit Hirnschädigungen eine Übertragung auf Rechenprozesse bei Personen ohne Schädigung des Gehirns bzw. bei Kindern ermöglichen, untersuchten Schmithorst & Brown (2004). Bei ihren Untersuchungen wurden 15 erwachsene Versuchspersonen aufgefordert, Additions- und Subtraktionsaufgaben mit Brüchen im Kopf zu lösen. Mittels funktioneller Magnetresonanztomographie wurde ermittelt, ob sich separate aufgabenbezogene Hirnareale ausmachen lassen. Schmithorst und Brown (2004) kommen zu folgendem Schluss:

> *„The results support the previously hypothesized triplecode model of analog magnitude, auditory verbal, and visual Arabic frames of number representation and processing, and indicate that the triple-code model may be a suitable neurocognitive framework for analyzing the performance of complex mathematical tasks“* (S. 1418).

Studien zur Überprüfung des Modells an Kindern stehen derzeit noch aus. Ebenfalls ist nicht eindeutig beschrieben, welche spezifischen Prozesse zwischen den einzelnen Modulen ablaufen, dadurch sind Aussagen über die Entwicklung des Zahlbegriffs bzw. von Rechenleistungen erschwert. Dennoch ist in dem triple-code-model nach Dehaene (1992) zur Erklärung von Rechenstörungen ein bedeutsamer theoretischer Ansatz zu sehen.

Aufbauend auf dem triple-code-model nach Dehaene (1992) beschreibt von Aster (2005) ein Vier-Stufen-Modell, welches „eine hierarchisch gegliederte Entwicklung der verschiedenen kognitiven Repräsentationen für Zahlen“ (von Aster et al., 2007, S. 85) postuliert. Neben spezifischen numerischen Funktionen werden in dem Modell die wachsende Kapazität und Verfügbarkeit unspezifischer, übergreifender Funktionen wie Aufmerksamkeit und Arbeitsgedächtnis berücksichtigt (von Aster, 2005). Auf der ersten Stufe des Modells sind angeborene Kompetenzen zur Erfassung hinreichend kleiner und der Unterscheidung hinreichend verschiedener Mengen verortet (Mengensinn) (Feigenson, Carey & Hauser, 2002; Feigenson, Carey & Spelke, 2002; Xu & Arriaga, 2007). Kennzeichnend für die zweite Stufe ist das Erlernen der Zahlwortreihe entsprechend der Prinzipien nach Gelman und Gallistel (1978) sowie der protoquantitativen Schemata des Vergleichens und der Veränderung nach Resnick (1989) durch den sozialen Austausch im Vorschulalter. Das Erlernen der Zahlen gemäß dem arabischen Notationssystem ist Inhalt der dritten Stufe (von Aster, 2005). Dies geschieht mit Eintritt in die Schule. „Die Kinder erlernen und

automatisieren damit auch kulturspezifische Übersetzungsregeln für das Übertragen eines gesprochenen oder geschriebenen Zahlwortes in die entsprechende arabische Symbolik und umgekehrt. In der deutschen Zahlwortsequenz ergibt sich hier insbesondere die Schwierigkeit der Zehner-Einer-Inversion (einundzwanzig - 21)“ (von Aster, Kucian & Martin, 2006, S. 3). Auf Grundlage des Wissens um das arabische Zahlensystem entwickelt sich nach von Aster (2005) auf der vierten Stufe, etwa in der zweiten bzw. dritten Klassenstufe, eine mentale Repräsentation eines Zahlenstrahls entsprechend der Ausführungen von Resnick (1989).

Zusammenfassend lassen sich verschiedene Bedingungen extrahieren, welche wesentliche Bedeutung für die Entwicklung arithmetischer Kompetenzen aus kognitions- bzw. neuropsychologischer Perspektive haben. Neben unspezifischen Komponenten wie Arbeitsgedächtnis- und Aufmerksamkeitsleistungen, spielen vor allem die spezifischen Komponenten wie eine analoge Repräsentation von Zahlen und Mengen auf einem mentalen Zahlenstrahl, ausgebildete Zählfähigkeiten, weitere Fähigkeiten der sprachlichen Verarbeitung von Zahlen, ein ausgeprägtes Wissensnetz arithmetischer Fakten sowie ein differenziertes Wissen über Zahlen, deren Strukturen und gültigen Rechenalgorithmen im arabischen Notationssystem eine zentrale Rolle (Ashcraft, 1992; von Aster, 2005; Dehaene, 1992). Diese spezifischen Kompetenzen sind in voneinander separaten Modulen kognitiv organisiert (Dehaene, 1992). Probleme beim Erwerb mathematischer Kompetenzen zu diagnostizieren, heißt somit diese Module – jedes für sich – auf Funktionalität hin zu überprüfen. Darüber hinaus muss sichergestellt werden, ob Transkodierleistungen von einer Repräsentationsform in die anderen möglich sind. Für Messverfahren zur Beurteilung mathematischer Leistungen in der ersten Klassenstufe sind vor allem die auditiv-sprachliche sowie die visuell-arabische Repräsentationsform

relevant. Ein solches Diagnostikum bedarf also Aufgaben zur Zahlenkenntnis (u. a. der Zahlwort-Ziffer-Korrespondenz) sowie zur Kenntnis und zum Verständnis des arabischen Notationssystems inklusive erster Rechenalgorithmen. Weiterhin sollten Zählkompetenzen sowie erworbene arithmetische Fakten geprüft werden. Obwohl der komplette Fähigkeitsumfang der analogen Repräsentation von Mengen und Zahlen auf einem mentalen Zahlenstrahl erst ab Klasse zwei bzw. 3 zu erwarten ist (von Aster, 2005), sollten dem zugrunde liegende Fähigkeiten wie das Kardinalitätsverständnis von Zahlen, die (Simultan-)Erfassung und die Diskrimination kleinerer Mengen im Zahlenraum bis 20 in einem diagnostischen Verfahren für Klasse eins operationalisiert sein.

2.8 Weitere allgemeine kognitive Faktoren und deren Zusammenhang mit der Entwicklung von Rechenkompetenzen

In den bereits beschriebenen entwicklungs- und kognitionspsychologisch orientierten Modellvorstellungen spielen bereichsspezifische Kompetenzen die zentrale Rolle beim Erlernen mathematischer Fähigkeiten. Es ist jedoch davon auszugehen, dass (schulische) Lernprozesse und so auch mathematisches Lernen durch weitere allgemeine kognitive Faktoren beeinflusst werden. Hasselhorn und Gold (2006) führen den Lernerfolg auf die Qualität der Informationsverarbeitung im Gehirn zurück. Lernen ist demnach als Prozess zu verstehen, bei dem externe Informationsangebote und intern vorhandene Wissensstrukturen miteinander verknüpft werden. An diesem Prozess sind kognitive Faktoren wie Wahrnehmung, Gedächtnis, Aufmerksamkeit und Intelligenz wesentlich beteiligt. Das Zusammenspiel dieser einzelnen kognitiven Faktoren soll im Folgenden durch das Modell zur Informationsverarbeitung im Gedächtnis nach Atkinson und Shiffrin (1968) exemplarisch

dargestellt und für den Bereich des mathematischen Lernens spezifiziert werden.

Abbildung 6 zeigt im Schritt 1, dass Informationen über visuell, haptisch und/oder akustisch dargebotene externe Reize vom Individuum über die Sinneskanäle wahrgenommen werden und ins sensorische Register gelangen. Die Informationen werden hier modalitätsbezogen gespeichert. Zwar kann das sensorische Register eine gewisse Menge Informationen bereithalten, jedoch nur einige Zehntelsekunden, daher die synonyme Bezeichnung Ultrakurzzeitgedächtnis. Mittels Fokussierung der Aufmerksamkeit werden relevante von irrelevanten Informationen getrennt (selektive Aufmerksamkeit). Die selektierten relevanten Informationen gelangen dann in das Kurzzeit- bzw. Arbeitsgedächtnis, wo sie in einem zweiten Schritt weiter verarbeitet werden. Die Gedächtnisspanne, also die Kapazität des Arbeitsgedächtnisses umfasst ca. sieben (± 2) Einheiten (Miller, 1956). Miller hält in diesem Zusammenhang fest: “By organizing the stimulus input simultaneously into several dimensions and successively into a sequence or chunks, we manage to break (or at least stretch) this informational bottleneck“ (1956, S. 84). Trotz solcher Strategien der Seriation oder Klassifikation von Informationen, ist die Kapazität des Arbeitsgedächtnisses begrenzt. In einem dritten Schritt werden die Informationen durch wiederholte Auseinandersetzung mit bereits vorhandenem Wissen im Langzeitgedächtnis verknüpft. Die Kapazität des Langzeitgedächtnisses ist sehr hoch, sodass sehr viele Informationen permanent gespeichert werden können. „Wenn also ein Schüler […] eine Rechenstrategie, wie das Verdoppeln und Halbieren oft nutzt, speichert er dabei Informationen im Langzeitgedächtnis ab. Zudem übt er gleichzeitig den schnellen Abruf von hierfür relevanten Informationen aus dem Gedächtnis“ (Hartke, 2010, S. 22).

Nach dem Modell von Atkinson und Shiffrin (1968) hängt der Prozess des Lernens wesentlich von altersgemäß entwickelten Wahrnehmungsfähigkeiten, von der Fähigkeit zur Fokussierung der Aufmerksamkeit sowie von der Kapazität des Arbeitsgedächtnisses ab. Speziell dem Arbeitsgedächtnis kommt in der aktuellen Lehr- und Lernforschung eine besondere Bedeutung zu (u. a. Büttner, 2003; Grube, 2005; Hasselhorn & Grube, 2003; Weinert & Helmke, 1997).

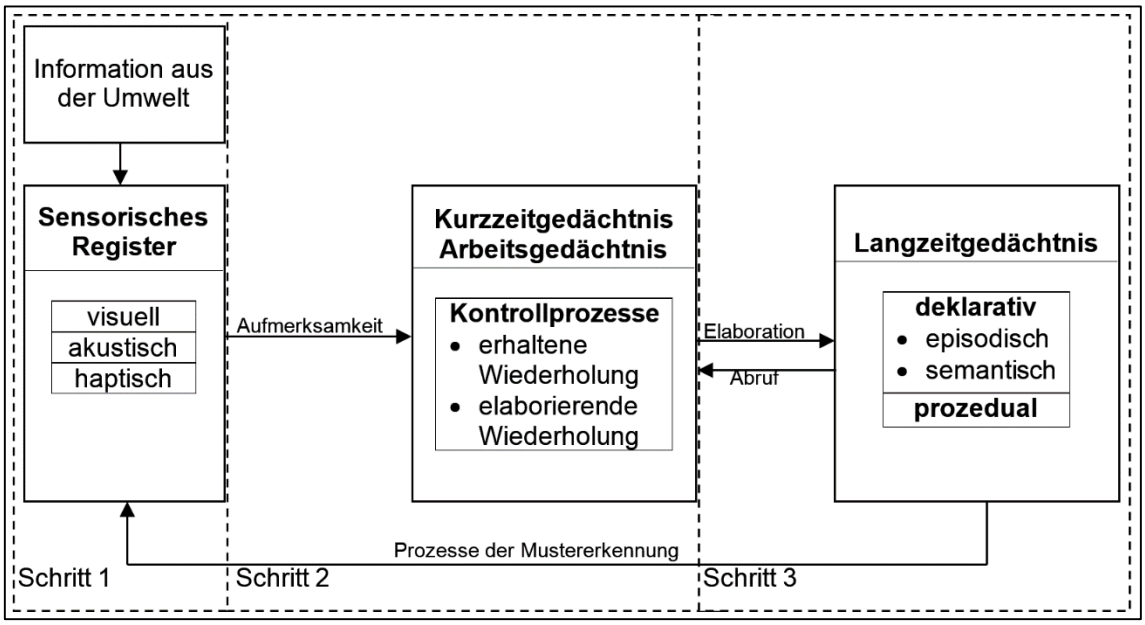


Abbildung 6: Verarbeitung einer Information gemäß dem Drei-Speicher-Modell (Atkinson & Shiffrin, 1968)

Die Informationsverarbeitung im Arbeitsgedächtnis wird im Modell des Arbeitsgedächtnisses von Baddeley (1986) genauer spezifiziert (Abbildung 7). Untersuchungen Baddeleys (1990) zeigten, dass zeitgleich verschiedene Typen von Aufgaben im Arbeitsgedächtnis verarbeitet werden können, z. B. das Merken einer Reihe von Worten während der Bearbeitung komplexer Rechenaufgaben. Darauf basiert die Annahme, dass das Arbeitsgedächtnis sich nicht als einfaches einheitliches System darstellt, sondern aus verschiedenen Komponenten besteht. Baddeley und Hitch (1974) unterscheiden zwei kapazitätsbegrenzte Subsysteme, welche diskret voneinander

sprachlich-akustische (in der *Phonologischen Schleife* bzw. *phonological loop*) und visuell-räumliche Informationen (im *Visuell-räumlichen Skizzenblock* bzw. *visual-spatial sketchpad*) verarbeiten. Als übergreifende Kontrolleinheit wird die *Zentrale Exekutive* (*central executive*) angesehen. Diese Komponente koordiniert und steuert aufmerksamkeitsabhängig alle bewusst ablaufenden kognitiven Prozesse. Im Jahr 2000 erweiterte Baddeley im Zuge einer Revision das Modell um den *Episodischen Puffer* (*episodic buffer*) als weitere Einheit. Der Episodische Puffer wird als eine Komponente mit begrenzter Speicherkapazität beschrieben, welche Informationen der beiden weiteren Subsysteme integriert und den Zugriff auf gespeicherte Inhalte aus dem Langzeitgedächtnis ermöglicht (Baddeley, 2000).

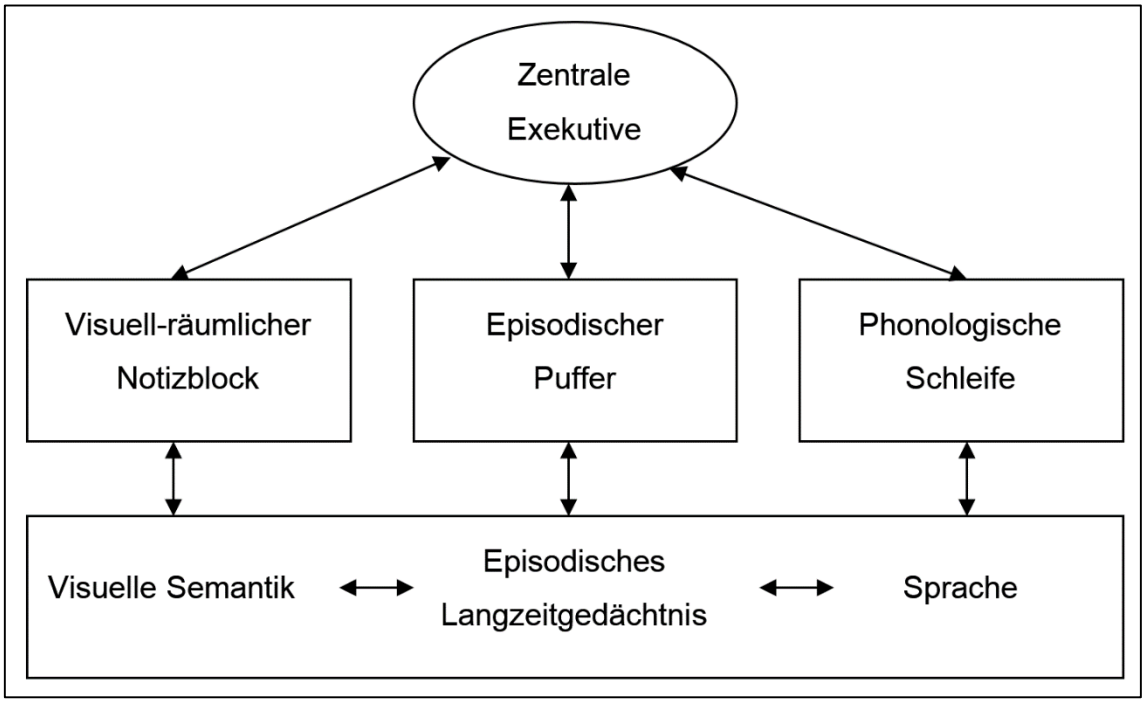


Abbildung 7: Arbeitsgedächtnismodell nach Baddeley (2000) (entnommen aus: Hasselhorn & Gold, 2009, S. 74)

Für die von Baddeley postulierte dreigliedrige Struktur des Arbeitsgedächtnisses fanden Schuchardt, Roick, Mähler und Hasselhorn (2008) innerhalb ihrer Untersuchung an 245 Grundschülern empirische Evidenzen.

Werden die Annahmen hinsichtlich des Einflusses der Gedächtnisleistungen auf das Lernen auf das Phänomen der Schwierigkeiten beim Erwerb von Rechenkompetenzen bezogen, lässt sich folgender Zusammenhang vermuten: Schüler mit Schwierigkeiten beim Rechnen nutzen meist zählende Strategie zur Berechnung von Aufgaben (Siegler, 1987), welche eine enorme Inanspruchnahme des Arbeitsgedächtnis bedeuten. Damit bleiben nur geringe bzw. keine Ressourcen frei, um übergeordnete Strategien wie bspw. den Faktenabruf aus dem Langzeitgedächtnis etablieren zu können, welche für die Flexibilisierung von Rechenkompetenzen unerlässlich sind. Tatsächlich konnte nachgewiesen werden, dass Kinder mit Rechenstörungen sowohl Beeinträchtigungen in der Zentralen Exekutiven (Gaupp, 2003; Geary et al., 2000; McLean & Hitch, 1999; Passolunghi & Siegel, 2001; Swanson & Sachse-Lee, 2001) als auch in der Phonologischen Schleife (Grube & Barth, 2004; Schuchardt, Kunze, Grube & Hasselhorn, 2006) und im Visuell-räumlichen Notizblock aufweisen (Gaupp, 2003; McLean & Hitch, 1999; Schuchardt et al., 2008). Diese Resultate sprechen dafür, mit schwachen Rechnern möglichst frühzeitig alternative Lösungsstrategien zu erarbeiten, um durch das zählende Rechnen blockierte Gedächtnisressourcen freizusetzen.

Auch wenn es in der gegenwärtigen pädagogisch-psychologischen Forschung sehr bedeutsam ist, gibt es berechtigte Einwände an dem Arbeitsgedächtnismodell Baddeleys. So erklärt Baddeley ausschließlich die Funktion der Subsysteme (Phonologische Schleife und Visuell-räumlicher Notizblock, später auch Episodischer Puffer). Die zwischen diesen Modulen ablaufenden Prozesse werden hingegen nicht beschrieben. Eines der in jüngerer Vergangenheit generierten Arbeitsgedächtnismodelle von Engle (2002) geht entgegen Baddeley und Hitch (1974) davon aus, dass nicht die Menge der Informationen, die eine Person speichern kann, sondern die Fokussierung der Aufmerksamkeit bzw. die Fähigkeit

Störvariablen auszublenden, die relevante Größe zur Beschreibung der Qualität der Arbeitsgedächtnisleistung ist. Der Aufmerksamkeit kommt die zentrale Aufgabe zu, arbeitsgedächtnisbeeinflussende, ablenkende bzw. interferierende Informationen zu kontrollieren. Gedächtniskapazität ist demzufolge kein Maß für die Speicherfähigkeit von Informationen gemäß Baddeley und Hitch (1974) bzw. Baddeley (2000), sondern ein Maß für die Fähigkeit, relevante von irrelevanten Informationen zu diskriminieren (Engle, 2002).

Resümierend kann festgehalten werden, dass sich Schwierigkeiten beim Lernen aus kognitionspsychologischer Perspektive als Resultat beeinträchtigter kognitiver Grundfunktionen ergeben können. Aufgrund von Schwierigkeiten in den Wahrnehmungsleistungen, Aufmerksamkeitsdefiziten bzw. Schwierigkeiten in der Aufmerksamkeitssteuerung, Überlastung des Arbeitsgedächtnisses bzw. dessen Einzelkomponenten oder aufgrund unzureichender Strukturierung und Vernetzung gespeicherten Wissens kann es zu Problemen beim Erlernen schulischer, speziell auch mathematischer Inhalte kommen. Die Förderung von Kindern mit Schwierigkeiten im Rechnen erfordert, neben einer differenzierten Analyse des bereits erworbenen mathematischen Vorwissens, ebenso eine Diagnostik kognitiver Einflussfaktoren, um Ursachen von Rechenproblemen zu lokalisieren. Generelles Ziel des Unterrichts und der Förderung sollte es sein, Kinder zu befähigen, nichtzählende Rechenstrategien zu erlernen und erfolgreich anzuwenden. Zentrale Rolle spielt dabei das wiederholte Üben (Automatisieren) grundlegender Rechenaufgaben. Folgt man den Modellen von Atkinson und Shiffrin (1968) sowie Ashcraft (1992), kann durch das Automatisieren von Grundrechenaufgaben ein breites Wissensnetz im Langzeitgedächtnis errichtet werden, aus dem Lösungen von Aufgaben abgerufen werden können. Wenn die Lösung vieler Grundrechenaufgaben bekannt ist, müssen diese nicht permanent

neu gelöst werden, sodass die Belastung des Arbeitsgedächtnisses reduziert werden kann und frei werdende Ressourcen für andere mathematische Kontexte genutzt werden können.

2.9 Zentrale Elemente des mathematischen Erstunterrichts in deutschen Lehrplänen

Die Ergebnisse internationaler Vergleichsstudien wie PISA, TIMSS und IGLU (vgl. Abschnitt 1.1) haben aufgezeigt, dass es zum Teil gravierende Unterschiede hinsichtlich der mathematischen Kompetenzen der Schüler zwischen den deutschen Bundesländern gibt. In diesem Zusammenhang wurden bundesweit geltende Bildungsstandards zur Entwicklung und Sicherung von Unterrichtsqualität beschlossen (Kultusministerkonferenz, 2004b). Darin werden u. a. konkrete Anforderungen an den Mathematikunterricht im Primarbereich spezifiziert. Dabei werden nicht primär konkret festgelegte Unterrichtsgegenstände und -inhalte aufgeführt, sondern es werden bestimmte Fähigkeiten, Fertigkeiten und Kompetenzen angeben, über welche die Kinder bis zum Ende der Grundschulzeit verfügen sollen (Kultusministerkonferenz, 2004b). Die Bildungsstandards für das Fach Mathematik in der Primarstufe gliedern die zu vermittelnden Kompetenzen in allgemeine (Problemlösen, Kommunizieren, Argumentieren, Modellieren und Darstellen) und inhaltsbezogene mathematische Kompetenzen (Zahlen und Operationen, Raum und Form, Muster und Strukturen, Größen und Messen sowie Daten, Häufigkeit und Wahrscheinlichkeit), welche in enger Verbindung gesehen werden müssen (vgl. Abbildung 8). Während inhaltsbezogene Kompetenzen auf das Aneignen von konkreten Faktenwissen und Fertigkeiten zur Anwendung von Lösungsalgorithmen dienen, werden mit den allgemeinen Kompetenzen eher übergreifende mathematische

Kompetenzen angesprochen, die das Problemlösen mathematischer Kontexte und Situationen begünstigen sollen.

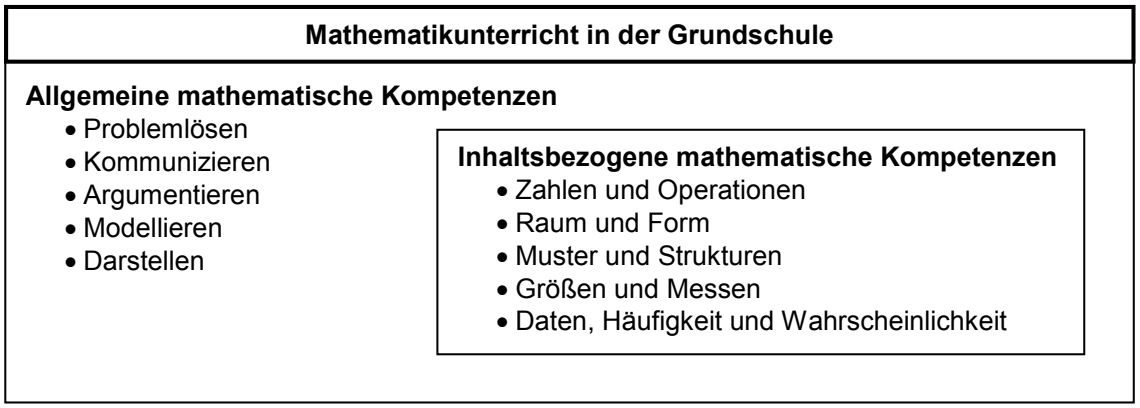


Abbildung 8: Allgemeine und inhaltsbezogene mathematischen Kompetenzen im Mathematikunterricht der Grundschule (Kultusministerkonferenz, 2004b, S. 7)

Neben diesen geltenden Bildungsstandards verfügt jedes deutsche Bundesland über Lehrpläne bzw. Curricula, welche die Lerninhalte des Mathematikunterrichts in der Grundschule beschreiben. Knopp (2010b) gibt einen komprimierten Überblick zu zentralen Elementen des mathematischen Erstunterrichts in Deutschland. Dazu fasst die Autorin die in den Bundesländern ausgeschriebenen Lehrplaninhalte für den Bereich der Arithmetik in den ersten Grundschulklassen sowie der geltenden Bildungsstandards zusammen. In Anlehnung an Knopp (2010b) stellen folgende Aspekte das Extrakt wesentlicher zu erlernender Kompetenzen für den mathematischen Erstunterricht dar:

- Strukturierung und Bündelung von Mengen
- Identifikation, Darstellung und Vergleich von strukturierten und unstrukturierten Mengen sowie Zahlen
- Erkennen der Eigenschaften von Zahlen (gerade/ungerade)
- Erkennen der Beziehungen zwischen Zahlen

- Zerlegen und Zusammenfügen von Mengen und Zahlen (z. B. Halbieren bzw. Verdoppeln)
- Verstehen, Erklären und Anwenden der Rechenoperationen Addition und Subtraktion (enaktiv, ikonisch und symbolisch)
- heuristische Strategien (Nachbaraufgaben, Tauschaufgaben, dekadische Analogien, Umkehraufgaben) zum Lösen von Aufgaben zur Addition und Subtraktion nutzen
- Lösen von Grundaufgaben durch gespeichertes Faktenwissen
- Erkennen, Vergleich und Bewertung verschiedener Rechenwege
- Erkennen und Beschreiben arithmetischer Muster

Die hier aufgeführten wesentlichen arithmetischen Aspekte sollten Kinder im Laufe des mathematischen Anfangsunterrichts beherrschen und stellen die Grundlage für den Mathematikunterricht in den weiterführenden Klassenstufen dar. Alle hier aufgelisteten Punkte finden sich im Wesentlichen in den in Kapitel 2 geschilderten Ausführungen und Modellvorstellungen zur Entwicklung mathematischer Kompetenzen vom Vorschul- bis zum frühen Grundschulalter wieder.

2.10 Schwierigkeiten beim Erwerb früher mathematischer Kompetenzen

Wie zu Beginn des Kapitels dargelegt, verfügen viele Schulanfänger schon über ausgereifte Kenntnisse und Fertigkeiten hinsichtlich mathematischer Inhalte (Schmidt, 2009). Dennoch gibt es Kinder, welche zum Schuleintritt noch nicht über die Kompetenzen verfügen die die Voraussetzung für den Erwerb von dem mathematischen Wissen sind, das in der Schule vermittelt wird. Bei diesen Kindern besteht die Gefahr, dass sich Rückstände zu Defiziten manifestieren, die die gesamte Schullaufbahn beeinträchtigen können (Aunola et al.,

2004; Gaupp et al., 2004; Geary et al., 2000; Helmke & Weinert, 1997; Krajewski, 2003; Krajewski & Schneider, 2006; Kurdek & Sinclair, 2001; Mazzocco et al., 2011; Mazzocco & Thompson, 2005; Stern, 2003; Weißhaupt et al., 2006). Diese Risikokinder unterliegen einer größeren Wahrscheinlichkeit, eine Rechenschwäche auszubilden. Was aber macht eine solche Rechenschwäche aus? Wer ist davon betroffen? Warum? Antworten auf diese Fragen sollen im folgenden Abschnitt auf Basis aktueller nationaler und internationaler Fachliteratur herausgearbeitet werden.

Das Phänomen der Rechenschwäche begrifflich einzugrenzen, ist schwierig. Lorenz und Radatz haben eine beachtliche, aber trotzdem „keineswegs vollständige Liste“ (1993, S. 17) an Begriffen zusammengestellt, welche den Rahmen für Minderleistungen im Rechnen zeichnen soll:

> *„Akalkulie, Alexie für Zahlen, Anarithmasthenie, Anarithmetrie, Anarithmie, asemantische Aphasie, Dyskalkulie, dysgraphische Dyskalkulie, dyslektische Dyskalkulie, Dysmathematica, Entwicklungsdyskalkulie, Fingeragnosie, Gerstmann-Syndrom, graphische Dyskalkulie, ideognostische Dyskalkulie, Kalkulasthenie, Lernstörung im arithmetischen Verstehen, lexikalische Dyskalkulie, motorisch-verbale Dyskalkulie, poerationale Dyskalkulie, Parakalkulie, parietale Dyskalkulie, postläsionale Dyskalkulie, praktognostische Dyskalkulie, Pseudo-Akalkulie, Pseudo-Dyskalkulie, Pseudo-Oligokalkulie, räumliche Akalkulie, sekundäre Akalkulie, sekundäre Dyskalkulie, sekundäre Oligokalkulie, sekundäre Parakalkulie, sensorisch-verbale Dyskalkulie, verbale Dyskalkulie, visuelle Agnosie, Zahlen-Aphasie, Zahlenblindheit, Zahlendysgraphie, Zahlendyslexie, Zahlendyssymbolismus“* (Lorenz & Radatz, 1993, S. 17).

Sowohl in der aktuellen deutschsprachigen Fachliteratur als auch im schulischen Alltag sind die Termini „Dyskalkulie“, „Akalkulie“, „Rechenstörung“ und „Rechenschwäche“ am geläufigsten. Die

Begriffs- und Definitionsvielfalt ist auch in englischsprachiger Fachliteratur zu beobachten. Dort sind u. a. Begriffe wie *mathematics difficulties*, *mathematics disabilities*, *dyscalculia*, *poor math achievement* etabliert (Mazzocco, 2005). Allen gemein ist, dass sie Schwierigkeiten in grundlegenden mathematischen Fähigkeiten beschreiben. Da diese Begriffe oftmals zu Unrecht synonym verwendet werden, soll zunächst eine Systematisierung der wesentlichen, für die schulische Praxis relevanten Termini erfolgen. Krajewski (2003) schlägt dafür die Einteilung in *Rechenschwäche*, *Dyskalkulie* und *Akalkulie* vor, zu welchen alle weiteren Begriffe zugeordnet werden können. Auf diese Einteilung wird in der vorliegenden Arbeit zurückgegriffen.

Akalkulien grenzen sich dabei als in Zusammenhang mit klinischen Störungsbildern stehende, sekundäre Rechenschwierigkeiten ab. Darunter sind erworbene Minderleistungen im Rechnen zusammengefasst, welche das Resultat hirnpathologischer Schädigungen sind oder als Begleitsymptom einiger neurologischer oder psychiatrischer Störungen einhergehen, z. B. bei Demenzen oder dem Gerstmann-Syndrom (Krajewski, 2003).
Unter Dyskalkulie fasst Krajewski (2003) das Auftreten schwacher mathematischer Leistungen zusammen, die in signifikanter Diskrepanz sowohl zum Alter und der allgemeinen Intelligenz als auch zu anderen Leistungen eines Schülers stehen (Diskrepanzkriterium). Dieses Diskrepanzkriterium ist erst erfüllt, wenn die Rechenleistungen unterdurchschnittlich sind, während Lese-Rechtschreibleistungen und der Intelligenzquotient mindestens durchschnittlich ausfallen. Beeinträchtigungen mathematischer Leistungen solcher Art werden in der International Classification of Diseases (ICD-10, Dilling et al., 2011) und im Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR, American Psychiatric Association, 2005) unter dem Begriff *Rechenstörungen* bzw. *Mathematics Disorders* zusammengefasst. Die Begriffe

Rechenstörung und Dyskalkulie können somit laut der von Krajewski (2003) vorgeschlagenen Kategorisierung synonym verwendet werden.

Rechenstörungen, in isolierter Form oder auch in Kombination mit Lese-Rechtschreib-Störungen, werden in der ICD-10 unter F81 als umschriebene Entwicklungsstörungen schulischer Fähigkeiten eingeordnet. Damit handelt es sich um eine Störung des normalen Fertigkeitserwerbs mit Beginn in der frühen Kindheit, welche nicht durch mangelnde Lerngelegenheit, eine Intelligenzminderung, eine erworbene Hirnschädigung oder -krankheit verursacht ist (Dilling et al., 2011). Die (isolierte) Rechenstörung ist mit F81.2 kodiert und wie folgt konkretisiert:

> *„Das Defizit betrifft vor allem die Beherrschung grundlegender Rechenfertigkeiten, wie Addition, Subtraktion, Multiplikation und Division, weniger die höheren mathematischen Fertigkeiten, die für Algebra, Trigonometrie, Geometrie oder Differential- und Integralrechnung benötigt werden.“* (Dilling et al., 2011, S. 277)

Im DSM-IV-TR finden sich Rechenstörungen (mathematics disorders) auf der Achse I, d. h. unter den klinischen Störungen als auch unter anderen klinisch relevanten Problemen, als Beeinträchtigungen, welche gewöhnlich im Kleinkinds-, Kindes- oder Jugendalter diagnostiziert werden. Neben der Lesestörung (*Reading Disorder*, 315.00) und der Störung des schriftlichen Ausdrucks (*Disorder of Written Expression*, 315.20) ist die Rechenstörung (315.10) als eine inhaltlich abgegrenzte Lernstörung ausgewiesen. Anders als in der ICD-10 werden zur Diagnose einer Rechenstörung im DSM-IV-TR nicht die vier Grundrechenarten als gestörte Leistungsbereiche angegeben. Die Definition des DSM-IV-TR beschränkt sich demnach nicht auf den Bereich der Arithmetik, sondern bezieht auch weitere Bereiche wie bspw. die Geometrie mit

ein. Auch hier werden die Rechenleistungen sowohl im Vergleich zur Intelligenz als auch zu weiteren schulischen Leistungen bewertet.

Mit dem Begriff *Rechenschwäche* umschreibt Krajewski (2003) das Phänomen schwacher Leistungen, welche zu den untersten 5 % bis 25 % der Leistungen Gleichaltriger zählen. Dabei verzichtet Krajewski (2003) bei der Definition von Rechenschwäche auf die Berücksichtigung einer Diskrepanz zum Alter, zur Intelligenz oder zu weiteren schulischen Leistungen als Kriterium. Rechenschwächen können somit als mildere, Rechenstörungen als gravierendere Form der Minderleistung im mathematischen Bereich angesehen werden.

Die der ICD-10 und dem DSM-IV zugrunde liegende Diskrepanzdefinition wurde in der Fachliteratur aus verschiedenen Gründen schon mehrfach kritisiert. So werden durch das Diskrepanzkriterium Kinder aus der Definition ausgeschlossen, die neben Intelligenzminderungen und/oder Lese- bzw. Rechtschreibproblemen auch Schwierigkeiten im Bereich des Rechnens aufweisen. Ein weiterer Kritikpunkt ergibt sich aus den Schwierigkeiten bei der Diagnose schwacher Mathematikleistungen. Je nach verwendetem Messinstrument und angelegtem Cutoff-Wert (Grenzwert), werden jeweils verschiedene Personengruppen als rechenschwach bzw. -gestört angenommen (u. a. Grube 2008; Hasselhorn, Mähler & Grube, 2008; Koch, 2005; Krajewski, 2003; Lorenz, 2003; Moser Opitz, 2004). Auch wenn der Einfluss der Intelligenz auf die Leistungen im Bereich der Mathematik nicht bestritten wird (Helmke & Weinert, 1997; Jordan, Hanich & Kaplan, 2003; Krajewski & Schneider, 2006), so hat sich doch gezeigt, dass die Entwicklung von Kindern mit schwachen Rechenleistungen sich nicht grundlegend von der normgerecht entwickelter Kinder unterscheidet, sondern lediglich verzögert verläuft (Goldman, Pellegrino & Mertz, 1988; Lorenz & Radatz, 1993; Werner, 2007). So wird auch von vielen Forschern vorgeschlagen, auf das

Diskrepanzkriterium zu verzichten und alle Kinder mit Schwierigkeiten – definiert bspw. als die unteren 20 % bzw. 25 % in bereichsspezifischen Testverfahren – als lern- bzw. konkreter als rechenschwach anzusehen (L.S. Fuchs, D. Fuchs & Prentice, 2004; Geary, 2004; Geary & Hoard, 2005; Jiménez Gonzalez & Garcia Espínel, 1999; McLean & Hitch, 1999; Wilson & Swanson, 2001). In diesem Zusammenhang postuliert Lorenz (2003) „all jene Kinder in die Förderung aufzunehmen, deren Lernfortschritte, durch welche Gründe auch immer, als unzureichend angesehen werden“ (S. 15). Gestützt werden diese Auffassungen durch zahlreiche Studien, in denen aufgezeigt wurde, dass sich Gruppen, welche das Diskrepanzkriterium erfüllen, mit den Gruppen überschneiden, bei denen auf das Diskrepanzkriterium verzichtet wird. Es ergeben sich also keine klar zu unterscheidenden Gruppenprofile (Aaron, 1997; Fletcher, Francis, Shaywitz et al., 1998; Fletcher, Lyon, Barnes et al., 2002).

Murphy, Mazzocco, Hanich und Early (2007) untersuchten verschiedene Gruppen von Kindern mit Schwierigkeiten im Bereich Mathematik über einen Zeitraum von drei Jahren (Kindergarten bis zweite Klassenstufe). Die Gruppen wurden durch Festlegung unterschiedlicher Cutoff-Werte definiert: Kinder mit Prozentrang < 10 (N = 22) und Kinder mit einem Prozentrang zwischen 11 und 25 (n = 42) in einem standardisierten Mathematiktest. Die Ergebnisse wiesen auf keine klar trennbaren Profile der beiden Gruppen hin. Resümierend kann festgehalten werden, dass der Fokus nicht darauf liegen sollte,

> *„Gruppen von Schülerinnen und Schülern zu definieren, sondern zu untersuchen, welche Schwierigkeiten diese beim Mathematikerwerb haben. Dieses Vorgehen ist darum von Bedeutung, weil es möglich wird, die Schwierigkeiten der Schülerinnen und Schüler bezogen auf den mathematischen Inhalt zu beschreiben und dadurch Grundlagen für die*

> *Diagnostik und Förderung abzuleiten. Es geht dabei nicht mehr um das Feststellen einer Störung bei einzelnen Individuen, sondern um die Beschreibung von mathematischen Inhaltsbereichen, bei deren Erwerb (häufig) Schwierigkeiten auftreten“* (Scherer & Moser Opitz, 2010, S. 12).

So vielfältig wie die Begrifflichkeiten zur Definition einer Rechenschwäche ausfallen, so groß ist auch die Anzahl differierender Angaben zur Häufigkeit des Auftretens. Wie zuvor erwähnt ergeben sich verschiedene Prävalenzraten von Rechenschwierigkeiten in Abhängigkeit von den verwendeten Definitionskriterien und den verwendeten Messverfahren. In einer Studie von Lewis, Hitch und Walker (1994), in der 1206 Kindern im Alter zwischen neun und zehn Jahren untersucht wurden, lag die Prävalenzrate isolierter Schwierigkeiten im Bereich des Rechens bei 1.3 %. Innerhalb beider Gruppen war das Geschlechterverhältnis gleich verteilt. Shalev, Auerbach, Manor und Gross-Tsur (2000) fanden in ihrer Metaanalyse zu den Auftretenshäufigkeiten in verschiedenen Ländern heraus, dass sich, obwohl die Definitionskriterien von Land zu Land unterschiedlich ausfielen, vergleichbare Prävalenzraten zwischen 3 % und 6 % ergaben. Auch hier zeigte sich jeweils ein relativ gleiches Verhältnis zwischen Mädchen und Jungen.

Bezogen auf den deutschsprachigen Raum variiert die Häufigkeit der Kinder mit Rechenschwächen von 4.4 % (Klauer, 1992) über 4.7 % (von Aster, Deloche, Dellatolas & Meier, 1997) bis 6.6 % (Hein, Bzufka & Neumärker, 2000). Kombinierte Rechen- und Lese-Rechtschreibstörungen weisen einer Studie von von Aster et al. (2007) zufolge 4.2 % der insgesamt 337 untersuchten Zweitklässler auf.

Die aufgeführten Studien beziehen sich auf Diskrepanzdefinitionen, sodass davon auszugehen ist, dass die Gruppe der Kinder mit

gravierenden Schwierigkeiten im Bereich des Rechnens in den untersuchten Kohorten größer ist. So gehen Lorenz und Radatz (1993) sowie Schipper (2003) davon aus, dass mindestens 15 % aller Schüler förderungsbedürftige Schwierigkeiten im Rechnen aufweisen. Die in Abschnitt 1.1 angeführten Ergebnisse von TIMSS und PISA zeigen sogar, dass etwa 20 % der Schülerschaft große Schwierigkeiten im Bereich der Mathematik haben. Übertragen auf eine durchschnittliche Klassengröße von 25 Schülern muss eine Lehrkraft davon ausgehen, dass immerhin fünf ihrer Schüler Schwierigkeiten und mindestens ein Schüler gravierende Probleme im Fach Mathematik aufweisen oder noch aufweisen werden.

Warum aber scheitern einige Kinder beim Erwerb mathematischer Kompetenzen? Einen gültigen Konsens über die Ursachen von Rechenstörungen gibt es gegenwärtig nicht, Ansätze hingegen viele. Jacobs und Petermann (2003) gehen von einem multikausalen Erklärungsmodell für Rechenstörungen aus, welches neben genetischen, psychologischen und psychosozialen Faktoren des Kindes auch umweltbezogene Faktoren, wie bspw. ungünstige Eltern-Kind- bzw. Lehrer-Kind-Interaktionen oder den Bildungsgrad der Eltern berücksichtigt. Fritz und Ricken (2008) gehen ebenso von einem multifaktoriell bedingten Ursachengefüge aus. „Als Kern der Problematik werden eine verzögerte Entwicklung rechenrelevanter Lernvoraussetzungen, bzw. Beeinträchtigungen im Verständnis mathematikspezifischer Inhalte angenommen“ (Fritz & Ricken, 2008, S. 15). In diesem Zusammenhang aktuell diskutierte Modellvorstellungen kommen aus der Entwicklungs- und der Neuro- bzw. Kognitionspsychologie und wurden bereits in Abschnitt 2.1 bis 2.8 dieser Arbeit näher erläutert (u. a. von Aster, 2003; von Aster et al., 2007; Dehaene, 1992; Fritz et al., 2007; Grube & Barth, 2004; Krajewski, 2003; Krajewski & Schneider, 2006).

Weiterhin werden jedoch auch genetische Faktoren für die Entstehung von Schwierigkeiten beim Erlernen des Rechnens angenommen. Zur Untersuchung genetischer Zusammenhänge bei Rechenstörungen wurde im Jahr 1997 eine Studie mit 40 eineiigen und 23 zweieiigen Zwillingsgeschwistern gleichen Geschlechts (mindestens einer der Zwillinge war von einer Rechenstörung betroffen) durchgeführt (Alarcón, DeFries, Light & Pennington, 1997). Dabei zeigten 58 % der eineiigen und 39 % der zweieiigen Zwillingsgeschwister ebenso Beeinträchtigungen im Rechnen. Shalev, Manor, Kerem, Ayali, Badichi, Friedlander und Gross-Tsur kamen im Jahr 2001 zu ähnlichen Ergebnissen. Sie fanden in Familienanalysen heraus, dass 66 % der insgesamt 21 Mütter, 40 % der insgesamt 22 Väter, 53 % der insgesamt 90 Geschwister und 44 % der insgesamt 16 Verwandten zweiten Grades von Kindern mit Dyskalkulie (N = 39), ebenfalls die Kriterien einer Rechenstörung erfüllten. Das Auftreten einer Rechenstörung innerhalb der nahen Verwandtschaft erhöht demnach das Risiko, eine Dyskalkulie zu entwickeln um das zehnfache verglichen zur Normalbevölkerung (Shalev et al., 2001).

Aus kognitionspsychologischer Perspektive werden durch die Modellannahmen zur kognitiven Repräsentation von Zahlen und Mengen (triple-code-model nach Dehaene, 1992; siehe Abschnitt 2.7), durch Modelle zur Informationsverarbeitung (Atkinson & Shiffrin, 1968; Baddeley, 1986; siehe Abschnitt 2.8) sowie durch Netzwerkmodelle zur Speicherung von Faktenwissen in Wissensnetzen (Ashcraft, 1992; siehe Abschnitt 2.7) Hinweise zur Entstehung von Rechenschwierigkeiten gegeben. Es zeigt sich, dass Kinder mit Rechenschwäche Beeinträchtigungen in den Repräsentationsmodulen von Zahlen und Mengen (triple-code-model nach Dehaene, 1992) sowie bei Transkodierprozessen zwischen diesen Modulen aufweisen. Besondere Schwierigkeiten zeigen sie bei einfachen Zählfähigkeiten, erstem Wissen über die arabischen

Zahlen sowie deren analogen Größenzuordnung (von Aster et al., 2007). Kinder ohne Schwierigkeiten im Rechnen gelingt es, im Laufe der Grundschulzeit ein differenziertes Wissensnetz über arithmetische Fakten aufzubauen, auf welches sie schnell zurückgreifen können. Rechenschwache Kinder hingegen scheinen Probleme bei der Speicherung solcher Informationen zu haben (Grube 2006; Ostad, 1998; Temple & Sherwood 2002). Sie nutzen den Abruf arithmetischer Fakten aus der Wissensbasis im Gegensatz zu Kindern ohne Rechenschwäche deutlich seltener als Lösungsstrategie (Bull & Johnston, 1997; Geary, 1990; Geary, Bow-Thomas & Yao, 1992; Geary & Brown, 1991; Geary, Brown & Samaranayake, 1991). Bull, Johnston und Roy (1999) weisen durch ihre Studie zudem darauf hin, dass es Kindern mit Rechenschwäche schwer fällt, die Aufmerksamkeit bei einer länger andauernden Aufgabe aufrecht zu erhalten. Bull et al. (1999) verweisen in diesem Zusammenhang auf eine eingeschränkte Funktion der zentralen Exekutiven im Sinne Baddeleys (1986), welche mit der Aufmerksamkeitssteuerung assoziiert ist. Von Aster et al. (2007) nehmen die „ineffiziente Hemmung irrelevanter Information […], die das Arbeitsgedächtnis zusätzlich beansprucht und zu einer anhaltenden Konkurrenz mit relevanten Informationen bei der Beantwortung von sowohl numerischen als auch sprachlichen Aufgaben führt“ (S. 94) als Ursache dafür an. Die Informationsverarbeitung, die wesentlich durch die Aufmerksamkeit eines Menschen determiniert ist, wird dadurch beeinträchtigt. Bei gedächtnisbeanspruchenden Aufgaben, wie bspw. dem Zählen, kann fehlende kontinuierliche Aufmerksamkeit schnell dazu führen, dass sich Kinder verzählen. Die Beeinträchtigung der Aufmerksamkeit kann jedoch nicht als spezifisches Merkmal und demnach nicht als alleinige Erklärung einer Rechenschwäche angenommen werden (von Aster et al., 2007; Grube, 2009). Je extremer die Minderleistung im Rechnen ist und je mehr der Einfluss der Intelligenz kontrolliert wird, desto stärker scheint die Rechenleistung der betroffenen Kinder

durch spezifisch zahlen- und mengenbezogene Defizite beeinträchtigt zu sein (Butterworth & Reigosa, 2007).

Folgt man aktuellen Forschungsergebnissen der Entwicklungspsychologie, ergeben sich Schwierigkeiten im Rechnen aufgrund nicht bzw. ungenügend ausgeprägter Einsichten in grundlegende mathematische Konzepte (Aunola et al., 2004; Gaupp et al., 2004; Geary et al., 2000; Helmke & Weinert, 1997; Krajewski, 2003; Krajewski & Schneider, 2006; Kurdek & Sinclair, 2001; Mazzocco et al., 2011; Mazzocco & Thompson, 2005; Stern, 2003; Weißhaupt et al., 2006). Diese Einsichten stellen Meilensteine in der Entwicklung von Rechenkompetenzen dar und besitzen eine gewisse Ordnung im Sinne einer hierarchischen Stufung (siehe dazu Abschnitt 2.4 bis 2.5). Ganz zentral sind dabei folgende Konzepte und Einsichten (Fritz et al., 2007; Ricken et al., 2011):

1. Zahlen können zum Zählen von Objekten benutzt werden (Zählzahl-Aspekt). Über das wiederholte Üben des Zählens gelangen Kinder zu einer gesicherten Kenntnis der Zahlwortreihe. Darüber hinaus werden Grundlagen für den gesicherten Zahlbegriff gelegt.
2. Kinder entwickeln ein inneres Bild über die Ordnung der Zahlen, ein mentaler Zahlenstrahl entsteht. Mit dieser Einsicht können erste Aufgaben zur Diskrimination von Zahlen (anhand der Position auf dem Zahlenstrahl) bzw. zur Addition (durch zählende Strategien) gelöst werden. Zahlen werden in dieser Phase der Entwicklung noch ordinal, d. h. als Rang in der Zahlenfolge betrachtet.
3. Mengen- und Zahlenwissen wird ineinander integriert. Kinder verstehen, dass Zahlen die Mächtigkeit von diskreten Mengen angeben, d. h. Zahlen werden als Anzahlen bzw. Kardinalzahlen bewusst. Dabei gibt die beim Zählen zuletzt genannte Zahl die Mächtigkeit der ausgezählten Menge an.

4. Mengen können in Teilmengen zerlegt und aus diesen zu einem Ganzen zusammengesetzt werden (Teile-Ganzes-Konzept). Additions- und Subtraktionsaufgaben können somit als Schritte des Zusammen- bzw. Zerlegens verstanden werden. Bezogen auf Nachbarzahlen ergibt sich die Einsicht, dass die nachfolgende Zahl immer um genau eins größer, die vorangegangene Zahl immer um genau eins kleiner ist.
5. Zahlen beschreiben ebenso Abschnitte auf dem Zahlenstrahl, d. h. Beziehungen zwischen Zahlen, bspw. repräsentiert die 5 den Abstand/die Relation zwischen z. B. 2 und 7, jedoch auch zwischen 11 und 16 (relationaler Zahlaspekt).

Das sichere und flexible Operieren mit Zahlen, bspw. beim Addieren, beruht demnach wesentlich auf einem Verständnis für Teile-Ganzes-Beziehungen zwischen Mengen und auf der Einsicht, dass Zahlen bestimmte Mengen repräsentieren. Erst durch die Integration dieser Konzeptvorstellungen werden Relationen zwischen Zahlen bewusst. Fehlen diese Einsichten und Konzeptvorstellungen bzw. sind sie nur eingeschränkt ausgeprägt, können Ergebnisse zu Aufgaben wie 3 + 8 = ? nur durch uneffektive Lösungsstrategien (siehe dazu auch Abschnitt 2.2 bis 2.5) ermittelt werden. In diesem Kontext merkt Ricken (2011) an: „Leistungsfähigkeit entsteht dann, wenn Mathematik nicht im Auswendiglernen von Rechenvorschriften besteht, sondern wenn Kinder Einsichten in Zusammenhänge zwischen Zahlen bzw. Mengen erwerben“ S. 254). Da rechenschwache Kinder Zahlen meist nur ordinal interpretieren (Gerster, 2009; Ricken, 2011), bleiben ihnen nur das Raten der Lösung oder das Abarbeiten von auswendiggelernten Lösungsalgorithmen als Strategien, jedoch ohne das grundlegende Verständnis. Studien belegen, dass zählbasierte Vorgehensweisen (manchmal unter Zuhilfenahme der Finger) auch noch bis in höhere Schuljahre hinein zu den häufigsten Lösungsstrategien rechenschwacher Kinder zählen (Kaufmann & Wessolowski, 2005,

2006; Ostad, 1997; Schipper, 2003; Wehrmann, 2007). Trotz intensiver Nutzung zählender Strategien können rechenschwache Kinder diese weder sicherer noch schneller anwenden als Kinder ohne Probleme im Rechnen (Bull & Johnston, 1997). Dazu kommt, dass diese Strategien im Laufe der Zeit eher ineffizient und aus kognitiver Perspektive übermäßig ressourcenaufwändig sind. Das hat hohe Bearbeitungszeiten und häufige Fehler zur Folge. Der Faktenabruf aus der Wissensbasis auswendiggelernter Aufgaben sowie deren Lösungen als weitere Lösungsmöglichkeit ist, aufgrund des bei rechenschwachen Kindern meist nur rudimentär ausgebildeten Wissensnetzes für mathematische Inhalte, meistens beeinträchtigt. Aus kognitions- und entwicklungspsychologischer Sicht sollte optimale (sonder-)pädagogische Förderung zum Ziel haben, einen sicheren Umgang mit Mengen und Zahlen und das Entwickeln effektiver Rechenstrategien zu ermöglichen und durch stetig wiederholtes automatiserendes Üben ein breites Netz an arithmetischem Faktenwissen im Langzeitgedächtnis aufzubauen (Ashcraft, 1992; Atkinson & Shiffrin, 1968; Krajewski & Schneider, 2006; Fritz et al., 2007).

2.11 Schlussfolgerungen

In Kapitel 2 wurden ausgewählte Annahmen und Modelle zur Entwicklung von mathematischen Kompetenzen vom Vorschul- bis zum Grundschulalter vorgestellt. Dabei wurden Ansätze aus verschiedenen Wissenschaftsdisziplinen erläutert. Aus entwicklungspsychologischer Perspektive wurden Meilensteine in der Entwicklung mathematischer Kompetenzen herausgearbeitet, welche Kinder im Entwicklungs- und Lernprozess erreichen sollten, um flexibel und sicher mit mathematischen Inhalten umgehen zu können. Aus kognitionspsychologischer Sicht wurde der Frage nachgegangen, wie Zahlen geistig repräsentiert und verarbeitet werden. Es wurde in diesem Zusammenhang auf die zentrale

Bedeutung des Arbeitsgedächtnisses sowie der Aufmerksamkeit und deren Steuerung aufmerksam gemacht. Die genannten Aspekte finden auch in den Bildungsstandards sowie den Lehrplänen der deutschen Bundesländer Berücksichtigung. Darüber hinaus wurden auf Grundlage dieser theoretischen Ausarbeitungen verschiedene Erklärungsansätze für die Entstehung von Rechenschwierigkeiten ausgearbeitet. Auch wenn nicht eindeutig geklärt ist, wodurch Rechenschwächen verursacht werden, kann zusammenfassend festgehalten werden, dass ihnen Beeinträchtigungen in der Entwicklung bereichsspezifischer als auch bereichsübergreifender Kompetenzen zugrunde liegen. Rechenschwächen persistieren langfristig, sofern Kinder mit Schwierigkeiten beim Erwerb mathematischer Kompetenzen nicht frühzeitig erkannt und nicht gezielt (sonder-)pädagogisch gefördert werden. Diese Kinder haben eine nur geringe Chance auf eine zukünftige Leistungsverbesserung und es besteht die Gefahr, dass ihre gesamte schulische Laufbahn negativ beeinträchtigt sein wird. Zu den Folgen können nicht nur eine erfolglose Schullaufbahn, sondern auch ein negativer Einfluss auf das weitere berufliche und psychosoziale Leben gehören. Dies wird durch (eingangs benannte) Studien belegt, die gezeigt haben, dass das bereichsspezifische Vorwissen am höchsten mit der zukünftigen Mathematikleistung von Kindern korreliert (Aunola et al., 2004; Gaupp et al., 2004; Geary et al., 2000; Helmke & Weinert, 1997; Krajewski, 2003; Krajewski & Schneider, 2006; Kurdek & Sinclair, 2001; Mazzocco et al., 2011; Mazzocco & Thompson, 2005; Stern, 2003; Weißhaupt et al., 2006). Zudem konnten Shalev, Manor und Gross-Tsur (2005) im Zuge einer Longitudinalstudie 104 Kinder mit diagnostizierter Dyskalkulie über sechs Jahre wissenschaftlich begleiten. Es stellte sich heraus, dass 40 % der Probanden, die zu Beginn der Studie 11 Jahre alt waren, nach sechs Jahren immer noch die Kriterien für eine Dyskalkulie erfüllten. Von den untersuchten Personen gehörten 95 % zum unteren Quartil bezogen auf die Normgruppe gleichen Alters. Parsons und Bynner (2005) berichten,

dass Erwachsene in einem Alter von 30 Jahren, welche nur schwache Rechenleistungen bzw. schwache Rechenleistungen kombiniert mit schwachen Leseleistungen aufweisen, einem höheren Risiko ausgesetzt sind, arbeitslos oder auch depressiv zu sein.

Welche pädagogischen Schlüsse lassen sich nun aus den Erkenntnissen der Forschungen zur Entwicklung von mathematischen Kompetenzen und Rechenschwächen ableiten?

Eine zentrale Aufgabe (sonder-)pädagogischer Praxis in der Schule ist, der Entstehung von Rechenschwächen vorzubeugen (primäre Prävention) und sich manifestierten Schwierigkeiten beim Erwerb mathematischer Kompetenzen mit adäquater spezifischer Förderung zu begegnen. Die Prävalenzraten, die Entwicklungsstabilität und die Folgen von Rechenschwächen sprechen dafür, schon im frühen schulischen bzw. bereits im vorschulischen Setting den Fokus der pädagogischen Arbeit auf das Vermeiden von Entwicklungsrückständen bzw. das Schließen von Wissenslücken im Lernbereich Mathematik zu legen. Der Mathematikunterricht der ersten Klasse kann nicht einfach bei null ansetzen und die im Lehrplan geforderten Inhalte im Gleichschritt erarbeiten. Vielmehr rückt das Anknüpfen an das individuelle Vorwissen der Kinder in den Vordergrund. Besonders Lerninhalten wie sichere Zählkompetenzen, die Einsicht in das Teile-Ganzes-Konzept sowie dem Verständnis für die Relationen von Zahlen zueinander sollte ein hoher Stellenwert eingeräumt werden. An dieser Stelle kommt der unterrichtenden Lehrkraft eine bedeutende Verantwortung zu, denn sie muss nicht nur den bereichsspezifischen Entwicklungsstand ihrer Schüler kennen, sondern diesen fortlaufend aktualisieren und darauf ausgerichtete individuelle und differenzierende Unterrichtsangebote schaffen. Dies bedarf einer hohen Sensibilität für die Bedürfnisse der heterogenen Schülerschaft. Dabei muss die pädagogische Tätigkeit nah an

anerkannten theoretischen Modellvorstellungen angeknüpft sein. So führt Ricken (2011) an:

> *„Um aus einer entwicklungspsychologischen Perspektive Rechenstörungen zu betrachten ist es wichtig, die ‚Entwicklungsstellen', ‚Meilensteine' oder auch ‚Nadelöhre' zu benennen, die die frühe Entwicklung in dieser Domäne bestimmen. Zu fragen ist einerseits, welche dies sind und wie diese in Beziehung zueinander stehen. Wenn sich eine Ordnung in den Meilensteinen herstellen lässt, wäre ein Raster gegeben, um den aktuellen Entwicklungsstand eines Kindes zu verorten, die Zone der nächsten Entwicklung qualitativ zu benennen und bereits erworbene Konzepte zu inventarisieren"* (S. 258).

Besonders essenziell erscheint die Identifikation der durchschnittlich etwa fünf Kinder einer Klasse, welche Schwierigkeiten beim Erwerb des Rechnens zeigen bzw. im Laufe der Zeit noch zeigen werden. Dazu müssen Lehrkräfte hinsichtlich der frühzeitigen und präzisen Erfassung der Problembereiche ihrer Schüler sensibilisiert und befähigt werden (Baumert & Kunter, 2006). Speziell vor dem Hintergrund beurteilungsverzerrender Effekte, wie dem Milde-, dem Strenge-, oder dem Halo-Effekt, stellt sich die Frage, wie gut und sicher Lehrkräfte die Leistungen ihrer Schüler einschätzen können. In diesem Kontext haben Hoge & Coladarci (1989) 16 Studien zur Urteilsgenauigkeit von Lehrkräften metaanalysiert. Es wurden Korrelationen zwischen r = .48 und r = .92 (Median bei .69) zwischen Lehrerurteilen zu den tatsächlichen Leistungen der Schüler ermittelt. Die Autoren kommen zu folgendem Schluss: „The achievement judgments revealed themselves to be generally accurate. Still, there was clearly some degree of error operating, and, further, levels of accuracy varied across teachers. There is, therefore, room for improvement" (S. 310). Ähnliche Ergebnisse konnten auch für Deutschland repliziert werden (Bromme, 1997; Schrader, 2006).

Erschwert wird die Einschätzung von Schülerleistungen durch den Fakt, dass rechenschwache Kinder oft über ausgefeilte Zählstrategien verfügen, welche zu dem Trugschluss führen können, dass wesentliche Konzeptvorstellungen schon erfolgreich bei diesen Kindern erarbeitet wurden. Darüber hinaus gibt Coladarci (1986) zu bedenken: „Because of the summary nature of such teacher judgments, little is disclosed about the teacher's specific knowledge of what the student has and has not mastered in some domain“ (S. 142). Für die erfolgreiche Förderung von Kindern mit Schwierigkeiten im Rechnen sind differenzierte Aussagen über Stärken und Schwächen der Schüler grundlegend. Um Rückstände im Rechnen möglichst frühzeitig zu erkennen, ist es nicht nur notwendig zu verstehen, wie sich erste mathematische Fähigkeiten entwickeln, damit eng verbunden sind auch Forderungen nach validen Methoden zur Standortbestimmung der Schüler, denn, so resümiert auch Gasteiger (2007):

> *„Erziehende und Lehrkräfte, die die mathematischen Kompetenzen der Kinder beobachten, gezielt erheben und verstehend dokumentieren, erhöhen gleichzeitig ihr Bewusstsein für mathematische Lerngelegenheiten. […] Der diagnostische Blick auf die mathematischen Kompetenzen der Kinder ist ein entscheidender Ansatzpunkt für die Zusammenarbeit der Institutionen und den fachlichen Austausch“* (S. 4).

Zentrale Aufgabe der wissenschaftlichen Forschung stellt demnach die Unterstützung der schulischen Praxis dar. Diese erfordert die Entwicklung und systematische Evaluation von Diagnose- und Förderprogrammen zur (Früh-)Identifikation und Förderung rechenschwacher Kinder bzw. derer, die einem Risiko unterliegen, eine Rechenschwäche auszubilden. Vor dem Hintergrund der nicht immer präzise ausfallenden Lehrerbeurteilungen erscheint vor allem eine Unterstützung bei der Erfassung von Schülern mit Lern- und Entwicklungsrisiken im Bereich der Mathematik angezeigt. Diese

Argumente sprechen für die Konzeption von Messverfahren, welche Lehrkräften helfen, objektive, reliable und valide Einschätzungen der Schüler zu ermöglichen und gleichzeitig einfach in den schulischen Alltag zu integrieren sind. Dies erscheint speziell für den Bereich der Erstmathematik im Schuleingangsunterricht notwendig zu sein. Dabei geht es jedoch nicht um die Identifikation von rechengestörten Kinder im Sinne einer Stigmatisierung oder Selektion bzw. Aussonderung. Vielmehr sollten die durch Tests gewonnen Informationen Grundlage bei der Planung entsprechender spezifischer Unterrichts- bzw. Förderangebote darstellen. Darüber hinaus dürfen die diagnostisch erhobenen Informationen zur Rechenkompetenz eines Kindes nicht als allgemeingültiger, überdauernder Status betrachtet werden. Der Lernverlauf des Schülers muss daher in regelmäßigen, kurzfristigen Abständen aktualisiert werden. Dies liefert zugleich wertvolle Hinweise zur Modifizierung bzw. Optimierung der Unterrichts- und Fördermaßnahmen unter Berücksichtigung des individuellen Entwicklungsstands des Schülers. Die zu diesem Zweck in den USA entwickelten und dort bereits etablierten curriculumbasierten Messungen (curriculum-based measurements, CBM) zur Unterstützung einer prozessbegleitenden und datenbasierten Diagnostik sollen im nächsten Kapitel eingehend beschrieben werden.

3 Diagnostik mathematischer Kompetenzen als zentrales Element schulischer Prävention

In den vorangegangenen Abschnitten wurde die Bedeutung von frühen zentralen mathematischen Einsichten als Meilensteine beim Erwerb von Rechenkompetenzen herausgearbeitet. Diese Meilensteine sind hierarchisch aufgebaut und in verschiedenen Modellen systematisiert worden (Fritz et al., 2007; Krajewski & Schneider, 2006). Eine Verortung von Schülerleistungen in einem solchen Modell liefert nicht nur Hinweise, welche Einsichten ein Kind noch nicht für sich erschlossen hat, sondern gibt ebenso Informationen darüber, welche Aspekte in Unterricht und Förderung vertieft werden müssen. Darüber hinaus wurde in Abschnitt 2 gezeigt, dass Kinder mit sehr unterschiedlichen Lernvoraussetzungen in die Schule kommen. Während die Mehrheit der Schulanfänger, entgegen der Lehrereinschätzungen, bereits ein ausdifferenziertes Mengen- und Zahlenwissen besitzen, verfügen einige Kinder lediglich über rudimentäres mathematisches Wissen, welches den unteren Entwicklungsstufen oben genannter Modelle entspricht. Das Anknüpfen an beim Schüler vorhandene Kompetenzen gilt als zentrale Aufgabe von Lehrkräften und als Voraussetzung für adaptiven Unterricht (Ingenkamp & Lissmann, 2008). Dies bedarf präziser Standortbestimmungen der Kinder durch den Lehrer. „Dazu sind diagnostisches Wissen (über Fähigkeiten und Leistungen von Schüler [sic!] und die Schwierigkeit von Aufgaben) und diagnostische Fertigkeiten (Beobachtungsfähigkeiten, Beherrschung von Diagnoseinstrumenten) erforderlich“ (Schrader & Helmke, 2002, S. 48). Diese Aspekte weist auch die Kultusministerkonferenz (KMK, 2004a) in ihren Standards für die Lehrerausbildung aus. Zentrale Kompetenz und alltägliche Anforderung an Lehrpersonen ist dabei neben dem *Unterrichten*, *Erziehen* und *Innovieren* das *Beurteilen*. Die Kompetenz des Beurteilens haben angehende Lehrkräfte der KMK zufolge dann erreicht, wenn sie u. a. nachfolgende Anforderungen erfüllen:

„Die Absolventinnen und Absolventen...

- wissen, wie unterschiedliche Lernvoraussetzungen Lehren und Lernen beeinflussen und wie sie im Unterricht berücksichtigt werden. […]
- erkennen Entwicklungsstände, Lernpotentiale, Lernhindernisse und Lernfortschritte.
- erkennen Lernausgangslagen und setzen spezielle Fördermöglichkeiten ein“ (KMK, 2004a, S. 11).

Dass es sich hierbei um Forderungen, nicht aber um die allgemeingültige Realität handelt, zeigt die in Abschnitt 2.11 (S. 58) erwähnte Metaanalyse von Hoge und Coladarci (1989) zur Einschätzung des Lehrerurteils, in der deutlich wird, dass diesbezüglich Verbesserungen notwendig sind. Wie aber kann eine solche Verbesserung des Lehrerurteils erfolgen?

Die bloße Wahrnehmung und auch die Schülerbeurteilung liefern kein exaktes Abbild der Realität, da dieses durch die Erwartungen und Hypothesen des Beobachters, hier der Lehrkraft, verzerrt wird (*Hypothesentheorie der Wahrnehmung*, Bruner, 1951). Um die Objektivierung von Beobachtungen zu ermöglichen, ist bei der validen Schülerbeurteilung die Anwendung von Hilfsmitteln notwendig. Im schulischen Alltag sind dazu eher informelle Verfahrensweisen, d. h. von Lehrern konstruierte Tests (Heller & Hany, 2002), üblich. Zu informellen Diagnosemaßnahmen zählen alle im schulischen Alltag gewonnenen subjektiven Urteile (Helmke, 2009; Ruiz-Primo & Furtak, 2007), die weniger planmäßig, sondern eher spontaner Natur sind (Bell & Cowie, 2001). So soll das regelmäßige Einfordern von Hausaufgaben, Schülerarbeiten oder Klausuren ein valides Bild über die Leistungsentwicklung der Schüler liefern. Dieses wenig bis nicht standardisierte Vorgehen hat den Vorteil, dass Denkwege und verwendete Strategien differenziert erfasst werden können (Scherer & Moser Opitz, 2010) und eine direkte Verknüpfung

zum Lehr-Lernprozess herstellt wird (Bell & Cowie, 2001), jedoch müssen Objektivitätseinbußen in Kauf genommen werden (Moosbrugger & Kevala, 2012).

Innerhalb der sich seit dem Ende der 1960er Jahren entwickelnden pädagogischen Diagnostik, die sich der Methoden der psychologischen Diagnostik bedient, wird hingegen besonderer Wert auf die wissenschaftliche Absicherung von Messungen zur Leistungsbeurteilung gelegt (Ingenkamp & Lissmann, 2008). Eine solche formale Diagnostik zeichnet sich durch eine planvolle, systematische Durchführung von Beobachtungen und Testungen aus (Helmke, 2009; Ruiz-Primo & Furtak, 2007). In diesem Zusammenhang müssen vor allem die geläufigen wissenschaftlichen Gütekriterien – Objektivität, Reliabilität und Validität – sichergestellt sein (Amelang & Schmidt-Atzert, 2006). Diese hohen Anforderungen an die Güte diagnostischer Verfahren werden verständlich, wenn man berücksichtigt welche Funktionen Tests in der schulischen Praxis haben. Im Allgemeinen sollen diagnostische Verfahren Hilfestellung bei Entscheidungsfindungen bzw. bei der Beantwortung von Fragestellungen leisten. Ingenkamp und Lissmann (2008) konkretisieren die Funktion von Diagnostik durch ihre Begriffsbestimmung der pädagogischen Diagnostik als „alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren“ (S. 13). Die Optimierung individueller Leistungen kann dabei durch unterschiedliche Strategien erreicht werden, wobei die Diagnostik unterschiedliche Funktionen einnimmt. Diagnostische Maßnahmen werden demnach in verschiedenen Dimensionen klassifiziert, einerseits nach ihrer Funktion (*Selektions-* vs. *Modifikationsdiagnostik*, Pawlik, 1982), anderseits nach der Art und Weise des Vorgehens (*formative* vs. *Statusdiagnostik*, Pawlik,

1982; Scriven 1967). Alle Ausprägungen werden nachfolgend erläutert.

Mit dem Ziel der Identifikation von Kindern mit Entwicklungsproblemen wie bspw. einer Rechen- oder Lernstörung und der anschließenden Zuweisung zu Sondermaßnahmen zielt Diagnostik eher auf einen selektiven Aspekt ab (*Selektionsdiagnostik*, Pawlik, 1982). Damit verbunden sind Ressourcenzuweisungen, wie die Teilnahme an spezifischen Förder- bzw. Therapieprogrammen oder gar eine Beschulung bspw. an einer Förderschule im Falle kumulierter Störungen. Um solch einschneidende Entscheidungen zu treffen, werden zumeist formale Tests eingesetzt, welche einen Risikostatus zu einem gegeben Zeitpunkt ermitteln. Eine derartige Herangehensweise wird als Statusdiagnostik bezeichnet (Pawlik, 1982; Scriven 1967). Dabei ist ein Risikostatus dann angezeigt, wenn die Leistung eines Schülers im Vergleich mit einer relevanten Bezugsgruppe als problematisch einzustufen ist (normorientiert, Wright, 2001). Obgleich statusdiagnostische Verfahren eine Reihe von Nachteilen für den schulischen Einsatz mit sich bringen, finden sie in der aktuellen diagnostischen Praxis häufig Anwendung. Ein statusmessender Test ist, abhängig vom zu prüfenden Leistungsbereich, recht umfangreich und zeitintensiv. Da solche Verfahren zudem selten einen direkten Bezug zum aktuell unterrichteten Curriculum haben, sind sie häufig schwierig in den täglichen Unterricht zu integrieren. Statusdiagnostische Verfahren bilden eine Momentaufnahme der zum Messzeitpunkt erhobenen Schülerleistungen ab. In der Regel sind derartige Verfahren für den punktuellen Einsatz zur Mitte bzw. zum Ende eines Schuljahres konzipiert und normiert und nicht für einen wiederholten, enger frequentierten Einsatz im Laufe des Schuljahres. D. h. kurzfristige Entwicklungen der Schülerleistungen (Verbesserung, Stagnation oder Abfall der Schülerleistungen) innerhalb des Schuljahres können

statusdiagnostische Verfahren nur bedingt wahrnehmen (Wright, 2001).

Eine andere Sicht ermöglicht die Modifikationsdiagnostik (Pawlik, 1982). Sie dient „der Untersuchung gestörter Lernprozesse sowie zur Planung und Evaluation pädagogischer Behandlungsmaßnahmen“ (Amelang & Schmidt-Atzert, 2006, S. 16). Hierbei wird das Ziel verfolgt, eine Passung von Unterrichts- bzw. Fördermaßnahmen mit den Voraussetzungen des Kindes zu erreichen. Diagnostik stellt somit die Grundlage für die organisatorische und inhaltliche Ausgestaltung von Lern- bzw. Förderangeboten dar. Dabei werden in der Regel nicht nur punktuelle diagnostische Informationen herangezogen, sondern durch fortlaufende, regelmäßige Untersuchungen wird der Lern- bzw. Förderprozess ständig begleitet und evaluiert. Eine solche regelmäßig frequentierte Leistungsmessung wird nach Scriven (1967) auch als formativ bezeichnet. Der Erfolg der Intervention wird dabei in der Regel nicht durch den Vergleich mit einer geeigneten Bezugsgruppe, sondern durch das Erreichen bestimmter vorgegebener Kriterien (kriteriumsorientiert) bemessen (Amelang & Schmidt-Atzert, 2006). Diese prozessbegleitende Diagnostik zur Optimierung von Lehr-Lern-Gefügen versucht, die oben genannten Nachteile der Statusdiagnostik, mit der meist das Ziel der Selektion verfolgt wird, zu überwinden.

Welche Bedeutung hat nun die Diagnostik bezogen auf den Mathematikunterricht in der Grundschule? Wie sollte Diagnostik im Sinne einer schulischen Leistungsmessung gestaltet sein?

Maier (2010) kommt nach Analyse deutscher (z. B. Klafki, 1994; Achtenhagen & Baumert, 1999; Sacher, 2006; Winter, 2006; Hascher & Astleitner, 2007) als auch internationaler Literatur (z. B. Broadfoot, 1998; Black & Wiliam, 1998a, b; Dochy, 2001; Pellegrino Chudowsky

& Glaser, 2001; Assessment Reform Group, 2002; Broadfoot & Black, 2004; Birenbaum, Breuer, Cascallar, Dochy, Dori, Ridgway, 2006) zu dem Schluss: „Es gibt einen breiten Konsens darüber, dass die traditionellen Funktionen und Verfahren der schulischen Leistungsmessung mit den gesellschaftlichen Anforderungen an institutionelles Lernen sowie neuen Lern- und Bildungskonzepten nicht mehr kompatibel sind“ (S. 294). Mit traditionellen Verfahren sind in diesem Kontext im schulischen Setting übliche Klausuren und Tests zur Überprüfung von Schülerleistungen, aber auch statusdiagnostische Maßnahmen im Allgemeinen gemeint. Erstere sind eher informelle Verfahrensweisen, die eher einen Status als vielmehr den Lernverlauf erfassen sollen. Bezüglich der Entwicklung mathematischer Kompetenzen in Vor- und Grundschule resümieren auch Fritz und Ricken (2009), „dass die weitere Entwicklung eines Kindes aus aktuellen diagnostischen Befunden nur begrenzt vorhergesagt werden kann. Diagnostische Untersuchungen müssen also wiederholt durchgeführt werden, um den Entwicklungsverlauf zu beurteilen, und Aussagen über den Entwicklungsstand haben immer nur hypothetischen Charakter“ (S. 49). Müller und Hartmann (2009) weisen in diesem Zusammenhang auf die Inkompatibilität von traditionellen schulischen Bewertungsmaßnahmen und der Forderung von Fritz und Ricken hin. So ist das Heranziehen von z. B. Hausaufgaben oder Klassenarbeiten zur Leistungsbeurteilung, auch wenn regelmäßig eingesetzt, eher summativer Natur, da dabei oft verschiedene Schwierigkeits- und Anforderungsniveaus Anwendung finden. Neben der Eignung zur Erfassung der Lernentwicklung wird somit auch deren Güte angezweifelt. Einen validen Schluss auf die Leistungsentwicklung im Laufe des Lernprozesses lassen diese Maßnahmen damit nicht zu. Daher, so Maier (2010),

> *„ist es in besonderem Maße notwendig, über Verfahren und Methoden der pädagogischen Diagnostik nachzudenken, die mit den Unterrichtsroutinen der Lehrkräfte kompatibel sind, zu Leistungssteigerungen beitragen können und sich bewusst von*

> *summativen Leistungsmessungen (zentralen Tests, Klassenarbeiten, Abschlussprüfungen etc.) abgrenzen bzw. diese sinnvoll ergänzen“* (S. 304).

Maier (2010) sieht in dem Konzept der formativen Leistungsmessung ebenfalls eine konsensfähige Alternative zur summativen Leistungsbeurteilung anhand von statusdiagnostischen Verfahren.

Die Fokussierung auf prozessbegleitende Diagnostik geht konform zu Präventionsansätzen im schulischen Rahmen. Grundlage dabei ist, dass schulische Leistungen, speziell im Bereich der Mathematik, in Abhängigkeit von verschiedenen Kontextfaktoren variieren bzw. gezielt variiert werden können. So kann durch eine frühzeitige Identifikation von Kindern mit Schwierigkeiten beim Rechenerwerb sowie anschließender spezifischer und kontinuierlicher Förderung in Anlehnung an entwicklungspsychologische Modelle, dem Entstehen bzw. dem Ausweiten von Wissenslücken entgegengewirkt werden. Es ergibt sich ein großes Entwicklungspotential hinsichtlich einer normgerechten Leistungsentwicklung (2008a). Wie in Abschnitt 1.2 bereits beschrieben, setzt sich schulische Prävention laut Hartke (2005) aus verschiedenen Komponenten zusammen. Einerseits werden im Rahmen primärer Prävention zur Vermeidung schulischer Minderleistungen im Bereich Mathematik unspezifische Schritte für alle Kinder ergriffen, wie z. B. den verstärkten Einsatz differenzierender Maßnahmen im regulären Unterricht. Andererseits muss es im Sinne der sekundären Prävention frühzeitig einsetzende spezifische Maßnahmen für Schüler bestimmter Risikogruppen geben, bspw. den Einsatz spezifischer Förderprogramme zum gezielten Aufbau mathematischer Einsichten. Bezogen auf diese Vorstellungen zur schulischen Prävention von Rechenschwächen erweist sich die Diagnostik als ein zentraler Baustein. Dabei sind diagnostische Maßnahmen an verschiedenen Stellen relevant und müssen verschiedene Funktionen einnehmen. Diagnostik muss demnach als Ensemble unterschiedlicher Verfahren angesehen

werden, denn, so konstatieren Heller und Hany (2002): „Das für alle möglichen Verwendungszwecke optimale Messinstrument gibt es in der Schulleistungsdiagnostik ebenso wenig wie in anderen Messbereichen“ (S. 93). Moser Opitz Reusser, Moeri Müller, Anliker, Wittich und Freesemann (2010) schlagen eine Kombination standardisierter Tests zur Sichtung von Kindern mit Schwierigkeiten beim Rechenlernen (Screening) mit weiteren wenig bis nicht-standardisierten Verfahren zur Abschätzung von konkreten Kompetenzen und Schwierigkeiten eines Kindes vor (Abbildung 9).

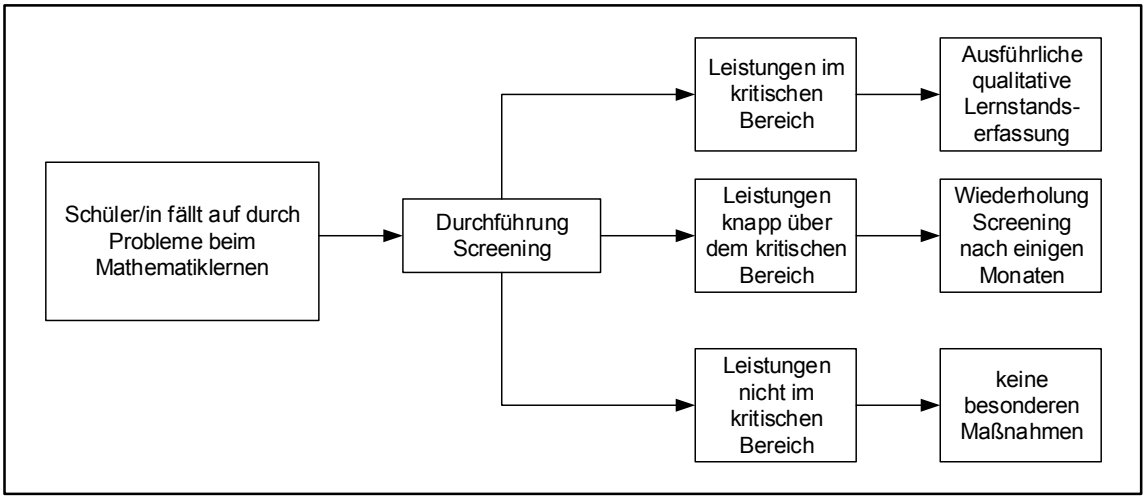


Abbildung 9: Zweistufiges Diagnoseverfahren nach Moser Opitz (entnommen aus Moser Opitz, 2009, S. 294)

Das US-amerikanische National Center on Response to Intervention (NCRTI, 2010) empfiehlt ebenfalls eine Kombination von Screenings zur Grobauslese potentiell gefährdeter Kinder und qualitativen Verfahren zur differenzierten Diagnose von auffällig gewordenen Kindern, erweitert dieses Vorgehen jedoch um eine fortlaufende Prozessüberwachung (Monitoring) im Sinne einer formativen Evaluation der Schülerleistungen. Im Konsens mit dem NCRTI (2010) erscheinen folgende drei Aspekte von Diagnostik innerhalb schulischer Prävention als wesentlich:

1. eine den Lernprozess fortlaufend begleitende Überwachung des Lernverlaufs aller Schüler im Sinne einer kontinuierlichen Evaluation der Unterrichts- bzw. Förderbemühungen.
2. der Einsatz regelmäßiger (zumindest einmal im Schuljahr) universaler Screeningverfahren mit allen Schülern zur Identifikation von Risikokindern.
3. der Einsatz von Methoden und Verfahren zur Inventarisierung von Stärken und Schwächen bei zuvor erkannten Risikokindern zur Ableitung von Verursachungshypothesen und Maßnahmen zur anschließenden Förderplanung.

Bei Vorgehensweisen entsprechend der Punkte 1. und 2. handelt es sich vorrangig um quantitative Evaluationen von Schülerdaten, welche sich den Maßnahmen primärer Prävention zuordnen lassen. Die unter Punkt 3. aufgeführten Maßnahmen sind eher qualitativer Natur und müssen als Teil sekundärer Prävention verstanden werden. Auf eine Auseinandersetzung mit qualitativen Verfahren entsprechend des Punkt 3. wird innerhalb dieser Arbeit verzichtet werden. Ausführungen zu Fehler- und Strategieanalysen sowie diagnostische Interviews finden sich u. a. bei Gerster (2009), Humbach (2008) Scherer und Moser Opitz (2010) sowie Wehrmann (2003). Besonders erwähnt seien an dieser Stelle die Bände der Diagnosereihe *Assessing math concepts* von Richardson (2002) für Kinder im Kindergartenalter bis zur dritten Klasse. In den neun Bänden werden detaillierte Diagnosestrategien und -blätter zu verschiedenen mathematischen Bereichen, wie bspw. Zählkompetenzen, Mengen- und Zahlenvergleiche oder Additions- und Subtraktionsrechnungen dargeboten. Diagnostische Maßnahmen der Punkte 1. und 2. werden innerhalb der vorliegenden Arbeit thematisiert. Dazu soll im folgenden Abschnitt zunächst allgemein das Konzept des formative assessment näher beleuchtet werden und anschließend differenziert auf curriculumbasierte

Messverfahren als eine mittlerweile bewährte Methode der Lernfortschrittsmessung eingegangen werden.

3.1 Das Konzept des formative assessment im Rahmen schulischer Prävention

Laut Duden bedeutet das Wort formativ so viel wie gestaltend. Im Zusammenhang mit diagnostischen Maßnahmen im schulischen Bereich wird der formativen Evaluation bzw. dem formative assessment damit ein wesentlicher Zweck zugesprochen: eine Leistungserfassung mit dem Ziel der Anpassung bzw. Modifikation laufender Förderbemühungen. Das Konzept des formative assessment ist kein Novum im pädagogischen Feld. Wie im Abschnitt zuvor bereits erwähnt, beschreibt Scriven schon im Jahr 1967 die wiederholte Leistungsmessung zum Zweck der Anpassung bzw. Modifikation von entwicklungs- bzw. leistungsförderlichen Maßnahmen im schulischen Setting als formative Evaluation. Damit definiert Scriven (1967) mit der formativen Evaluation den Gegenpol zur summativen Evaluation. In einem Artikel von Black und Wiliam aus dem Jahr 2003 wird die Entwicklung des Konzeptes in England differenziert aufgeschlüsselt. Demnach haben Maßnahmen des formative assessment über die 70er, 80er und 90er Jahre ein stetiges Auf und Ab hinsichtlich ihrer Nützlichkeit für das Schulwesen erlebt (Black & Wiliam, 2003). Jedoch erst zum Ende der 90er Jahre gewann das Thema erheblich an Bedeutung (Black & Wiliam 1998a; 1998b). Es soll innerhalb dieser Arbeit jedoch keine detaillierte Beschreibung des historischen Werdeganges des formative assessment erfolgen, sondern nur einige Meilensteine benannt werden. Interessierte Leser seien auf den Artikel von Black und Wiliam (2003) verwiesen.

Laut Black und Wiliam (1998b) zeichnet sich assessment und konkreter das formative assessment folgendermaßen aus:

> *„In this paper, the term 'assessment' refers to all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged. Such assessment becomes 'formative assessment' when the evidence is actually used to adapt the teaching work to meet the needs"* (S. 141).

Mit ihrer 1998 veröffentlichten Literatursynthese liefern Black und Wiliam (1998a; 1998b) neben einer Definition des formative assessment auch eine Zusammenschau von Studien, die Maßnahmen im Sinne des formative assessment thematisierten. Insgesamt, so halten die Autoren fest, wiesen die Untersuchungen auf eine hohe Effektivität solcher Maßnahmen für die pädagogische Arbeit im Schulwesen aus, wodurch deren Bedeutung weiterhin unterstrichen wurde. Auch in den USA ist das Konzept des formative assessment mittlerweile von großer Wichtigkeit. Der Council of Chief State School Officers (CCSSO, 2008a), eine unabhängige und länderübergreifende Organisation von Amtsleitern der US-amerikanischen Schulbehörden, welche bei bedeutenden Fragen der Bildung eine beratende Funktion einnimmt, startete eine Offensive, um den Einsatz formativer Verfahren im Unterricht gezielt zu verbreiten. In diesem Zusammenhang wurde innerhalb dieser Organisation eine Arbeitsgruppe gebildet, deren Hauptaufgabe die verstärkte Auseinandersetzung mit dem Thema *Formative Assessment for Students and Teachers* (FAST) darstellte. Innerhalb dieser Gruppe wurde 2007 eine allgemeingültige Definition für das Konzept des formative assessments publiziert, welche nun in den USA national anerkannt wurde (CCSSO, 2008a): „Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and

learning to improve students' achievement of intended instructional outcomes" (S. 3). Diese Definition beschreibt nicht ein spezifisches formatives Verfahren, sondern umfasst eine Bandbreite von Strategien, die Informationen über den Erfolg des Lernprozesses liefern, seien sie formeller (z. B. standardisierte Beobachtungen oder Tests) oder informeller (z. B. unstandardisierte Beobachtungen oder Interviews) Natur. Der Einsatz formativer Verfahren soll den Lernprozess durch regelmäßiges Feedback unterstützen und die Schülerleistungen in übergreifenden Lernstanderhebungen verbessern. Werden Kinder in solchen regelmäßig eingesetzten formativen Verfahren auffällig, bedeutet dies, dass sie spezifischere Unterstützungsmaßnahmen benötigen. Weist ein hoher Anteil der Schüler einer Klasse hingegen Schwierigkeiten auf, so kann das als Indikator für eine notwendige Umstrukturierung des ganzen Unterrichts gedeutet werden, wodurch ein Erreichen des grade levels aller Schüler ermöglicht werden soll (L.S. Fuchs & D. Fuchs, 2007).

Frey und Schmitt (2007) weisen darauf hin, dass der Begriff des formative assessments in vielen Publikationen unterschiedlich aufgefasst und verwendet wird. Im Folgenden wird die oben genannte Definition der CCSSO (2008a) zur Umschreibung des formative assessment verwendet, welche Maier (2010) treffend zusammenfasst: „Formativ ist somit kein Attribut der Leistungsmessung selbst, sondern der Funktion einer Leistungsmessung. Werden die Rückmeldungen genutzt, um den weiteren Lehr-Lernprozess zu informieren, handelt es sich um eine formative Leistungsmessung" (S. 299).

In den letzten Jahren hat der Ansatz des formative assessment in den USA an Bedeutung gewonnen, da dort alle Schulen als direkt verantwortlich für den Lernfortschritt ihrer Schüler gelten. D. h., dass anhand der Schülerleistungen unmittelbar Rückschlüsse auf die Qualität der pädagogischen Arbeit an einer Schule gezogen werden

(Spellings, 2005). Dies wird in einem Schreiben über die Schlüsselrichtlinien der Bildungsministerin Spellings an die Chief State School Officers vom 21.11.2005 dargestellt:

> „*The accountability model must ensure that all students in the tested grades are included in the assessment and accountability system. Schools and districts must be held accountable for the performance of student subgroups. The accountability model, applied statewide, must include all schools and districts"* (Spellings, 2005, core principle 4).

Diese Forderung geht einher mit dem No-Child-Left-Behind-Gesetz, welches im Jahre 2001 in den USA verabschiedet wurde. Demnach sind US-amerikanische Schulen dazu angehalten und dafür verantwortlich, dass zumindest 95 % aller Schüler die Normen in Lesen und in Mathematik des jeweiligen Bundesstaates (grade level) erreichen. In ihrem Schreiben erweitert Spellings diese Anforderung dahingehend, dass bis zum Schuljahr 2013/2014 sogar 100 % der Schüler dieses Ziel erreichen sollen. „The accountability model must ensure that all students are proficient by 2013-14 and set annual goals to ensure that the achievement gap is closing for all groups of students" (Spellings, 2005, core princinple 1). Dieses deklarierte Verantwortungsbewusstsein der Schulen ist Grundlage für die in den USA vorliegende bildungspolitische Situation. Dabei sollen regelmäßige, flächendeckende Testungen der schulischen Leistungen Aufschluss darüber geben, inwieweit es den Schulen gelingt, ihren Schülern zu den erwarteten Lernerfolgen zu verhelfen. Diese Art des regelmäßigen Leistungsmessens wird als *high-stakes testing* (high stake, zu Deutsch: hoher Einsatz) bezeichnet „because of the consequences to schools (and of course to students) that fail to maintain a steady increase in achievement across the subpopulations of the schools (i.e., minority, poor, and special education students)" (The Center for Public Education, 2006, o. S.). Underachievement, also das Nichterreichen der geforderten

Standards, bedeutet für betroffene Schüler das Nichtversetzen in höhere Klassenstufen. Darüber hinaus erhalten Schulen mit hohen Raten von sog. „underachievern“ erhebliche Sanktionen, u. a. die Kürzung oder Streichung finanzieller Mittel oder den Austausch von angestelltem Lehrpersonal. Derartige Sanktionen erhöhen das Bestreben der Schulen, ihrer Verantwortung möglichst nachzukommen. Dieses Vorgehen wird jedoch auch kritisch betrachtet (Amrein & Berliner, 2002; Nichols & Berliner, 2007; Noddings, 2004). So berichten Amrein und Berliner (2002), dass es durch den enormen Druck der den Schulen auferlegt ist, häufig zu Testmanipulationen durch bspw. Vorbereitung der Schüler auf die Messungen (*teaching to the test*) oder Ausschluss bestimmter Schüler von den Tests kommt. Weiterhin wird die Validität der dazu eingesetzten Verfahren angezweifelt, so dass Amrein und Berliner (2002) zu dem Schluss kommen: „Both the uncertainty associated with high-stakes testing data, and the questionable validity of high-stakes tests as indicators of the domains they are intended to reflect, suggest that this is a failed policy initiative“ (S. 58).

Um ein mögliches Scheitern von Schülern in den high-stakes tests möglichst frühzeitig zu erkennen und dem entgegenzuwirken zu können, hat sich das Interesse der Lehrkräfte an formativen Verfahren zur lernbegleitenden Erfassung von Schülerleistungen in den letzten Jahren deutlich verstärkt (CCSSO, 2008a).

3.1.1 Eigenschaften und Anforderungen des formative assessments

Die CCSSO (2008a) führt fünf wesentliche Aspekte auf, welche einen effektiven Einsatz des formative assessment ermöglichen sollen und sich wie folgt zusammenfassen lassen:

1. Durch das wiederholte Erheben von Schülerleistungen können der Verlauf des Lernprozesses verfolgt werden. Lernfortschritt, -abfall

bzw. -stagnation kann festgestellt werden und ein entsprechend zeitnahes Reagieren wird ermöglicht (CCSSO, 2008a). Unterrichtsanpassungen, bspw. die Wiederholung des Stoffes, das Schaffen weiterer Übungsmöglichkeiten oder die alternative Unterrichtsgestaltung, die aufgrund unzureichendem Lernzuwachs eingeleitet wurden, führen zur Steigerung der Schülerleistungen (Boston, 2002). Dabei können längerfristige Gesamtziele (z. B. das sichere Rechnen im Zahlenraum bis 20 zum Ende der Klasse eins) oder auch kurzfristigere Teilziele (z. B. das Verständnis für Menge-Zahl-Beziehungen) definiert und die Leistungsentwicklungen dahingehend analysiert werden.

2. Die festgelegten Ziele sollten mit den Schülern kommuniziert werden. Die Kinder sollten wissen, welche konkreten Ziele der Unterricht verfolgt werden und welche Kriterien den Erfolg der Lernentwicklung anzeigen (Black & Wiliam, 2009; CCSSO, 2008a; Harris, 2007). Durch das Einbeziehen der Schüler werden diese sensibilisiert, ihre eigenen Leistungen gezielt wahrzunehmen, zu analysieren und entsprechend darauf zu reagieren, wie bspw. verstärktes Lernen bei einem zu niedrigen Leistungsniveau (Bell & Cowie, 2001).
3. Dabei ist fortlaufendes Feedback über den eigenen Entwicklungsstand für die Schüler unerlässlich. Jedoch sollte dies nicht durch eine Bewertung im Sinne einer Notengebung geschehen. Vielmehr ist eine Rückmeldung deskriptiver Art sinnvoll (Black & Wiliam, 2009; Crooks, 1988; Harris, 2007). Feedback sollte also nicht einfach durch Anmerkungen der Art „Das hast du gut/schlecht gemacht“, sondern durch informative, konstruktive Hinweise an den Schüler gegeben werden, welche die individuellen Probleme aufzeigen und Aufschluss darüber geben, was getan werden kann, um die eigenen Leistungen weiter zu steigern. Dazu müssen die angelegten Bewertungskriterien anspruchsvoll, aber dennoch realistisch sein. Dies führt zur Stärkung der internalen Kausalattribution und des Fähigkeitsselbstkonzepts (Crooks,

1988). Darüber hinaus werden Schüler motiviert, sich intensiv mit dem Unterrichtsstoff auseinanderzusetzen, mit dem Ziel ihn zu durchdringen, anstelle nur einfach die richtige Antwort nennen zu wollen. Dies scheint insbesondere für schwache Lerner zuzutreffen (Bangert-Drowns, Kulik & Morgan, 1991; Elawar & Corno, 1985). Außerdem sollte die individuelle Bezugsnorm, also die eigene Leistung im Vergleich zu einem früheren Zeitpunkt, zur Beurteilung herangezogen werden, weniger der soziale Vergleich mit anderen Schülern. Die Rückmeldungen sollten zudem möglichst zeitnah gegeben werden und den Lernvoraussetzungen der Schüler angepasst sein (Crooks, 1988).

4. Während Schülerbeurteilungen normalerweise durch die Lehrkraft erfolgen, können Schüler durch Maßnahmen des formative assessment ihre Leistungen selbst evaluieren (Black & Wiliam, 2009; CCSSO, 2008a; Harris, 2007). Das angeregte metakognitive Überdenken des eigenen Lernens, führt zu der Einsicht, dass jeder selbstverantwortlich für seinen Lernfortschritt ist. Durch aktives Beobachten und Bewerten des eigenen Lernfortschritts werden persönliche Lernstrategien kritisch hinterfragt und ggf. optimiert. Fontana und Fernandes (1994) sowie Frederiksen und White (1996) haben in experimentellen Studien den positiven Effekt einer solchen Selbstevaluation der Schüler nachweisen können. Erstere Forschergruppe (1994) untersuchte die Mathematikleistungen von insgesamt 667 Grundschülern (Experimentalgruppe: N = 354; Kontrollgruppe: N = 313), Frederiksen und White (1996) untersuchten die Leistungen im Bereich Physik bei Schülern von 12 siebenten bis neunten Klassen. Dabei stellte sich heraus, dass Schüler mit klaren Vorstellungen ihrer zu erreichenden Lernziele sowie von den Kriterien zur Leistungsbeurteilung einen größeren Lernfortschritt aufwiesen als Schüler ohne solcherlei Anschauungen. McCurdy und Shapiro (1992) konnten in einer Studie mit 48 lernbehinderten Kindern im Grundschulalter zeigen, dass die Schüler die größten Leistungssteigerungen im lauten

Lesen aufwiesen, wenn sie ihren Lernverlauf selbst überwachten. Zu ähnlichen Resultaten führt die Studie von Sawyer, Graham und Harris (1992), welche eine Gruppe von 43 lernbehinderten Kindern der fünften und sechsten Klasse hinsichtlich ihrer Rechtschreibleistungen untersuchten.

5. Die zuvor benannten Punkte verdeutlichen insgesamt die Notwendigkeit der Kooperation von Lehrern und Schülern, um den Einsatz des formative assessment im Unterricht zu ermöglichen (CSSOO, 2008a; Harris, 2007). Speziell zur Realisierung einer unter Punkt 4. beschriebenen Selbstevaluation der Schüler ist eine besondere Struktur- und Strategievermittlung sowie Unterstützung durch die Lehrkraft notwendig.

Im nachfolgenden Abschnitt sollen Beispiele für das formative assessment näher erläutert werden.

3.1.2 Beispiele des formative assessment

Formative Leistungsbeurteilung in der Schule äußert sich in einer Vielzahl an Verfahrensweisen und Instrumenten. Dazu zählen die Durchführung von Tests, wie z. B. Lernstandserfassungen oder Rätselaufgaben, das Auswerten einzelner gestellter Aufgaben, Hausaufgaben sowie weitere unterrichtsinterne Beobachtungen oder gezielte Fragen und Unterrichtsgespräche, sofern die extrahierten Resultate zur Optimierung des Unterrichtsgeschehens genutzt werden (CCSSO, 2008b; Miller, Linn & Gronlund, 2008; McMillan, 2010). Alle genannten Möglichkeiten lassen sich dabei auf einem Kontinuum von informellen zu formellen Maßnahmen einordnen (Bell & Cowie, 2001). So können aus jeder Schüler-Lehrer-Interaktion, ganz gleich ob in Einzel-, Gruppen- oder Klassensituationen informelle Informationen gezogen werden (Bell & Cowie, 2001). Je planvoller und systematischer Schülerleistungsdaten erhoben werden, desto formaler kann das Vorgehen eingeschätzt werden, wie

z. B. bei Klassenarbeiten. Durch Standardisierung eines diagnostischen Verfahrens wird ein besonders hohes Maß an Formalität erreicht. Eine solche Fokussierung auf die Standardisierung eines Tests ist vor allem in der in Abschnitt 1.2 erwähnten summativen Diagnostik geläufig, welche sich im Gegensatz zur formativen Diagnostik auf punktuelle Leistungsmessungen bezieht, die nicht primär zur Unterrichtsmodifikation, sondern häufiger zur Selektion von Personen oder Bedingungen herangezogen werden. Aus Perspektive der pädagogischen Diagnostik ist eine Standardisierung und Orientierung an wissenschaftlichen Gütekriterien, wie sie summative Verfahren (zumindest oft) vorweisen können, essenziell. Auch bei der Fortschrittsdiagnostik durch formative Leistungsbeurteilungen sollte nicht auf die Güte der eingesetzten Verfahren verzichtet werden, da gemessene Ergebnisse sonst nicht klar interpretiert werden können. Scherer und Moser Opitz (2010) postulieren, dass jede professionelle Diagnostik, so auch eine formative Diagnostik, an Gütekriterien orientiert sein muss, jedoch werden diese nicht im Sinne der pädagogischen Diagnostik definiert: „Mit dieser Forderung ist nicht gemeint, dass die Gütekriterien der klassischen Testtheorie anzuwenden sind, sondern es geht darum, den Diagnoseprozess theoriegeleitet, transparent und intersubjektiv nachvollziehbar zu planen, durchzuführen und zu evaluieren“ (S. 37). Um das Dilemma zwischen summativen und formativen Vorgehensweisen aufzulösen, schlagen Black, Harrison, Lee, Marshall und Wiliam (2003) sowie Wininger (2005) als eine Möglichkeit vor, summative Verfahren wie zentrale Tests und Klassenarbeiten formativ einzusetzen (*formative use of summative assessments*). Die Ergebnisse solcher summativen Verfahren sollen laut den Autoren genutzt werden, um Unterrichtsadaptionen vorzunehmen, wodurch auch die Vorteile der ihnen inhärenten Standardisierung nutzbar werden. Ansätze, die formative und summative Ansätze miteinander verbinden, gewinnen

aktuell an Bedeutung im schulischen Setting (Drouin, 2010; Wininger, 2005).

3.1.3 Effekte des formative assessment

Bezüglich der Effektivität formativer Leistungsmessungen als unterrichtsinterne Maßnahme in Hinblick auf die Steigerung der schulischen Leistungen, der Motivation und des Fähigkeitsselbstkonzepts der Schüler liegen empirische Evidenzen durch zahlreiche Studien vor (Crooks, 1988; L.S. Fuchs & D. Fuchs, 1986; Natriello, 1987). Blacks und Wiliams Publikationen im Zeitraum von 1988 bis 1998 zum Thema formative assessment im schulischen Setting geben hierzu einen guten Überblick. Von 681 Beiträgen konnten durch die Autoren ca. 250 als relevante Publikationen eingestuft werden. Darunter findet sich neben den in Abschnitt 3.1.1 bereits beschriebenen Studien zur Effektivität von Selbstevaluationsmaßnahmen bei Schülern (Fontana & Fernandes, 1994; Frederiksen und White, 1996) auch eine Metaanalyse von L.S. Fuchs und D. Fuchs (1986), welche zunächst exemplarisch vorgestellt wird.

L.S. Fuchs und D. Fuchs (1986) analysierten 21 Untersuchungen mit Vorschulkindern bis hin zu Schülern der zwölften Klassenstufe. Das Hauptaugenmerk lag auf Schülern, die leichte Beeinträchtigungen des schulischen Lernens zeigten, es gab jedoch auch Studien, die ausschließlich die Leistungen normal entwickelter Kinder untersuchten Bei allen Studien sollte der Effekt des Feedbacks für und von Lehrern, welches durch zwei- bis fünfmal in der Woche stattfindende Leistungserhebungen ermittelt wurde, bezüglich der schulischen Leistungen der Kinder überprüft werden. Aus methodischer Sicht wurde dabei sichergestellt, dass diese Untersuchungen jeweils eine Experimental- als auch eine Kontrollgruppe einbezogen. Insgesamt wurden durchschnittliche

Effektstärken von d = 0.70 berichtet. Zwischen Studien, die Kinder mit Beeinträchtigungen und solchen, die ausschließlich normal entwickelte Kinder untersuchten, wurden keine signifikanten Unterschiede in den ermittelten Effektstärken festgestellt, obwohl der Wert bei Studien einschließlich Kindern mit Beeinträchtigungen etwas höher lag (d = 0.73 vs. d = 0.63). Die größten Effekte, so L.S. Fuchs und D. Fuchs (1986), wurden dann erzielt, wenn die Lehrkräfte Regeln zur Datenanalyse und den daraus resultierenden Konsequenzen mit den Schülern aufstellten (d = 0.92). Im Gegensatz dazu fiel die Effektstärke mit d = 0.42 in Gruppen, in denen die Lehrkraft auf ein solch differenziertes Feedback verzichtete, deutlich geringer aus. Studien, in welchen die regelmäßig erhobenen Leistungsdaten diagrammartig zusammengefasst wurden, erzielten durchschnittlich höhere Effektstärken (d = 0.70) im Gegensatz zu Untersuchungen, in denen solche Aufbereitung der Daten nicht stattfand (d = 0.26). Zusammengefasst lässt sich durch diese Metaanalyse festhalten, dass bereits durch den bloßen Einsatz formativer Leistungserhebungen zur Adaption des Unterrichts in der Schule ein positiver Effekt auf die Leistungen aller untersuchten Schüler festzustellen ist. Außerdem wird deutlich, dass vor allem leistungsschwächere Schüler von formativen Leistungsmessungen profitieren. Je differenzierter und systematischer die Daten sowie das daraus resultierende Feedback aufbereitet wurden, umso höher fielen auch die Effekte auf die Schülerleistungen aus (L.S. Fuchs & D. Fuchs, 1986).

Diese Ergebnisse spiegeln auch das Resultat der von Black und Wiliam (1998a, b) analysierten Untersuchungen wieder. Die von den Autoren zusammengetragenen Studien umfassen einen Querschnitt verschiedener Testbedingungen. So streut das Alter, der in den Studien untersuchten Probanden von fünf Jahren bis hin zum Universitäts- bzw. College-Alter. Es finden sich darüber hinaus Studien zu verschiedenen Fachbereichen, wie Mathematik, Lesen,

Schreiben oder Naturwissenschaften. Zudem umfassen die Untersuchungen Probanden verschiedener Länder und auch Schultypen. Die Ergebnisse dieser unterschiedlichen Studien zusammenfassend, kommen Black und Wiliam (1998a) zu dem Schluss, dass die formative Leistungserfassung im Sinne des Konzeptes des formative assessments zu Steigerungen schulischer Leistungen führt. Es zeigt sich ebenso, dass dies vor allem für leistungsschwächere Schüler gilt. „Many of them [den Studien] show that improved formative assessment helps the (so-called) low attainers more than the rest, and so reduces the spread of attainment whilst also raising it overall" (Black & Wiliam, 1998b, S. 3). Die Autoren berichten für die von ihnen gesichteten Studien durchschnittliche Effektstärken formativer Leistungsmessungen von d = 0.40 bis 0.70 (Black & Wiliam, 1998b). Auf welche Studien sich diese Werte konkret beziehen, wird nicht geklärt. Um einen praxisrelevanten Bezug herzustellen, beschreiben Black und Wiliam (1998b) die große Bedeutung einer im Kontext pädagogischer Maßnahmen als hoch anzusehende Effektstärke von d = 0.70 folgendermaßen: Leistungssteigerungen entsprechend einer Effektstärke von d = 0.70 würden England in der TIMS-Studie des Jahres 1996 (Beaton, Mullis, Martin, Gonzalez, Kelly & Smith, 1996) von einem damals mittleren Platz in die Liste der stärksten fünf der 41 teilnehmenden Länder manövrieren. Black und Wiliam (1998b) weisen darauf hin, dass Maßnahmen im Sinne des formative assessments somit als zentraler Baustein eines effektiven Unterrichts angesehen werden können: „this feature is at the heart of effective teaching" (Black & Wiliam, 1998b, S. 1).

Die von Black und Wiliams (1998a, b) vorgelegte Analyse ist jedoch nicht nur unkritisch betrachtet worden, obgleich sie seit der Veröffentlichung knapp 3000 Mal rezitiert wurde (Bennett, 2009; Kingston & Nash, 2011; Shepard, 2010). Shepard (2010) spricht von einer ritualisierten Ehrerbietung dahingehend, dass die der Analyse

zugrundeliegenden Studien in aktueller Literatur zum formative assessment nicht weiter differenziert betrachtet werden. “Attention has been focused on the positive effect sizes they reported, rather than the underlying theories that explain how formative assessment works” (S. 32). So ist der Analyse von Black und Wiliam (1998a, b) vorzuwerfen, dass sie Studien umfasste, die verschiedene Ansätze verfolgten, bspw. der Einfluss von mastery learning Konzepten, des Einsatzes curriculumbasierter Messverfahren, des Einsatzes verschiedener Formen des Feedbacks als auch des Einsatzes selbstevaluierender Maßnahmen (Kingston & Nash, 2011). Zwar hängen alle diese Ansätze mit dem Konzept des formative assessments zusammen, dennoch unterscheiden sie sich zum Teil grundlegend in ihrer Herangehensweise. Bei der Arbeit von Black und Wiliam (1998a, b) handelt es sich, Bennett (2009) zufolge, lediglich um eine qualitative Sichtung, nicht aber um eine quantitative Metaanalyse.

Im Jahr 2011 lieferten Kingston und Nash eine methodisch sorgfältig strukturierte, quantitative Metaanalyse. Dabei identifizierten die Autoren lediglich 13 von ursprünglich über 300 Untersuchungen, welche den zuvor fest definierten Auswahlkriterien (Einsatz formativer Messungen zur Steigerung des Leistungsniveaus der Schüler, Probanden vom Kindergartenalter bis Klassenstufe 12, Vorhandensein einer Kontroll- oder Vergleichsgruppe, Vorhandensein notwendiger Statistiken zur Berechnung der Effektstärke, Veröffentlichung nach 1988) entsprachen. Die errechneten Effektstärken dieser Studien liegen deutlich unter den Angaben von Black und Wiliam (durchschnittliche Effektstärken von d = 0.40 bis 0.70, 1998a, b). Für den Bereich Mathematik ergibt sich laut Berechnungen von Kingston und Nash (2011) eine mittlere Effektstärke von d = 0.17 mit einem Konfidenzintervall von 0.14 bis 0.20. Für den Sprach- und Literaturunterricht konnten durchschnittliche Effektstärken von d = 0.32 (Konfidenzintervall: 0.30

– 0.34.) und für Naturwissenschaften von d = 0.09 (Konfidenzintervall: −0.09 – 0.25) ermittelt werden. Der Einsatz computerbasierter Maßnahmen zur formativen Leistungserhebung der untersuchten Schülergruppen ergab eine Effektstärke von d = 0.28 mit einem Konfidenzintervall von 0.26 bis 0.30 (Kingston & Nash, 2011). Obgleich die ermittelten Effektstärken deutlich unter den in der Literatur angegebenen Werten liegen, so sind diese doch von großer praktischer Bedeutung für die pädagogische Arbeit in der Schule. Kingston und Nash (2011) kommen zu dem Ergebnis: „Results, though, do indicate formative assessment can be a significant and readily achievable source of improved student learning" (S. 33).

3.1.4 Ein Konzept zur strukturierten Umsetzung des formative assessment – Response-to-Instruction/Response-to-Intervention

In den USA ist die gemeinsame Beschulung starker und schwächerer Lerner mit der Verabschiedung des Education for All Handicapped Children Act von 1975 schon seit knapp 40 Jahren gängige Praxis. Seitdem wurde eine Überrepräsentation von Minderheitsgruppen in sonderpädagogischen Förderprogrammen und -einrichtungen festgestellt (Donovan & Cross, 2002). Überdies konnte seither über die Jahre hinweg ein kontinuierlicher Anstieg des Risiko-Index für Lernbehinderung festgestellt werden. Dabei zeigte sich, dass viele Kinder zu Unrecht als lernbehindert (falsch positiv) klassifiziert wurden (Jenkins & O'Connor, 2002; MacMillan, Siperstein & Gresham, 1998; Vaughn, Linan-Thompson & Hickman, 2003). „Sixty-nine percent of the children predicted to be at risk in Felton's (1992) study, for example, were good readers by third grade; only 58 % of Badian's (1994) at-risk preschoolers had confirmed reading problems" (Jenkins & O'Connor, 2000, S. 112). Diese Fakten gaben

Anlass dazu, die vorherrschenden Zuweisungskriterien zu sonderpädagogischen Maßnahmen in Frage zu stellen und nach neuen, passenderen Kriterien zu suchen, welche der überproportionalen Repräsentation von Minderheiten in Risikogruppen, sowie dem Anstieg des Risikoindexes für Lernbehinderungen Rechnung tragen sollten.

Abbildung 10 stellt die von Donovan und Cross (2002, S. 49) berichteten Daten zum Anstieg des Risikoindexes für Lernbehinderungen für die Jahre 1974 bis 1998 grafisch dar.

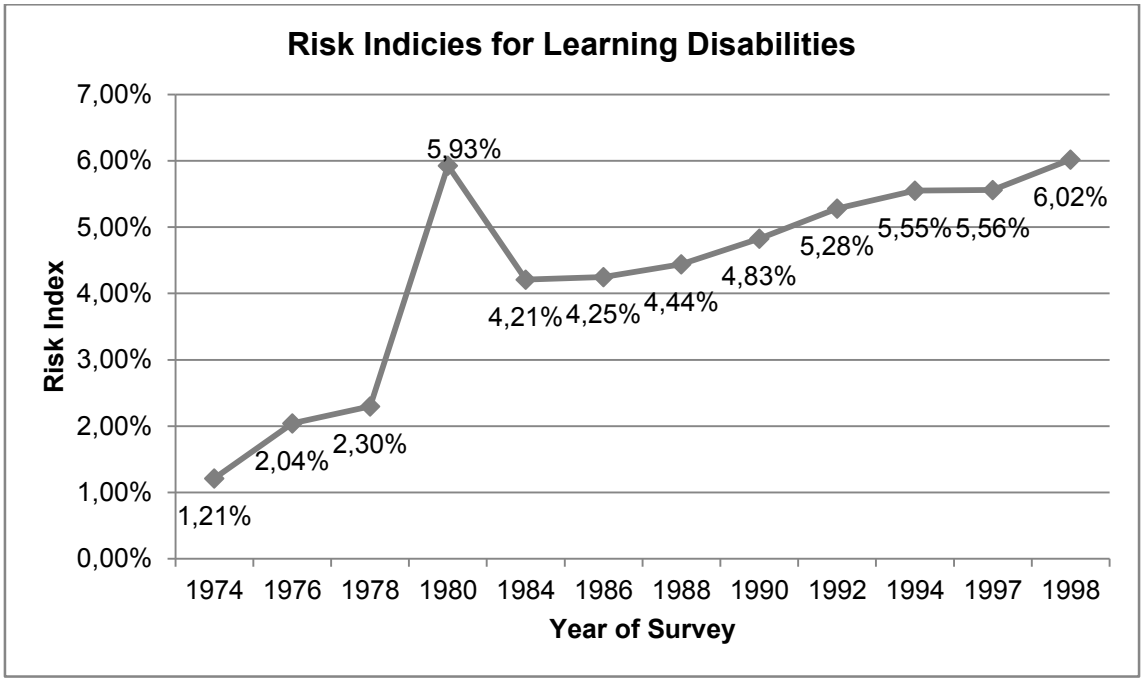


Abbildung 10: Risiko Indices für Lernbehinderung (learning disability) in den USA für die Jahre 1974 bis 1998 (Daten entnommen aus: Donovan & Cross, 2002, S. 49)

Der Grafik ist zu entnehmen, dass der Risikoindex über die 20 Jahre hinweg einen positiven Anstieg zeigt. „Twenty years later, disproportion in special education persists (Donovan & Cross, 2002, S. 1). Weiterhin kommen die Autoren zu dem Schluss, dass „while about 5 percent of Asian/Pacific Islander students are identified for special education, the rate for Hispanics is 11 percent, for whites 12

percent, for American Indians 13 percent, and for blacks over 14 percent" (S. 1f.).

Diese Ergebnisse mündeten in einem erneuten Überdenken der bis dahin geläufigen Zuweisungskriterien zu sonderpädagogischen Maßnahmen. Eine ganz zentrale Rolle spielt dabei das bis dato in den USA vorherrschende Diskrepanzkriterium (Reschly, Hosp & Schmied, 2003), nach dem eine Lernbehinderung erst dann vorliegt, wenn eine Diskrepanz zwischen kognitiven Leistungen und Schulleistungen diagnostiziert wird (siehe auch Abschnitt 2.10). Dieses Diskrepanzkriterium wurde in der Fachwelt von vielen Seiten kritisiert (Fletcher, Shaywitz, Shankweiler, Katz, Liberman, Fowler, Francis, Stuebing & Shaywitz, 1994; Fletcher, et al. 1998; Foorman, Francis, Fletcher & Lynn, 1996; Siegel, 1989a, b; K.E. Stanovich & Siegel, 1994; Vellutino, Scanlon & Lyon, 2000). Zur Argumentation wurden dabei vor allem folgende Gründe aufgeführt:

1. Aus methodologischer Sicht wird die Validität von Testverfahren zur Ermittlung des IQ angezweifelt. Wenn der IQ nicht valide erfasst werden kann, ist die gemessene Diskrepanz ebenfalls fehlerbehaftet (Vellutino et al., 2000).
2. Kognitive und schulische Leistungen sind nicht unabhängig voneinander, sondern bedingen sich. Die Diskrepanz beider Bereiche als Kriterium ist somit unzulässig (Siegel, 1989a, b).
3. Die Förderung setzt erst an, wenn Kinder signifikant schlechte Ergebnisse aufweisen und gescheitert sind (wait-to-fail), wodurch der Leistungsabstand zwischen starken und schwachen Lernern immer größer wird (Fletcher, et al. 1998; Vaughn & L.S. Fuchs, 2003).
4. Eine Unterscheidung der Gruppen, welche das Diskrepanzkriterium erfüllen und derer, die es nicht erfüllen, jedoch Schwierigkeiten aufweisen, ist nicht möglich (Fletcher et al., 1994; Foorman et al., 1996; K.E. Stanovich & Siegel, 1994).

Eine mögliche Antwort auf diese Aspekte soll der Ansatz *Response-to-Instruction* bzw. *Response-to-Intervention* (Antwort/Reaktion auf den Unterricht bzw. die Intervention, RtI) darstellen (u. a. L.S. Fuchs, 2003; L.S. Fuchs & D. Fuchs, 2007; Gersten & Dimino, 2006). RtI soll helfen, Kinder mit „echter" Lernbehinderung von denen zu trennen, die „lediglich" einer intensivierten pädagogischen Zuwendung bedürfen und somit der Stigmatisierung lernschwacher Schüler vorbeugen. Dabei wird auf die zur Klassifikation von Lernbehinderung vorherrschende Diskrepanzdefinition von IQ zur schulischen Leistung (siehe dazu auch Abschnitt 2.10) verzichtet. Hingegen wird angenommen, dass „responsiveness-to-intervention can differentiate between two explanations for low achievement: poor instruction versus disability" (L.S. Fuchs, 2003, S. 172). Somit werden Kinder erst dann als lernbehindert klassifiziert, wenn sie nicht die erwartete Reaktion bzw. Antwort (response) auf einen Unterricht zeigen (instruction), der so gestaltet ist, dass die überwiegende Mehrheit eine erfolgreiche schulische Leistungsentwicklung erreichen kann. Damit wird gleichzeitig ein Umdenken hinsichtlich der Selektion von Lernbehinderten in spezielle Förderprogramme bzw. –einrichtungen initiiert, denn in erster Linie wird der Unterricht bzw. die Lehrkraft oder die Schule für die Leistungsentwicklung eines jeden Schülers in Verantwortung gezogen. Beeinträchtigungen im Lernen werden immer im Zusammenhang zum Unterricht bzw. dessen Qualität gesehen. Eine Lernbehinderung kann demnach erst auftreten, wenn ein qualitativ hochwertiger Unterricht gesichert ist und ein Schüler dennoch nicht adäquat reagiert. Lern- bzw. konkreter auch Rechenschwächen können laut RtI-Definition nur als Resultat einer dualen Diskrepanz festgestellt werden, d. h. wenn Leistungsniveau *und* Lernfortschritt eines Schülers deutlich unter dem seiner Mitschüler liegt (L.S. Fuchs & D. Fuchs, 2007). „The idea is that students who respond poorly to otherwise effective instruction have a disability that limits their response to conventional instruction and instead require specialized treatment to affect schooling outcomes

associated with success in life" (L.S. Fuchs & D. Fuchs, 2007, S. 2). Mit dem Individuals with Disabilities Education Improvement Act (IDEA) im Jahr 2004 wurde RtI als eine alternative Möglichkeit zur Identifikation von Lernbehinderung offiziell anerkannt (Senate and House of Representatives of the United States of America, 2004).

Obwohl die Begriffe Response-to-Instruction und Response-to-Intervention als synonym betrachtet werden können, wird nachfolgend nur der Begriff Response-to-Instruction benutzt. Dies liegt in der präventiven Ausrichtung der vorliegenden Arbeit begründet, welche bezogen auf das schulische Setting eine besondere Akzentuierung des Unterrichts (instruction) als primäre Prävention vornimmt und Interventionen (intervention) als sekundäre bzw. tertiäre Prävention versteht.

Was aber verbirgt sich hinter RtI konkret? Mit Einführung des RtI-Konzeptes bestanden dahingehend diverse Unschärfen. So werden verschiedene Umsetzungen des RtI-Modells berichtet. Die implementierten Konzepte unterschieden sich hinsichtlich verschiedener Aspekte, wie z. B. in schulorganisatorischen Prozessen. Jedoch auch ganz grundlegende Fragen kamen auf, wie bspw. nach dem Zweck von RtI. Geht es dabei um die frühe Förderung im Sinne von Prävention? Oder um die Identifikation von Lernbehinderungen? Die Entwicklung zeigt, es gibt nicht das RtI-Konzept als solches, vielmehr bildet Response-to-Instruction einen konzeptuellen Rahmen, in dem verschiedene Komponenten zu einem präventiv ausgerichteten Beschulungsansatz miteinander verbunden werden. Das National Center on Response to Intervention (NCRTI, 2010) schlägt folgende Definition vor:

> „*Response to intervention integrates assessment and intervention within a multi-level prevention system to maximize student achievement and to reduce behavioral problems. With RTI, schools use data to identify students at risk for poor*

learning outcomes, monitor student progress, provide evidence-based interventions and adjust the intensity and nature of those interventions depending on a student's responsiveness, and identify students with learning disabilities or other disabilities" (S. 2).

Dabei werden vier wesentliche Bestandteile des RtI-Ansatzes benannt, die allen Konzepten eigen sind:

- „A school-wide, multi-level instructional and behavioral system tor preventing school failure
- Screening
- Progress Monitoring
- Data-based decision making tor instruction, movement within the multi-level system, and disability identification (in accordance with state law)" (NCRTI, 2010, S. 1).

In vorliegender Arbeit wird die vom NCRTI angegebene Gliederung wesentlicher Komponenten des RtI-Ansatzes modifiziert. Wie in Abbildung 1 (vgl. Abschnitt 1.2) aufgeführt, werden dazu der zweite und vierte Aspekt der Gliederung des NCRTI zu dem Punkt „datengeleitete Entscheidungsfindung" zusammengefasst und zusätzlich wird „evidenzbasierte Praxis" (evidence-based practice) als Extrapunkt hervorgehoben. Ein derartiges Vorgehen wird in den Ausführungen des NCRTI vorausgesetzt, erscheint aber im Rahmen schulischer Prävention hervorhebenswert. Die genaue Bedeutung der einzelnen Komponenten sowie deren Beziehungen zueinander werden in den nachfolgenden Abschnitten näher erläutert.

3.1.4.1 Mehrebenenprävention im RtI-Ansatz

Alle im RtI-Ansatz getroffenen Maßnahmen pädagogischen und sonderpädagogischen Charakters sind in Analogie zu Caplan (1964)

einer von drei Präventionsarten zuzuordnen: primäre, sekundäre und tertiäre Prävention (NCRTI, 2010). Alle schulischen Maßnahmen zur Unterstützung des Lernens sind dabei auf verschiedenen Ebenen organisiert und nehmen von Stufe zu Stufe in ihrer Intensität zu. Die Erhöhung der Intensität von Unterstützungsmaßnahmen kann dabei erreicht werden, indem Unterstützungsangebote

- häufiger
- verlängert
- in kleineren Gruppen
- durch höher qualifiziertes Personal

angeboten werden (D. Fuchs & L.S. Fuchs, 2006). Ein Kind, welches nicht in ausreichendem Maß Lernfortschritte zeigt, wird zusätzlich auf nächst höherer Stufe gefördert, solange bis sich Lernerfolge einstellen. Dabei gibt es keine feste Regelung zur Anzahl dieser Ebenen. So gibt es RtI-Ansätze mit zwei bis vier Ebenen (D. Fuchs et al., 2003), häufig werden jedoch drei Stufen definiert (L.S. Fuchs & D. Fuchs, 2007). In der vorliegenden Arbeit wird daher für die weiteren Ausführungen ein dreistufiges Modell zugrunde gelegt. Dabei gibt es keine Vorgaben zur inhaltlichen Ausgestaltung dieser Stufen, da dies von Schule zu Schule variiert (L.S. Fuchs & D. Fuchs, 2007).

Ebene 1 – Primäre Prävention

Als Maßnahme der primären Prävention auf Ebene 1 wird ein qualitativ hochwertiger Unterricht für alle Kinder angeboten, welcher den individuellen Anforderungen des Großteils der Kinder gerecht wird. Risikokinder werden durch zu Schuljahresbeginn eingesetzte Screeningverfahren identifiziert. Der Lernverlauf der Kinder wird durch regelmäßige, den Unterricht begleitende formative Leistungsmessungen überwacht (L.S. Fuchs & D. Fuchs, 2007).

Ebene 2 – Sekundäre Prävention

Die Ebene 2 richtet sich an identifizierte Risikokinder, also diejenigen Kinder, welche nicht in ausreichendem Maß von Maßnahmen der ersten Stufe profitieren konnten (Nichtresponder). Vaughn und Roberts (2007) gehen von einem Anteil von etwa 20 % bis 30 % von Kindern aus, die eine Förderung entsprechend der Ebene 2 bedürfen. Laut Reschly (2005), zeigt sich der Hauptunterschied zwischen den Förderebenen in „intervention intensity and measurement precision" (S. 511). Das bedeutet, dass diese Schüler einerseits zusätzliche Unterstützung in gefährdeten Teilbereichen (ca. 20-30min/täglich zusätzlich zur Förderebene 1), wie bspw. dem Lesen oder der Mathematik erhalten (intervention intensity), damit sich Schwierigkeiten nicht manifestieren (sekundäre Prävention). Auf Ebene 2 werden dazu auch evidenzbasierte Interventionsmaßnahmen eingesetzt. Die Förderung ist in der Regel in Kleingruppen von vier bis fünf Kindern organisiert (Vaughn & Roberts, 2007). Darüber hinaus werden weitere Diagnoseverfahren eingesetzt, um ein differenzierte Profile der Risikokinder zu erhalten (measurement precision), auf deren Grundlage Fördermaßnahmen adjustiert werden (NCRTI, 2010). Die Lernverlaufsdiagnostik wird höher frequentiert eingesetzt (bis zu zweimal wöchentlich), um zeitnahe Rückmeldung über Interventionseffektivität zu garantieren (L.S. Fuchs & D. Fuchs, 2007). „The goal is that students will 'catch up' with their peers after secondary intervention, which is typically implemented in 20- to 30-min sessions over a period of 10 to 20 weeks" (Vaughn & Roberts, 2007, S. 42). Zeigt ein Kind den erwünschten Lernfortschritt, können zusätzliche Maßnahmen der Förderebene 2 beendet werden und das Kind kann ausschließlich auf Förderebene 1 weiter unterrichtet (L.S. Fuchs & D. Fuchs, 2007).

Ebene 3 – Tertiäre Prävention

Kinder, bei denen die Maßnahmen der Ebene 2 keinen Effekt erzielen, erhalten auf Ebene 3 sonderpädagogische Förderung, um

dem Entstehen übergreifender Probleme vorzubeugen. Die Maßnahmen dieser Förderebene bewegen sich im Rahmen tertiärer Prävention. Grundlage der Förderarbeit bildet ein individuell abgestimmter Förderplan, in welchem Förderziele und -maßnahmen festgehalten sind. Diese Maßnahmen zeichnen sich durch eine hohe Systematik aus, deren Grundlage die erhobenen Schülerleistungen sind. Die Förderung findet etwa 50 Minuten täglich in Gruppen, deren Größe auf zwei bis drei Kinder begrenzt ist, statt. Oft wird Förderung auch im Einzelsetting angeboten (Vaughn & Roberts, 2007). Auf der Förderebene 3 wird der Lernverlauf weiter mittels regelmäßiger Testungen überwacht, deren Ergebnisse in die Entscheidung über Fortlauf der Förderung (Förderebenenzuweisung, Anpassung von Fördermaßnahmen) eingehen. Responder der Förderebene 3 bedürfen keine weiteren Maßnahmen dieser Ebene und werden auf Förderebene 1 und ggf. 2 weiter unterrichtet (L.S. Fuchs & D. Fuchs, 2007).

3.1.4.2 Datenbasierte Entscheidungsfindung im RtI-Ansatz

Die Entscheidung, auf welcher Förderebene ein Kind unterrichtet werden soll, wird auf Grundlage diagnostischer Maßnahmen getroffen. Dabei wird sich nicht auf ein einzelnes diskretes Ergebnis gestützt, sondern Ergebnisse verschiedener diagnostischer Maßnahmen werden aggregiert. Laut NCRTI (2010) gelten, wie schon in Kapitel 3 beschrieben, drei Aspekte von Diagnostik als wesentlich im RtI-Ansatz:

1. der Einsatz von Screeningverfahren (screening)
2. eine fortlaufend begleitende Überwachung des Lernprozesses (progress monitoring)
3. der Einsatz von Methoden und Verfahren zur Inventarisierung von Stärken und Schwächen bei zuvor erkannten Risikokindern

zur Ableitung von Verursachungshypothesen und Maßnahmen zur anschließenden Förderplanung.

Wo diese diagnostischen Maßnahmen im RtI-Konzept verortet sind und welchen Zweck sie dabei erfüllen, ist in Abbildung 11 schematisch dargestellt. Zunächst wird ein allgemeines bereichsspezifisches Screeningverfahren zu Schuljahresbeginn eingesetzt, an welchem alle Schüler teilnehmen. Durch ein solches Screening wird im Sinne der Ökonomie lediglich ein Risikostatus abgefragt, differenzierte Informationen liefert ein solches Verfahren nur begrenzt. Screenings müssen die Gütekriterien Objektivität, Reliabilität und Validität erfüllen (NCRTI, 2010). Kinder, die eine Leistung unterhalb eines festgelegten kritischen Wertes (Cutoff-Wert) erzielen, werden als Risikokinder eingestuft. D. Fuchs und Deshler (2007) geben bspw. als kritischen Wert bezüglich Leseleistungen ein Ergebnis entsprechend eines Prozentranges von unter 25 an. Diese Kinder mit Hinweisen auf Risiken werden zunächst auf Förderebene 1 weiter unterrichtet. Nach einer Zeit von etwa sechs bis zehn Wochen wird der Risikostatus des Kindes erneut überprüft. Dazu werden die Ergebnisse regelmäßig (wöchentlich bis zweiwöchentlich) eingesetzter Verfahren zur Dokumentation des Lernverlaufs herangezogen, mit deren Hilfe die individuelle Leistungsentwicklung abgeschätzt wird. Kinder, welche im Laufe der Zeit nicht den erwarteten Fortschritt zeigen, werden dann zusätzlich auf Förderebene 2 unterrichtet.

Durch zusätzliche diagnostische Maßnahmen werden in einem nächsten Schritt differenziertere Informationen über Stärken und Schwächen der identifizierten Risikokinder eingeholt, auf deren Basis ein umfassendes Schülerprofil für den betroffenen Bereich erstellt werden kann, dass als Grundlage für die weitere Förderarbeit dient (NCRTI, 2010). Die Effektivität der Förderbemühungen wird durch fortlaufende, regelmäßige Lernverlaufsmessungen quantifiziert,

welche innerhalb der Förderebene 2 mit höherer zeitlicher Frequenz eingesetzt werden. Dadurch ist eine zeitnahe Rückmeldung über den Erfolg der Maßnahmen gegeben. Auf Basis der erhobenen Daten wird der weitere Verlauf der Förderung bestimmt. Kinder, die sich als responsiv erweisen, werden lediglich auf Förderebene 1 weiter unterrichtet, Schüler, welche keine Leistungssteigerungen im erwarteten Ausmaß zeigen oder deren Leistungen einem sehr niedrigen Niveau entsprechen, erhalten zusätzliche Unterstützungsangebote auf Förderebene 3 (L.S. Fuchs & D. Fuchs, 2007).

Gemäß Reschly (2005) werden Maßnahmen der Diagnostik innerhalb der Förderebene 3 weiter präzisiert. Ziel ist es, einen umfassenden individuellen Förderplan zu generieren, der die innerhalb der ersten beiden Förderebenen identifizierten Problembereiche eines Schülers aufgreift. Dazu werden, analog zur Förderebene 2, hoch frequentierte (wöchentliche) lernverlaufsdiagnostische Maßnahmen durchgeführt. Die Ergebnisse werden genutzt, um Hinweise abzuleiten, wie Förderung weiter gestaltet sein muss und was realistisch erreichbare Ziele sind. Wie in den unteren Ebenen, dienen die Ergebnisse darüber hinaus der Einschätzung des Lernerfolges eines Schülers und stellen somit die Basis für weitere Förderebenenentscheidungen dar (L.S. Fuchs & D. Fuchs, 2007). Auf Ebene 3 erfolgreich lernende Kinder benötigen keine weitere derart intensive Förderung und werden auf Ebene 1 bzw. 2 weiter unterrichtet. Nichtresponder durchlaufen einen erneuten Förderdurchgang auf Förderebene 3. Innerhalb der Förderebene 3 wird, je nach Art und Ausprägung des Problemfeldes eines Schülers, eine Klassifikation in „Kind mit einer Lernbehinderung“ (learning disabilitiy, LD), „Kind mit Auffälligkeiten in der emotional-sozialen Entwicklung“ (emotional/behavioral disorders, EBD) oder „Kind mit geistiger Behinderung“ (mental retardation, MR) vorgenommen (L.S. Fuchs & D. Fuchs, 2007).

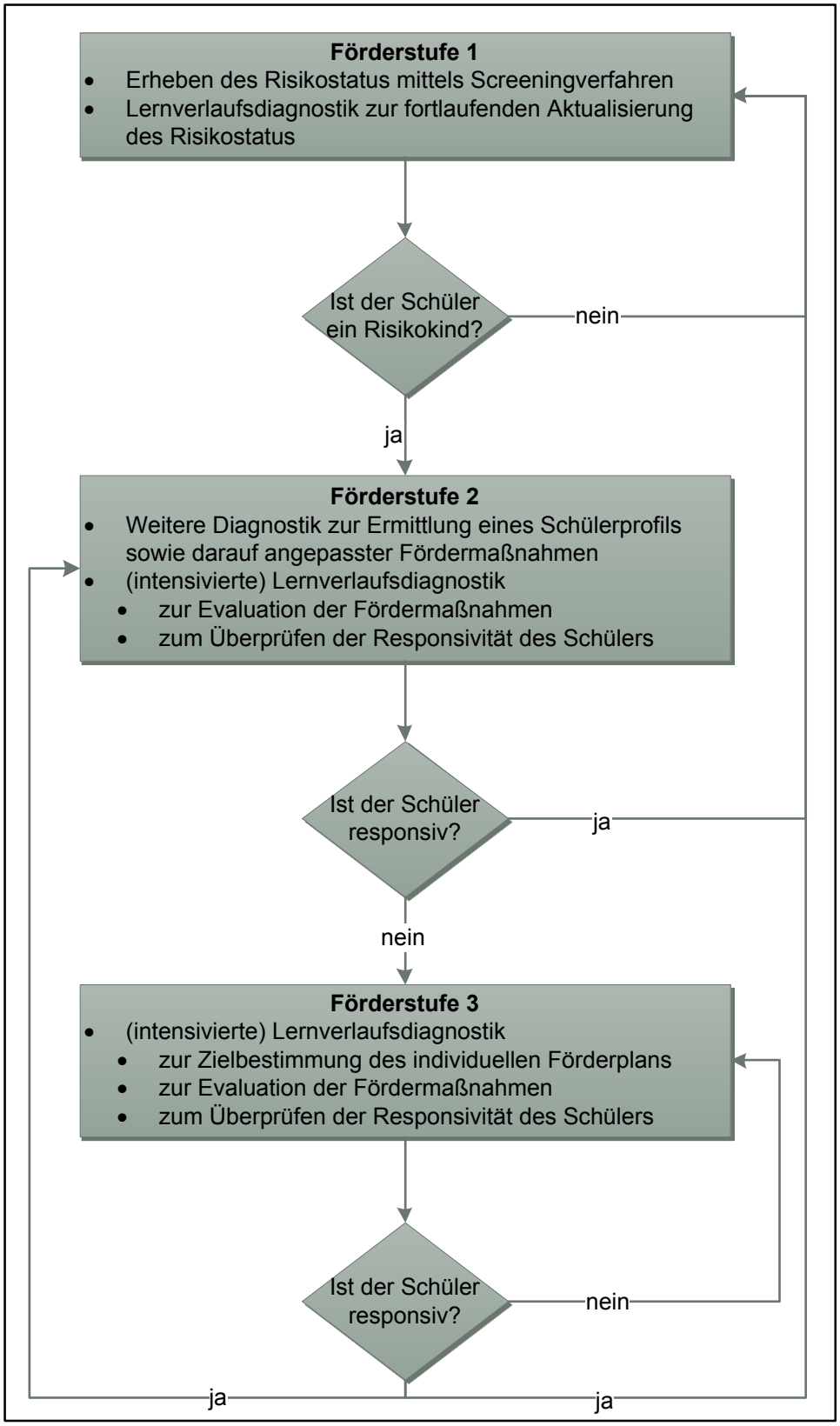


Abbildung 11: Flussdiagramm zur schematischen Darstellung des Einsatzes diagnostischer Maßnahmen im RtI in Anlehnung an L.S. Fuchs und D. Fuchs (2007) sowie NCRTI (2010) (entnommen und modifiziert aus: L.S. Fuchs & D. Fuchs, 2007, S. 26)

Bei den hier genannten Methoden der Lernverlaufsdiagnostik handelt es sich um curriculumbasierte Messverfahren (CBM), welche im Sinne des formative assessment eingesetzt werden, jedoch formale Kriterien wie Objektivität, Reliabilität sowie Validität erfüllen müssen (Deno, 1985). Curriculumbasierte Messverfahren werden in Abschnitt 3.2 als wesentlicher Aspekt dieser Arbeit differenziert dargestellt.

3.1.4.3 Evidenzbasierte Praxis im RtI-Ansatz

Um das erfolgreiche Lernen der Schüler sicherzustellen, sollen innerhalb eines RtI-Konzeptes eingesetzte Maßnahmen, Verfahren und Programme, gleich welcher Ebene zugeordnet, evidenzbasiert sein. Das bedeutet, dass alle einsetzenden Interventionen sorgfältig hinsichtlich ihrer Effektivität ausgewählt werden müssen. Im Sinne einer Qualitätssicherung müssen zur Einschätzung der Effektivität Beweise bzw. Nachweise (evidence) vorliegen. Die Fokussierung auf die Notwendigkeit von Maßnahmen, welche nachweislich als wirksam eingestuft werden können, entstammt der Medizin und wird dort folgendermaßen definiert:

> „*Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research"* (Sackett, Rosenberg, Muir Gray, Haynes & Richardson, 1996, S. 71).

Um eine Einschätzung der Güte einer Maßnahme zu ermöglichen, wurden Klassifikationssysteme erarbeitet, welche nach Validitätskriterien hierarchisch geordnete Evidenzebenen systematisieren (Deutsches Cochrane Zentrum, 2011; Oxford Centre for Evidence-based Medicine, 2009). In Tabelle 2 ist das Klassifikationssystem des Deutschen Crochane Zentrums

dargestellt. Dabei entsprechen Maßnahmen der Stufe I dem höchsten Evidenzniveau.

Tabelle 2: Klassifikationssystem zur Einschätzung von Evidenzen gemäß dem Deutschen Crochane Zentrum (2011, o. S.)

Stufe	Evidenz-Typ
Ia	wenigstens ein systematischer Review auf der Basis methodisch hochwertiger kontrollierter, randomisierter Studien (RCTs)
Ib	wenigstens ein ausreichend großer, methodisch hochwertiger RCT
IIa	wenigstens eine hochwertige Studie ohne Randomisierung
IIb	wenigstens eine hochwertige Studie eines anderen Typs quasi-experimenteller Studien
III	mehr als eine methodisch hochwertige nichtexperimentelle Studie
IV	Meinungen und Überzeugungen von angesehenen Autoritäten (aus klinischer Erfahrung); Expertenkommissionen; beschreibende Studien

Ein solches beweis- bzw. nachweisgestütztes Arbeiten ist mittlerweile auch im Handlungsfeld der Psychotherapie weitgehend etabliert, weniger jedoch im Bereich der Pädagogik bzw. Sonderpädagogik. In den vergangenen 10 Jahren nehmen die Forderungen nach evidenzbasiertem Vorgehen jedoch auch im pädagogischen Feld zu (P.J. Stanovich & K.E. Stanovich, 2003). Speziell in den USA wird mit dem No Child Left Behind (NCLB) Act ab dem Jahr 2001 der verstärkte Einsatz forschungsbasierter Unterrichtsmaßnahmen (research-based instruction) von Seiten der Regierung vorgeschlagen, speziell auch für Kinder mit Hinweisen auf Entwicklungsrisiken. Darunter sind Methoden zu verstehen, welche unter strengen Standards auf Effektivität hin geprüft wurden (Simpson, LaCava & Sampson Graner, 2004), die Begriffe forschungsbasiert und evidenzbasiert können somit, zumindest im Feld der Pädagogik, als synonym angesehen werden. Um den Kriterien des NCLB Act zu genügen, müssen folgende Punkte gegeben sein:

- „employ systematic, empirical methods that draw on observation or experiment;

- involve rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions;
- rely on measurements or observational methods that provide valid data across evaluators and observers, and across multiple measurements and observations; and
- be accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparatively rigorous, objective, and scientific review" (National Institute for Literacy, 2006, S. 1).

Um solche Standards im schulischen Bereich zu erreichen, sollte auf Anraten der Coalition for Evidence-Based Policy (2002) der Einsatz evidenzbasierter Verfahren in der Schule durch die Regierung finanziell unterstützt werden. Des Weiteren schlägt die Coalition for Evidence-Based Policy (2002) vor, eine Datenbank pädagogischer Interventionsmaßnahmen aufzubauen, welche durch randomisierte, kontrollierte Studien als effektiv eingestuft wurden und diese durch öffentlichen Zugang zu verbreiten. In diesem Zusammenhang sind in den USA diverse Websites der Regierung oder weiteren freien Trägern zur Verbreitung evidenzbasierter Maßnahmen und zur Unterstützung bei deren Umsetzung entstanden, welche auf Evidenz geprüfte Vorgehensweisen und Programme zu verschiedenen Handlungsfeldern (schulisches Lernen, Verhaltensauffälligkeiten, Kriminalität, Drogen- bzw. Alkoholkonsum etc.) aufführen, wie bspw.:

- http://www.whatworks.ed.gov
- http://www.findyouthinfo.gov
- http://www.evidencebasedprograms.org
- http://www.bestevidence.org

In der deutschen Pädagogik bzw. Sonderpädagogik ist ebenfalls eine vermehrte besondere Akzentuierung auf empirische Evidenz von Unterrichts- und Fördermaßnahmen und -programmen festzustellen.

So sind in der sonderpädagogischen Fachliteratur systematische Kategorisierungen von Kriterien zur Einschätzung der wissenschaftlichen Evidenz von Unterrichts- und Förderprogrammen vorgenommen worden (Fingerle & Ellinger, 2008; Hartke, 2005; Koch, 2008; Nußbeck, 2007). Exemplarisch sind die Überlegungen von Hartke (2005) sowie Fingerle und Ellinger (2008) in Tabelle 3 aufgeführt. In beiden Publikationen wird eine Systematisierung in drei Kategorien vorgenommen, wobei die Erfüllung der Kriterien in Kategorie 1 den höchsten Bewährungsgrad bzw. den bewährtesten Ansatz belegen. Erfüllt ein Programm bzw. eine Maßnahme die Kriterien der Kategorie 2, kann nur von einer vermuteten Effektivität ausgegangen werden (Fingerle & Ellinger, 2008), welche nur bedingt empfehlenswert ist (Hartke, 2005). Während die von den Autoren vorgeschlagenen Unterteilungen in Kategorie 1 und 2 noch recht ähnlich definiert sind, gibt es wesentliche Unterschiede in der Kategorie 3. So werden die Kriterien der Kategorie 3 bei Hartke (2005) so gefasst, dass ein Erfüllen dieser auf keine bzw. gegenteilige Effekte einer Maßnahme oder eines Programms hinweisen würde. Bei Fingerle und Ellinger (2008) hingegen weisen die Merkmale innerhalb der Kategorie 3 noch auf einige, jedoch wenige wissenschaftlich fundierte Evidenzen hin, wodurch nur eine potentielle Effektivität angenommen werden kann. Die von Fingerle und Ellinger (2008) beschriebenen Bewertungskriterien fallen im Vergleich zu den von Hartke postulierten Bewährungsgraden strenger aus, sodass eine Anwendung zu unterschiedlichen Einschätzungen führen könnte.

Tabelle 3: Vorgeschlagene Kategorisierung von Evidenzkriterien für (sonder-)pädagogische Förderprogramme nach Hartke (2005) sowie Fingerle und Ellinger (2008) (entnommen aus: Hartke, 2005, S. 17; Fingerle & Ellinger, 2008, S 9f.)

	Bewährungsgrade nach Hartke (2005)	**Evidenzkriterien nach Fingerle & Ellinger (2008)**
Kategorie 1	*Ausreichender Bewährungsgrad – empfehlenswert* • Maßnahme durch eine wissenschaftliche Theorie begründet und • Wirksamkeit der Maßnahme in der Praxis durch mehrere empirische Studien belegt	*„Bewährte Ansätze“* • Förderansatz durch eine wissenschaftliche Theorie begründet • Wirksamkeit des Ansatzes durch mehrere empirische Studien belegt • Ansätze müssen in randomisierten Kontrollgruppen-Designs überprüft worden sein • die Ansätze sollen in schulischen Settings überprüft worden sein • vor dem Beginn der Intervention sollte es keine bedeutsamen Unterschiede in den Randvariablen gegeben haben • es werden ausreichend gute Effektstärken berichtet • es werden Angaben zu Gruppenkennwerten und Signifikanzen, zu Design und Methoden gemacht, die den Standards wissenschaftlicher Publikationen entsprechen • es wurden valide Outcome-Maße verwendet bzw. die quantitativen Maße werden durch qualitative Daten ergänzt, welche die Nachteile der quantitativen Maße kompensieren • idealerweise sollte auch eine Follow-up-Erhebung stattgefunden haben • die Ergebnisse müssen in der Fachliteratur publiziert worden sein

	Bewährungsgrade nach Hartke (2005)	Evidenzkriterien nach Fingerle & Ellinger (2008)
Kategorie 2	*Knapp ausreichender Bewährungsgrad – bedingt empfehlenswert* • Maßnahme durch eine wissenschaftliche Theorie begründet und • erste Hinweise in empirischen Studien auf eine erfolgreiche Implementation in die Praxis oder auf Wirksamkeit in der Praxis	*„Vermutlich effektive Ansätze“* • Förderansatz durch eine wissenschaftliche Theorie begründet • Wirksamkeit des Ansatzes durch mehrere empirische Studien belegt • Ansätze müssen in Kontrollgruppen-Designs überprüft worden sein • es werden ausreichend gute Effektstärken berichtet • es werden Angaben zu Gruppenkennwerten und Signifikanzen, zu Design und Methoden gemacht, die den Standards wissenschaftlicher Publikationen entsprechen • es wurden valide Outcome-Maße verwendet bzw. die quantitativen Maße werden durch qualitative Daten ergänzt, welche die Nachteile der quantitativen Maße kompensieren • die Ergebnisse müssen in der Fachliteratur publiziert worden sein
Kategorie 3	*Nicht ausreichender Bewährungsgrad – nicht empfehlenswert* • bisher keine empirischen Hinweise auf Implementationschancen oder Wirksamkeit trotz theoretischer Begründung oder • deutliche empirische Hinweise auf problematische Effekte oder • Praktiker berichten über Erfolge ohne ausreichend theoretische oder empirische Begründung, also ohne Hinweise auf eine allgemeine Gültigkeit	*„Potentiell effektiv“* • Förderansatz durch eine wissenschaftliche Theorie begründet • Wirksamkeit des Ansatzes durch eine empirische Studien belegt • dies können auch mehrere qualitative Fallstudien sein • Ansätze müssen in Kontrollgruppen-Designs überprüft worden sein • ausreichend gute Effektstärken (bzw. bei qualitativen Fallstudien nachvollziehbare Erfolgskriterien) • es werden Angaben zu Gruppenkennwerten und Signifikanzen, zu Design und Methoden gemacht, die den Standards wissenschaftlicher Publikationen entsprechen (bzw. bei qualitativen Fallstudien nachvollziehbare Beschreibungen der Methode, des Samplings, der Erhebung und Auswertung) • valide Outcome-Maße • die Ergebnisse müssen in der Fachliteratur publiziert worden sein

3.1.4.4 Bedeutung und Effekte des RtI-Ansatzes

Durch die präventive Ausrichtung des Response-to-Instruction-Ansatzes, werden alle Kinder durch qualitativ hochwertige Unterrichts- und Fördermaßnahmen zeitnah spezifisch gefördert – ein wait-to-fail wird somit vermieden. Durch regelmäßig wiederholte Leistungserfassungen der Schüler zur Evaluation und Optimierung des Unterrichts werden weitreichende Prognosen von Schulleistungen vermieden. Ziel dieses Ansatzes ist, die Kluft zwischen schwachen und stärkeren Lernern zu schließen. Dabei werden Konzepte und Ansätze systematisch mit einander verbunden, die sich zuvor theoretisch und zum großen Teil empirisch bewährt haben:

- der Einsatz früher spezifischer Hilfen zur Prävention von manifesten Minderleistungen (Hinweise u. a. bei Aunola et al., 2004; Gaupp et al., 2004; Geary et al., 2000; Helmke & Weinert, 1997; Krajewski, 2003; Krajewski & Schneider, 2006; Kurdek & Sinclair, 2001; Mazzocco et al., 2011; Mazzocco & Thompson, 2005; Stern, 2003; Weißhaupt et al., 2006)
- der Einsatz regelmäßiger Leistungserhebungen, deren Resultate als Feedback zur Förderung für Lehrer und Schüler genutzt werden, im Sinne des formative assessment (u. a. Black & Wiliam, 1998a, b; L.S. Fuchs & D. Fuchs, 1986; Kingston & Nash, 2011)
- Kooperation von verschiedenem schulischen Personal (Lehrkraft, Sonderpädagoge und Schulpsychologe) bei der Förderplanung und -entscheidung (u. a. Burns & Symington, 2002; Kovaleski & Pedersen, 2008; Tilly, 2008)
- der Einsatz evidenzbasierter Unterrichts- und Fördermaßnahmen (u. a. Shapiro, 2004; Shinn, Walker & Stoner, 2006; Swanson, 1999, 2000; Swanson, Hoskyn & Lee, 1999)

Das lässt augenscheinlich den Schluss zu, dass es sich bei dem Response-to-Instruction-Ansatz, verstanden als Verknüpfung von Maßnahmen im Sinne des formative assessment sowie darauf abgestimmter evidenzbasierter Maßnahmen, organisiert auf verschiedenen nach Intensität gestuften Förderebenen, um ein sinnvoll strukturiertes System handelt, welches in der Lage ist, schwachen Lernern eine positive schulische Lernentwicklung zu ermöglichen. Doch reicht eine solche Annahme aus, um eine Implementation dieses Konzeptes in der Schule zu rechtfertigen? Letztlich bleibt die Frage offen, ob RtI als Komposition verschiedener vielversprechender Komponenten sich genauso wirksam zeigt wie seine einzelnen Teile. Die Forschung hat sich bis dato recht intensiv mit der Evaluation der Einzelkomponenten befasst, jedoch wenig mit den im RtI systematisierten Prozessen verstanden als Ganzes (VanDerHeyden, Witt & Gilbertson, 2007). Dies ist sicherlich auch durch die uneinheitliche Konzeptualisierung des RtI-Gerüsts (Abschnitt 3.1.4) und damit verbundenen methodischen Schwierigkeiten bei der Evaluation verschuldet.

In einer Meta-Analyse von Burns, Appleton & Stehouwer (2005) wurden die Ergebnisse von Studien zur Effektivität eines flächendeckenden Einsatzes eines RtI-Konzeptes zusammengetragen. Das Resultat einer umfassenden Literaturanalyse waren 31 publizierte Artikel anhand derer 24 verschiedene relevante Studien ermittelt werden konnten. Acht der Studien beschrieben dabei RtI-Konzepte, welche an Universitäten zu Forschungszwecken konzipiert wurden (universitäre Umsetzung). Bei den anderen 16 Studien handelte es sich um bereits implementierte und etablierte Umsetzungen von RtI-Ansätzen im schulischen Setting (schulische Umsetzung). Dabei waren Studien vertreten, welche sich auf gemessene Schülerleistungen (students outcome) bezogen und solche, die systemische Daten, bspw. die Anzahl von Kindern mit festgestellter Lernbehinderung fokussierten (systemic outcome).

Burns et al. (2005) kommen zu dem Schluss, dass alle Studien einen großen Effekt auf das schulische Lernen zeigen konnten.

> „*RTI models currently implemented in the field (n = 11) led to a mean ES* [Effektstärke] *of 1.73 (SD = 0.99) for systemic variables but resulted in a mean ES of 0.62 (SD = 0.33) for student outcomes (n = 5). A reverse relationship was noted for RTI models implemented by university faculty for research, with a mean ES of .47 (SD = .07) for systemic outcomes (n = 2) and 1.23 (SD = .95) for student outcomes (n = 6)"* (Burns et al., 2005, S. 387).

Von Burns et al. (2005) errechnete korrigierte Schätzungen der Effektstärken ergaben Werte von d = 0.94 (schulische Umsetzung) und d = 1.14 (universitäre Umsetzung) auf die Schülerleistungen und Werte von d = 1.80 (schulische Umsetzung) und d = 0.47 (universitäre Umsetzung) der systemic outcomes.

In einer longitudinal angelegten Studie zur Wirksamkeit eines RtI-Ansatzes (Simmons, Coyne, Kwok, McDonagh, Harn & Kame'enui, 2008) wurde die Entwicklung der Lesekompetenzen von 41 im Kindergartenalter als riskobelastet eingestuften Kinder (Cuttoff-Wert bei einem Prozentrang unter 30 in standardisierten Lesetests) bis zum Ende der dritten Klassenstufe verfolgt. Alle Probanden wurden intensiv in Kleingruppen gefördert. Zu Beginn jeder Klassenstufe wurde der Risikostatus der Kinder mittels standardisierter Screenings aktualisiert. Entsprach ein Ergebnis einer Leistung unterhalb des 30sten Perzentils, wurde die zusätzliche Förderung weiter fortgesetzt. Erzielte ein Kind bessere Leistungen, bedurfte es keiner weiteren Förderung zu diesem Zeitpunkt. Es zeigte sich, dass die Mehrheit der Kinder schon zum Ende des Kindergartens nicht mehr der Risikogruppe angehörte. Nach dem dritten Schuljahr erzielten 38 der 41 (93 %) ehemals identifizierten Riskokinder Resultate entsprechend einer Leistung oberhalb des 30sten Perzentils, d. h. Ergebnisse, welche nicht mehr dem Risikokriterium entsprachen. 37

der Kinder (90.2 %) zeigten über die Zeit von Anfang erster bis Ende dritter Klasse unauffällige Leistungen in einem Subtests, bei welchem Kunstwörter vorgelesen werden mussten, in einem Subtest, in welchem reale Wörter vorgelesen werden sollten, waren es 32 Kinder (78 %).

Resümierend kann festgehalten werden, dass durch eine Verbindung formativer Evaluation der Leseleistungen und anschließender entsprechender Förderung im Rahmen eines RtI-Ansatzes die Leseleistungen der untersuchten Kinder bis zum Ende der dritten Klasse normalisiert werden konnten (Simmons et al., 2008). Ähnliche Ergebnisse konnten auch Vellutino, Scanlon, Small & Fanuele (2006) in ihrer Studie mit 1373 Kindern nachweisen. Sie untersuchten die Wirksamkeit von in einem RtI-Ansatz geförderten Lesekompetenzen vom Kindergarten bis zum Ende der dritten Klasse.

Die hier präsentierten Daten stimmen zuversichtlich und so kommt auch Burns (2010) zu dem Schluss: „RTI may be a collection of parts accumulated and pieced together over decades of research and practice, but the result of this compilation of parts is a sum that equals positive outcomes for kids" (S. 14). Dennoch muss die Effektivität von Response-to-Instruction-Ansätzen in der Schule durch weitere systematische Forschung gestützt werden.

An der Universität Rostock wird derzeit ein inklusives Beschulungskonzept für die Grundschule, angelehnt an den US-amerikanischen Response-to-Instruction-Ansatz, implementiert und evaluiert: Das Rügener Inklusionsmodell (RIM; Mahlau et al., 2011). Dabei handelt es sich um die erste und zu diesem Zeitpunkt einzige flächendeckende Adaption des RtI-Ansatzes auf deutsche Verhältnisse, innerhalb welcher Kinder mit Förderbedarf in den Bereichen Lernen, Sprache sowie Emotionale und soziale Entwicklung präventiv gefördert und inklusiv beschult werden. In

einem Zwischenbericht (Voß, Blumenthal, Diehl, Ehlers, Mahlau & Hartke, 2012) wurden erste Ergebnisse der Arbeit nach dem Konzept RIM auf die Schulleistungen sowie die sprachliche und Emotionale und soziale Entwicklung dargestellt. Dazu wurde ein Zweigruppen-Versuchsplan (Kontroll- und Experimentalgruppe) mit Prä-Post-Testung gewählt. An dieser Stelle soll nur ein zusammenfassender Überblick gegeben werden, für die detaillierte Ergebnisdarstellung sei auf Voß et al. (2012) verwiesen. Im Bereich des Lesens und der Mathematik weisen die Experimentalgruppe und die Kontrollgruppe, bezogen auf die Eichstichproben eingesetzter Testverfahren, im Mittel durchschnittliche Leistungen auf, wobei die Schüler der Experimentalgruppe etwas geringere mathematische Leistungen erzielen. Ebenso weisen die Experimentalgruppenkinder Nachteile in der Rechtschreibung auf, jedoch sind die Ergebnisse beider Untersuchungsgruppen hier durch einen geringen Leistungsstand gekennzeichnet. In den Förderschwerpunkten Sprache sowie Emotionale und soziale Entwicklung deuten die Ergebnisse auf positive Effekte des RIM hin. So erweist sich die inklusive sprachliche Förderung in den Grundschulklassen als ebenso erfolgreich wie in den Sprachheilgrundschulklassen. Ähnlich sieht es im Bereich der Emotionalen und sozialen Entwicklung aus. So zeigen beide Untersuchungsgruppen bezogen auf die Normstichprobe eine positive Entwicklung auf, mit Vorteilen hinsichtlich des prosozialen Verhaltens für die Kontrollgruppe.

> *„Die Wirksamkeit des auf Rügen praktizierten RTI-Ansatzes trägt gegenwärtig zwar zu einer gelingenden leistungs- und inklusionsorientierten Grundschulpädagogik bei, seine weiterreichenden in der Fachliteratur beschriebenen Potenziale werden gegenwärtig noch nicht ausgeschöpft. Dies ist verständlich, wenn man berücksichtigt, dass die Forschung zum RTI-Ansatz in Deutschland erst am Anfang steht, während international seit mehreren Jahrzehnten in Anschluss an diesen Ansatz bzw. innerhalb einer stärker empirisch ausgerichteten Unterrichtsforschung Messverfahren sowie Unterrichts- und*

Fördermethoden wissenschaftlich erarbeitet wurden“ (Voß et al., 2012, S. 100).

Der Bericht hat zu diesem Zeitpunkt jedoch den Charakter eines Zwischenstandes, da eine differenzierte Darstellung der Forschungsergebnisse zum Ende der Klasse drei noch aussteht.

3.2 Curriculumbasierte Messverfahren als Spezialform des formative assessment

In nachfolgenden Abschnitten sollen Curriculumbasierte Messverfahren als eine Spezialform des formative assessment näher beschrieben werden. Neben einer Übersicht zur Definition und zur Abgrenzung von weiteren Verfahren ähnlicher Natur soll dabei auf die praktische Anwendung im schulischen Alltag sowie auf verfügbare CBM-Verfahren im US-amerikanischen als auch deutschen Raum eingegangen werden. Darüber hinaus werden die Bedeutung als auch Grenzen von CBM dargelegt und eine Zusammenschau von Forschungsbemühungen und deren Ergebnissen gegeben.

3.2.1 Definition Curriculumbasierter Messverfahren (CBM)

Im Zeitraum von 1977 bis 1985 entwickelten und testen Stanley L. Deno und Phyllis K. Mirkin mit *Data-Based Program Modification* (DBPM) ein Interventionsprogramm zur Unterstützung von Sonderpädagogen bei der Erstellung individueller Förderpläne für Kinder mit Lern- oder Verhaltensschwierigkeiten (Deno, 1985; Deno & L.S. Fuchs, 1987; Deno & Mirkin, 1977). Computergestützt wurden dazu in regelmäßigen Abständen empirisch gewonnene Schülerdaten gespeichert, zusammengefasst und grafisch ausgewertet. Auf der Basis der vorliegenden Daten konnte der Anwender schnell und einfach die Schülerentwicklung verfolgen und

ggf. systematische Förderanpassungen bzw. -änderungen vornehmen (Deno & Mirkin, 1977). Zentrale Komponente des Programms waren dabei die regelmäßig durchgeführten Lernstandserhebungen. Es handelte sich hierbei um kurze fachspezifische Testverfahren, welche durch die Schüler bearbeitet wurden und an die besondere Bedingungen geknüpft wurden, welche Hosp, Hosp und Howell (2007) wie folgt beschreiben:

1. Die Tests fragen curricular zu vermittelnde Inhalte ab. Diese Basierung auf dem Curriculum war Grundlage für die Namensgebung der Testverfahren: *curriculum-based measurement* (curriculumbasierte Messungen, CBM).
2. An die CBM werden wissenschaftliche Gütekriterien angelegt, d. h. sie müssen reliabel und valide sein. Es handelt sich demnach nicht um informelle Verfahren, sondern um im Sinne der pädagogischen Diagnostik entwickelte Tests.
3. Zur Bewertung der mittels CBM erhobenen Schülerleistungen werden die jeweiligen Ergebnisse im Vergleich zu einem zuvor definierten Kriterium, nicht primär im sozialen Vergleich mit den Leistungen anderer Kinder betrachtet.
4. Bei der Konzeption und Durchführung von CBM werden festgelegte Standards eingehalten, um die Objektivität der Tests zu gewährleisten.
5. Zur Leistungseinschätzung werden die richtig gelösten Aufgaben innerhalb eines fest definierten Zeitintervalls bestimmt.
6. Feste Regeln zur Interpretation der Daten werden bestimmt, damit der Anwender die Bedeutung unterschiedlicher Schülerleistungen einschätzen kann.
7. CBM-Daten werden regelmäßig wiederholt eingesetzt. Durch die Aggregation der Daten lässt sich ein Lernverlauf der Schülerleistungen abbilden. Dies ermöglicht die Überwachung des Lernprozesses und gibt direktes Feedback über

verwendete Lehrmethoden und -strategien. Ein Leistungszuwachs des Schülers zeigt sich demzufolge in einer gestiegenen Punktzahl. Eine gleichbleibende Punktzahl deutet auf eine Stagnation des Schülers hin, eine abfallende Punktzahl zeigt einen Leistungsabfall an. In letzteren beiden Fällen sind Anpassungen der Lehr- bzw. Fördermaßnahmen angezeigt.

8. Der Einsatz von CBM soll ökonomisch gestaltet sein. Eine Einweisung in die Durchführung der Verfahren ist schnell und einfach möglich. Dies liegt auch darin begründet, dass die erhobenen Rohwertdaten einfach verarbeitet und ausgewertet werden können und nicht in andere Formate (T-Wert, PR-Wert etc.) transformiert werden müssen. Eine Vermittlung der Durchführung, Auswertung und Interpretation ist auch an Personen ohne differenziertes testtheoretisches Wissen gut möglich.
9. Die mittels CBM erhobenen Daten können auf verschiedene Weise ausgewertet werden. Die grafische Ausgabe in einem Diagramm bspw. erlaubt eine einfach verständliche Interpretation für Lehrer und Schüler. Durch den Einsatz des Computers kann diese Arbeit deutlich vereinfacht werden (Hosp et al., 2007).

Die hier aufgeführten Punkte beschreiben nicht ein konkretes Testverfahren, sondern ziehen den Rahmen für eine breite Palette an Methoden zur Ermittlung des Lernverlaufs von Schülern. Allen gemein ist, dass die Schülerleistungen durch das Bearbeiten von vorgelegten konkreten Materialien direkt erhoben werden. Meist handelt es sich dabei um Arbeitsblätter (*probes*), welche innerhalb einer festgelegten Zeitvorgabe bearbeitet werden sollen (Wright, 2001). Je nachdem, welche Fähigkeit abgefragt wird, ist eine Bearbeitungszeit zwischen einer und fünf Minuten vorgegeben. Durch die kurze Bearbeitungszeit sind CBM, ähnlich wie eine tägliche Übung, gut in den laufenden Unterricht zu integrieren (L.S. Fuchs &

D. Fuchs, 2007). Die Schülerleistung kann im Anschluss hinsichtlich der Bearbeitungsflüssigkeit (fluency) und Korrektheit (accuracy) ausgewertet werden. Der Lernfortschritt des einzelnen Kindes wird über jeweils erreichte Punktzahlen beim bis hin zu wöchentlich frequentierten Einsatz der CBM abgebildet (L.S. Fuchs, D. Fuchs & Courey, 2005), wodurch eine regelmäßige Leistungseinschätzung erfolgt und ggf. kurzfristig entsprechende Fördermaßnahmen eingeleitet werden können. Dabei muss sichergestellt sein, dass die Tests zwar aus verschiedenen Aufgaben bestehen, welche sich jedoch hinsichtlich des zu prüfenden Konstrukts und des inhärenten Schwierigkeitsgrades gleichen, also parallel sind (Deno, 2003a,b). Erst wenn diese Voraussetzungen gegeben sind, können die Lernverlaufsergebnisse klar interpretiert werden und erst dann ist der Rückbezug von Schülerleistungen auf den Unterricht zulässig (Deno, 2003a, b).

Die innerhalb von CBM erzielten Ergebnisse können gemäß L.S. Fuchs et al. (2005) auf verschiedene Weise interpretiert werden. „Through this linking of paradigms, CBM simultaneously provides information about relative standing as well as progress" (S. 34). Die Messergebnisse lassen einerseits eine interindividuelle Beurteilung über die mittels regelmäßiger Testungen erhobene Leistungsentwicklung im Laufe der Zeit zu (*individuelle Bezugsnorm*, Schrader & Helmke, 2002). Andererseits – auch wenn nicht zu diesem Zweck konzipiert (Hosp et al., 2007) – können die Daten vor dem Hintergrund von Vergleichswerten anderer Kinder, wie z. B. anhand des Klassendurchschnitts oder weiterer Normwerte (*soziale Bezugsnorm*, Schrader & Helmke, 2002), eingeordnet und interpretiert werden. Somit können CBM sowohl als kriteriumsorientierte als auch als normorientierte angesehen werden (Ingenkamp & Lissmann, 2008).

Denos Arbeit begann unabhängig von den in den Abschnitten 3.1 und 3.1.4 erwähnten bildungspolitischen Gegebenheiten in den USA, wurde aber dadurch umso bedeutsamer. Im Laufe der Zeit verbreiteten sich die Verfahren und werden nun auch im Regelunterricht in unterschiedlichen Fachbereichen eingesetzt. Am verbreitetsten ist der Einsatz von CBM in den Bereichen Lesen und Rechtschreibung sowie Mathematik, wobei die Entwicklung und Erforschung von mathematischen Verfahren erst in jüngerer Vergangenheit an Bedeutung gewann (Foegen, Jiban & Deno, 2007; VanDerHeyden & Burns, 2005; VanDerHeyden & Burns, 2008).

3.2.2 CBM, CBA, DIBS, MM, GOM, SBM – Systematisierung und Abgrenzung zentraler Begriffe

Wie zuvor beschrieben, sollen CBM Lernfortschritte von Schülern mittels regelmäßig wiederholter Messungen quantifizieren und den sich abzeichnenden Verlauf dokumentieren, um daraus Schlüsse auf die Effektivität des Unterrichts bzw. der Förderung als auch das Ansprechen des Schülers darauf ziehen zu können. Damit gliedern sich CBM in die Methoden des formative assessment ein. Durch die besondere Akzentuierung der wissenschaftlichen Gütekriterien (Deno, 1985) im Sinne der pädagogischen Diagnostik heben sich CBM jedoch von vielen anderen informellen Verfahren des formative assessment ab. Die Abkürzungen CBM, CBA, DIBS, MM, GOM sowie SBM werden oft in diesem Zusammenhang beschrieben. Welche Konzepte sich konkret hinter diesen Abkürzungen verbergen und in welcher Beziehung diese zueinander stehen, soll im folgenden Abschnitt herausgearbeitet werden.

Das in Verbindung mit CBM häufig benannte c*urriculum-based assessment* (CBA, Tucker, 1985) umschließt alle Maßnahmen des

Informationsgewinns zur pädagogischen Entscheidungsfindung (Deno, 2003a), wie bspw. „scoring a student's worksheets to determine the percentage of questions answered correctly on a worksheet; doing an error analysis of a student's oral reading from instructional text, or establishing 'mastery' of a new skill based on performance on an end of unit test (Deno, 2003a, S. 4f.). Aus dieser Perspektive betrachtet, werden CBM als ein Teil des CBA verstanden. Dabei unterscheiden sich die beiden Bereiche wesentlich in zwei Punkten: hinsichtlich ihres Zwecks und ihrer wissenschaftlichen Güte (Shinn & Bamonto, 1998). Während CBM in erster Linie den Lernverlauf von Schülern dokumentieren sollen und gleichzeitig als Evaluationswerkzeug für den Unterricht fungieren, dienen Maßnahmen des CBA primär nur dem Zweck, Informationen zu liefern, wie eine Passung des Unterrichts an individuelle Schülervoraussetzungen erreicht werden kann. Im Gegensatz zu der Entwicklung von CBM, besteht bei der Konzeption von Maßnahmen des CBA nicht die Notwendigkeit des Einsatzes standardisierter, d. h. wissenschaftlich geprüfter, Testverfahren (Shinn & Bamonto, 1998). Als Testmaterial für das curriculum-based assessment wird das aktuell verwendete Unterrichtsmaterial herangezogen, um den direkten Bezug zum Unterrichtsgeschehen herzustellen (Shinn & Bamonto, 1998), getreu dem Motto: „test what you teach and teach what you test" (Hosp et al., 2007, S. 3). Dies war zunächst auch eine übliche Vorgehensweise bei der Konzeption von CBM (Deno, 2003a), jedoch mangelte es dabei an wissenschaftlicher Güte. L.S. Fuchs und Deno (1994) stellten im Zuge der Weiterentwicklung der CBM fest, dass auch Verfahren, welche sich nicht direkt am aktuell behandelten Lehrmaterial orientieren, gleiche Effekte erzielen und darüber hinaus auch wissenschaftliche Gütekriterien erfüllen. L.S. Fuchs et al. (2005) beschreiben diese Form der CBM folgendermaßen: „every skill in the annual curriculum is represented on each weekly test. So, like a norm-referenced test, CBM samples a broad range of skills, and each repeated measurement is an alternate form of equivalent difficulty,

assessing the same constructs" (S. 34). Die daraus resultierenden Vorteile beschreibt Deno (2003a) wie folgt:

> „*This 'uncoupling' of CBM from the local school's curriculum has made it increasingly possible, both in research and practice, to capitalize on using standardized stimulus materials without the loss of the relevance of CBM for making everyday instructional programming decisions"* (S. 5).

Erst durch das Einhalten wissenschaftlicher Gütekriterien, vor allem der Parallelität der Verfahren, so L.S. Fuchs und Deno (1994), können die erhobenen Daten der Kinder als Maß der allgemeinen Leistung der Schüler (general outcome) interpretiert werden. Aus diesem Grund werden diese Verfahren auch als g*eneral outcome measures* (GOM) bezeichnet (L.S. Fuchs & Deno, 1994). GOM umfassen Aufgabenformate, welche nur durch das Zusammenspiel erfolgreich ausgeprägter Teilkompetenzen gelöst werden können (Hosp et al., 2007). Eine weitere, als synonym zu den GOM betrachtbare Form von CBM beschreibt Shinn (1995, 1998) mit seinen dynamic indicators of basic skills (DIBS, Shinn, 1995). Beiden CBM-Typen gemein ist das Vorgehen bei der Testentwicklung. Shinn (1995) schlägt dazu vor, Aufgaben zu entwickeln, die in hohem Maße mit den verschiedenen für den jeweiligen schulischen Bereich bedeutsamen Kompetenzen bzw. dem Lernziel korrelieren (robust indicators). „Knowing students' scores on the indicator can give one an accurate picture of their performance on a broader [Herv. d. Verf.] number of tasks in the same domain" (Shinn & Bamonto, 1998, S. 6). Ein Beispiel dafür stellt die Leseflüssigkeit dar, welche sich aus einer Vielzahl von Teilkompetenzen (Buchstaben- und Wörterkenntnis, Lautsynthese etc.) zusammensetzt. „Success or improvement on the GOM is assumed to reflect the synthetic application of the contributing skills" (Hosp et al., 2007, S. 11). Einerseits muss dadurch auf den Einsatz verschiedener Verfahren zur Erfassung der Einzelkompetenzen verzichtet werden, was im Sinne der Ökonomie

als großer Vorteil angesehen werden kann. Auf der anderen Seite liefern GOM bzw. DIBS jedoch kein differenziertes Feedback darüber, welche Teilkompetenzen eines Kindes evtl. noch ungenügend ausgeprägt sind. Insgesamt lässt sich jedoch festhalten, dass GOM bzw. DIBS sowohl als Screening als auch zur Diagnostik des Lernverlaufs geeignet sind (Hosp et al., 2007).

Als eine weitere Form des CBM führen Hosp et al. (2007) sog. skills-based measures (SBM) an. Im Gegensatz zu GOM werden hier einzelne Teilkompetenzen eines schulischen Bereichs abgefragt. Zur Entwicklung von SBM werden alle für einen festgelegten Zeitraum (bspw. ein halbes oder ein ganzes Schuljahr) relevanten Lernziele ermittelt. Für jedes dieser Lernziele werden eine Reihe Items gleicher Schwierigkeit generiert und gleichmäßig auf verschiedene parallele Formen verteilt (curriculum sampling). „This produces a set of equivalent measures providing balanced coverage of the same content" (Hosp et al., 2007, S. 12). Bspw. könnte ein Verfahren zur Überprüfung des Lernfortschritts im Rechnen Aufgaben zu den einzelnen Bereichen Addition, Subtraktion, Multiplikation und Division in jeweils unterschiedlich hohen Schwierigkeitsausprägungen auf einem Aufgabenblatt zusammenfassen. Mittels regelmäßigen Einsatzes des SBM kann der Lernverlauf über die verschiedenen Lernziele hinweg erhoben werden. Dies bedeutet jedoch, dass die Schüler zunächst nur wenige Aufgaben richtig lösen, da entsprechendes Wissen bzw. entsprechende Kompetenzen zum Lösen aller Aufgaben (z. B. die Multiplikation oder Division) noch nicht unterrichtet wurden. Zum Ende des festgelegten Zeitraums dreht sich dieses Verhältnis um, die Kinder können nun die meisten der Aufgaben lösen, da sie das entsprechende Wissen bzw. entsprechende Kompetenzen bereits erlernt haben. Je nachdem wie groß der Testzeitraum gewählt wird, kann es dabei sein, dass die meisten der Testaufgaben für den jeweils aktuell unterrichteten Lernstoff irrelevant sind (Hosp et al., 2007). Ein Vorteil dieser

Verfahren ist, dass sie nah am Schuljahrescurriculum konzipiert sind und der unterrichtenden Lehrkraft demnach direktes Feedback zur Unterrichtsgestaltung liefern. Die Kehrseite ist, dass dadurch für jedes Schuljahr verschiedene, auf die Klassenstufe angepasste Verfahren konzipiert werden müssen, die Dokumentation des Lernverlaufs ist dadurch nur innerhalb eines Schuljahres, nicht aber über mehrere Klassenstufen hinweg möglich (Foegen, Jiban & Deno, 2007). Skills-based measures können als Screeningverfahren genutzt werden, um Risikokinder zu identifizieren. Darüber hinaus können sie auch hinsichtlich einzelner Lernziele ausgewertet werden und liefern somit auch qualitative Hinweise über die Schülerleistungen (Hosp et al., 2007). So könnte sich zum Bespiel herausstellen, dass ein Kind Additions- und Subtraktionsaufgaben lösen kann, jedoch noch Schwierigkeiten bei der Multiplikation und Division zeigt.

Ein weiteres Format von CBM stellen mastery measures (MM) dar (Hosp et al., 2007). Dabei handelt es sich um Verfahren, ähnlich den SBM, welche jedoch nicht mehrere Kompetenzbereiche zusammengefasst abfragen, sondern jeden Kompetenzbereich diskret erheben. VanDerHeyden & Burns (2005) beschreiben die Vorgehensweise bei der Entwicklung von mastery measures wie folgt: „specific hierarchical sequence of skills and subskills are defined and the performance of each specific subskill is measured to ensure that the skill has been mastered before providing instruction on the next skill in the hierarchy" (S. 16). Bezogen auf den Bereich der Mathematik stellen Verfahren, welche diskret voneinander Kompetenzen hinsichtlich der Addition, der Subtraktion sowie Multiplikation erheben, ein Beispiel für mastery measures dar. Ein hypothetisches Beispiel für ein mittels mastery measures ermitteltes Lernverlaufsprofil ist in Abbildung 12 demonstriert.

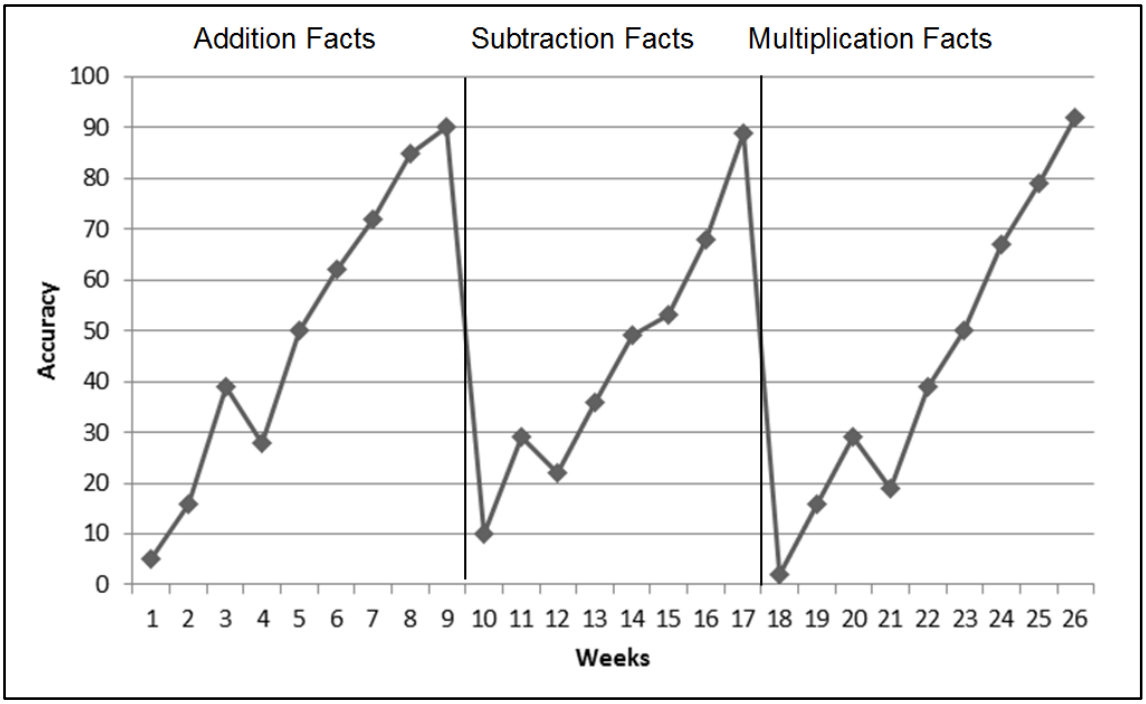


Abbildung 12: Hypothetisches Beispiel für ein mittels mastery measures ermitteltes Lernverlaufsprofil in Anlehnung an Hosp et al. (2007) (entnommen aus: Hosp et al., 2007, S. 13)

Mastery measures werden eingesetzt, wenn man den Entwicklungsverlauf einzelner diskreter Kompetenzen nachvollziehen möchte, z. B. beim Vermitteln von Basiskompetenzen, welche Grundlage für weitere Kompetenzen sind. Ebenso eignen sich mastery measures bei der Suche nach Problembereichen eines Schülers. Trotz dieser Vorzüge sind MM in der Literatur in Kritik geraten, da angezweifelt wird, dass durch das isolierte Betrachten einzelner Kompetenzen Rückschluss auf die schulische Leistung innerhalb eines Schulfaches gezogen werden kann (L.S. Fuchs, 2004; Hosp et al., 2007). Darüber hinaus ist ein längerfristiger Lernverlauf mittels MM nur in der in Abbildung 12 dargestellten Form möglich. Dies kann jedoch demotivierend auf Schüler wirken, denn sobald sie ein Ziel erreicht haben, sinkt die erreichte Punktzahl sofort wieder ab (Hosp et al., 2007).

Abbildung 13 gibt den Zusammenhang zwischen den hier beschriebenen Konzepten des curriculum-based asessment, des

curriculum-based measurements der skills-based measures, der general outcome measures, der dynamic indicators of basic skills und des mastery measures schematisch wieder. Zusammenfassend lässt sich festhalten, dass sich alle hier vorgestellten Konzepte in das CBA eingliedern lassen. Während Hosp et al. (2007) das Konzept des mastery measures als eine Art von CBM darstellen, bringen VanDerHeyden und Burns (2005) es mit dem weiter gefassten CBA in Verbindung. Alle unter CBM zusammengefassten Verfahren unterliegen strengen wissenschaftlichen Gütekriterien, so auch SBM, GOM, DIBS sowie ein Teil der MM. Laut Ausführungen von Shinn (1995), Shinn und Bamonto (1998) sowie L.S. Fuchs und Deno (1994) und Deno (2003a, b) handelt es sich bei General Outcome Measures und Dynamic Indicators for Basic Skills um Konzepte, die nicht klar voneinander zu trennen sind. Somit kommt es zu Überschneidungen. Skills-based measures (Hosp et al., 2007) hingegen lassen sich klar von den anderen Konzepten abgrenzen. Während GOM sowie auch DIBS relativ komplexe Kompetenzen abfragen, erheben SBM und MM einfache Kompetenzen, wobei SBM verschiedene Bereiche anhand eines Verfahrens überprüfen, MM jedoch jeweils nur einen diskreten Bereich fokussieren (Hosp et al., 2007).

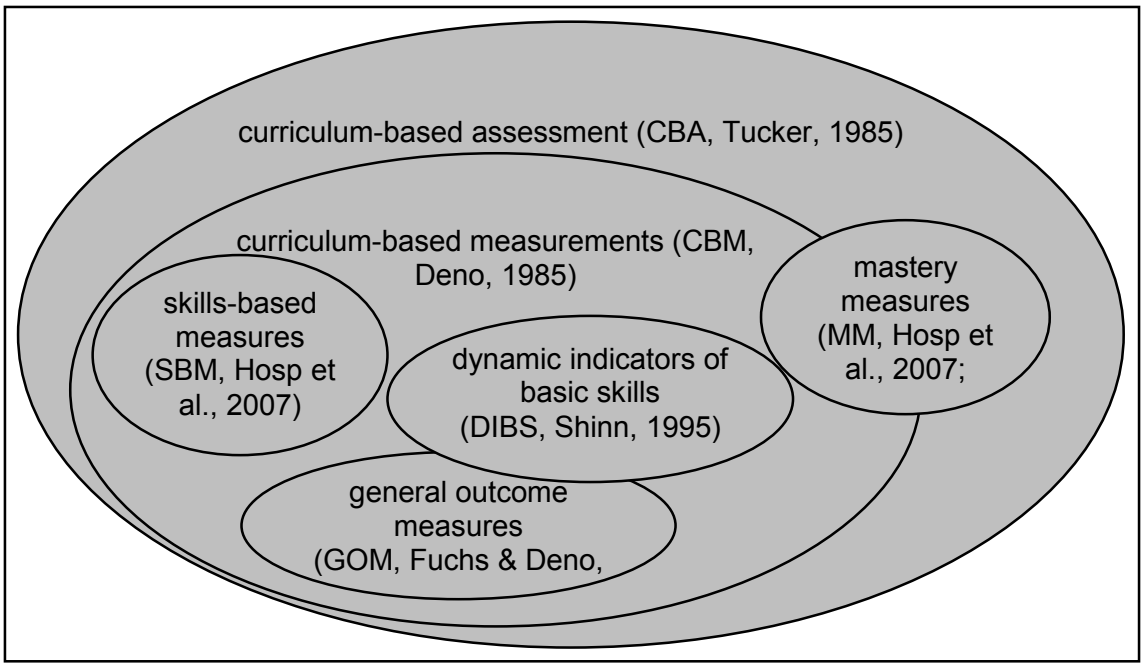


Abbildung 13: Schematische Darstellung des Zusammenhangs von curriculum-based asessment (CBA), curriculum-based measurements (CBM), skills-based measures (SBM), general outcome measures (GOM), dynamic indicators of basic skills (DIBS) und mastery measures (MM) in Anlehnung an Tucker (1985), Deno (1985), Hosp et al. (2007), L.S. Fuchs und Deno (1994), Shinn (1995) sowie VanDerHeyden & Burns (2005)

3.2.3 Praktischer Einsatz von CBM im Bereich Mathematik

Die Durchführung und Auswertung von CBM ist sehr einfach gestaltet, um den regelmäßigen Einsatz im schulischen Alltag zu ermöglichen. In der Regel liegen CBM in Paper-and-Pencil-Form vor, d. h. die Schüler erhalten Testmaterial in Form von Arbeitsblättern, welche eine Serie an Aufgaben bereitstellen. Es gibt jedoch auch Internetdienste, die eine internetgestützte Durchführung der CBM ermöglichen, wie bspw. www.easyCBM.com. Der Vorteil dabei ist die automatisierte Auswertung und Verarbeitung der Daten, was eine erhebliche Vereinfachung und Zeiteinsparung bei der Durchführung und Auswertung mit sich bringt.

Im Gegensatz zum Lesen können im Bereich der Mathematik die meisten Tests als Gruppenverfahren genutzt werden, was zu einer deutlichen Vereinfachung des praktischen Einsatzes beiträgt. Die Testdauer beläuft sich meist auf eine bis zwei Minuten. Wie oft und wann CBM zur Erhebung von Schülerleitungen eingesetzt werden, hängt vom Zweck ihres Einsatzes ab. Im Sinne eines Screening werden CBM normalerweise drei- bis viermal im Schuljahr eingesetzt (Hosp et al., 2007). Zur detaillierteren Beobachtung des Lernverlaufs bieten sich darüber hinaus monatliche Testungen an, dadurch können Leistungsentwicklungen zeitnah erfasst werden. Laut Auffassung von Hosp et al. (2007) sollten Schüler, deren Leistungen dem unteren Quartil entsprechen, höher frequentiert getestet werden. Die Autoren schlagen dafür einen wöchentlichen bis hin zu zweimal wöchentlichen Einsatz vor.

Für die Durchführung der CBM liegen in der Regel vorgegebene wörtliche Instruktionen vor, sodass die Objektivität bei der Anwendung gewährleistet ist. Auch die Auswertung der CBM gestaltet sich recht einfach. Dabei gibt es verschiedene Möglichkeiten. Es können einfach die richtig gelösten Aufgaben oder aber auch die Anzahl richtiger Ziffern bestimmt werden Zur Auswertung der Daten eines Mathematik-CBM schlagen Hosp et al. (2007) vor, anstatt die richtig gelösten Aufgaben, die Anzahl der korrekt gelösten Ziffern zu zählen.

> *„When scoring Math CBM, the number of correct digits in the solution to the problem (which includes critical processes, not just the answer) rather than the number of correct problems is used because it is a more sensitive measure to change. It is also considered a fairer metric because the student is awarded more points for correctly solving more complex problems“* (Hosp et al., 2007, S. 104).

In Abbildung 14 ist die Auswertung von Schülerlösungen anhand der Anzahl korrekter Ziffern beispielhaft dargestellt. Pro richtige Ziffer wird dabei ein Punkt vergeben, falsche Ziffern werden nicht gezählt. Das Gesamtergebnis erhält man dann durch das Summieren der erreichten Punkte. Die erste Rechnung erbringt ein Ergebnis von vier Punkten, die zweite Rechnung ergibt drei, die dritte zwei Punkte. Eine Auswertung anhand der richtig gelösten Aufgaben würde zu einem Ergebnis von einem Punkt für die erste Rechnung ergeben, die beiden anderen Rechnungen würden jeweils nicht bepunktet werden, da sie insgesamt als falsch zu werten sind.

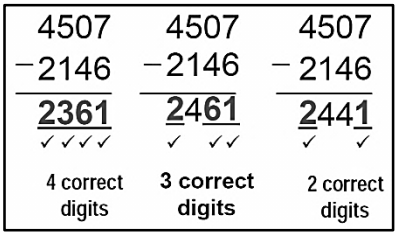


Abbildung 14: Auswertung von Schülerlösungen anhand der Anzahl korrekter Ziffern (entnommen aus: L.S. Fuchs & D. Fuchs, 2005, S. 14)

Obwohl mittlerweile allgemein üblich (L.S. Fuchs & D. Fuchs, 2005; Hosp et al., 2007), ist diese Form der Auswertung jedoch auch kritisch zu betrachten. So hinterfragt bspw. Knopp (2010a) in diesem Zusammenhang,

> *„ob die dadurch stärkere Gewichtung von Aufgaben mit einem zweistelligen Ergebnis und die damit vorgenommene Stufung im Schwierigkeitsgrad auf Basis der Erkenntnisse zum Erwerb erster arithmetischer Kenntnisse gerechtfertigt ist. […] Wenn eine Gewichtung von Aufgaben erfolgt, sollten dann nicht neben dem Kriterium ‚das Ergebnis ist größer bzw. kleiner als 10' andere Kriterien ebenfalls berücksichtigt werden"* (S. 69).

Welcher Ansatz auch gewählt wird, die ermittelten Punktzahlen für die richtig gelösten Ziffern bzw. Aufgaben werden nach der Bearbeitung zusammengezählt. Zur besseren Veranschaulichung kann das Ergebnis, ähnlich wie in Abbildung 15 gezeigt wird, in einem

Diagramm abgetragen werden. Auf der Abszisse (x-Achse) sind die entsprechenden Messzeitpunkte bzw. Schulwochen einzutragen, die erreichten Punktzahlen auf der Ordinate (y-Achse). Zur Interpretation der erhobenen Daten können diese mit einem zuvor festgelegten Ziel (das Kreuz in Abbildung 15) verglichen werden. Dazu bietet es sich an, eine sog. Ziel- oder Richtlinie zu ermitteln und in demselben Diagramm abzutragen, wie die gestrichelte Linie in Abbildung 15. Zur Bestimmung dieses Ziels geben L.S. Fuchs und D. Fuchs (2005) drei verschiedene Vorgehensweisen an: das sog. *End-of-Year-Benchmarking,* den Vergleich anhand von Normwerten für einen durchschnittlichen Leistungsanstieg oder den sog. *Intra-Individual Framework.*

Das End-of-Year-Benchmarking stellt die einfachste Methode dar. Dabei geben vorhandene Normwerte das zum Schuljahresende zu erreichende CBM-Ergebnis an. Zur Ermittlung der Ziellinie wird der Median der ersten drei Messwerte mit dem entsprechend ermittelten Zielwert verbunden (L.S. Fuchs & D. Fuchs, 2005). In diesem Fall wird der Median als Ausgangspunkt herangezogen, da er robust gegen Ausreißerwerte ist und somit eine reliable, biasunanfälligere Grundrate widerspiegelt. Da in diesem Fall die Normwerte einer großen Schülerkohorte Grundlage der Leistungsinterpretation sind, wird bei dieser Vorgehensweise die soziale Bezugsnorm als Kriterium herangezogen.

Auch bei der zweiten Option zur Ermittlung des Lernziels zum Schuljahresende wird auf eine soziale Bezugsnorm zurückgegriffen. Grundlage sind dabei Normwerte für den durchschnittlichen wöchentlichen Leistungsanstieg. Diese werden als Faktor mit der Anzahl der bis zum Ende des Schuljahres verbleibenden Schulwochen multipliziert. Das Ergebnis ist der Wert für das gesuchte Lernziel. Wird dieser Wert mit dem Median der ersten drei CBM-Ergebnisse des Schülers verbunden, erhält man die Ziellinie.

Die für diese beiden Verfahrensweisen benötigten Normwerte konnten in den USA in den letzten Jahrzehnten erhoben und analysiert werden. Da die CBM-Forschung in Deutschland derzeit noch an ihrem Anfang steht, liegen derartig differenzierte Daten zu diesem Zeitpunkt nicht vor. So beschreiben L.S. Fuchs und D. Fuchs (2005) bspw. für das CBM „Monitoring Basic Skills Progress“ ((MBSP; L.S. Fuchs, Hamlett & D. Fuchs, 1998, 1999), dass ein Schüler der ersten Klasse einen wöchentlichen Leistungsanstieg von 0.35 Punkten und zum Schuljahresende ein Ergebnis von mindestens 20 Punkten in einem Rechen-CBM erreichen sollte. Aufgrund kultureller Unterschiede, speziell auch in Hinblick auf die verschiedenen Schulsysteme der USA und Deutschland, ist fraglich, ob solche in den USA erhobenen Normwerte auch für deutsche Schüler gültig sind. Dies gilt für die anhand von Normen ermittelten durchschnittlichen Lernzielwerte zum Schuljahresende sowie für die durchschnittlichen Leistungsanstiege.

Eine weitere Vorgehensweise zur Ermittlung des Lernziels zum Schuljahresende stellt der Intra-Individual-Framework dar. Wie der Name schon zu erkennen gibt, wird hierbei eine individuelle Bezugsnorm zur Bewertung der Schülerleistungen herangezogen. Zur Bestimmung des Lernziels nach dieser Methode sind mindestens acht CBM-Werte notwendig. Aus diesen Werten wird der durchschnittliche wöchentliche Anstieg des Schülers berechnet. Der errechnete mittlere Anstieg multipliziert mit dem Faktor 1.5 wird auf die bis zum Schuljahresende verbleibenden Schulwochen hochgerechnet (durch Multiplikation mit der Anzahl verbleibender Schulwochen). Das daraus resultierende Ergebnis stellt den avisierten Soll-Leistungsstand Lernziel dar. Verbunden mit dem Median der ersten acht CBM-Ergebnisse des Schülers, erhält man mit diesem Wert die Ziellinie. Warum bei dieser Vorgehensweise der Faktor 1.5 gewählt wurde, bleibt in den Ausführungen von L.S. Fuchs und D. Fuchs (2005) offen.

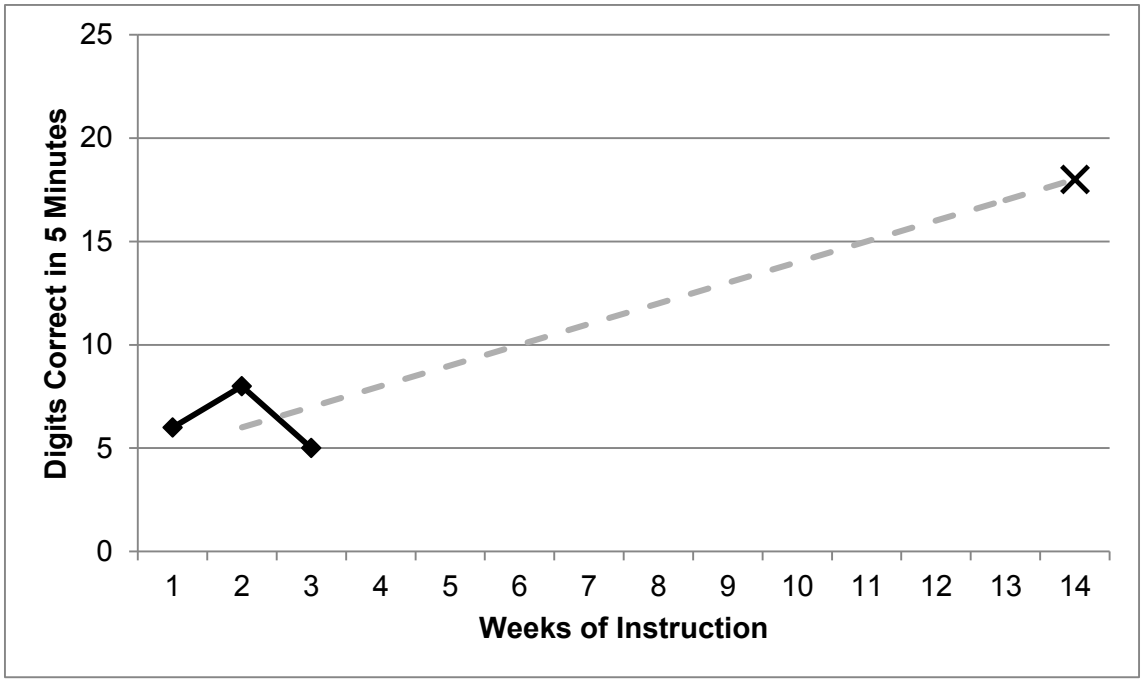


Abbildung 15: Beispiel für die grafische Darstellung von CBM-Ergebnissen (schwarze Linie) im Vergleich zu einer festgelegten Ziellinie (gestrichelte Linie) und zu einem zuvor definierten Ziel (schwarzes Kreuz) (entnommen und modifiziert aus: L.S. Fuchs & D. Fuchs, 2005, S. 66)

Die ermittelte Ziellinie in einem solchen Auswertungsdiagramm ermöglicht eine schnelle Einschätzung der Schülerleistungen. Durch Vergleich der Schülerergebnisse mit der Ziellinie lässt sich Lernfortschritt, -stagnation oder -rückgang leicht und objektiv ablesen. Laut L.S. Fuchs und D. Fuchs (2005) erscheint vor allem die Objektivierung der Schülereinschätzung als ein wesentlicher Vorteil der grafischen Auswertung von CBM-Ergebnissen.

> „*Frequently, teachers underestimate the rate at which students can improve (especially in special education classrooms), and the CBM graphs help teachers set ambitious, but realistic, goals. Without graphs and decision rules for analyzing the graphs, teachers often stick with low goals"* (L.S. Fuchs & D. Fuchs, 2005, S. 61).

Anhand des sich im Diagramm abzeichnenden Verlaufs kann die Lernentwicklung abgelesen werden, welche zugleich Aufschluss über

die Effektivität des Unterrichts gibt. Zur Interpretation des Graphen geben L.S. Fuchs und D. Fuchs (2005) zwei Möglichkeiten an:

a) Die letzten vier Datenpunkte entscheiden über weitere Maßnahmen. Liegen sie über der Ziellinie, wurde die Schülerleistung unterschätzt, das festgesetzte Ziel kann angehoben werden. Liegen sie unterhalb der Ziellinie, entspricht der Lernfortschritt des Schülers nicht den Erwartungen, der Unterricht muss den Bedürfnissen des Schülers angepasst werden, das festgesetzte Lernziel wird dabei nicht nach unten korrigiert. Ein erwartungsgemäßer Lernfortschritt zeigt sich, wenn die Datenpunkte und die Ziellinie übereinander liegen, dann müssen keine Maßnahmen vorgenommen werden (L.S. Fuchs & D. Fuchs, 2005).
b) Der sich in den Daten abzeichnende Trend entscheidet über weitere Maßnahmen. Liegt die Trendlinie über der Ziellinie, wurde die Schülerleistung unterschätzt, das Lernziel kann angehoben werden. Liegt die Trendlinie unterhalb der Ziellinie, entspricht der Lernfortschritt des Schülers nicht den Erwartungen, der Unterricht muss den Bedürfnissen des Schülers angepasst werden. Liegen Trend- und Ziellinie übereinander, zeigt der Schüler den erwarteten Lernfortschritt, es müssen keine Maßnahmen vorgenommen werden (L.S. Fuchs & D. Fuchs, 2005).

Abbildung 16 zeigt ein hypothetisches Beispiel einer Lernverlaufskurve eines Schülers (schwarze Linie). Die Daten wurden durch wöchentliche CBM erhoben. Die Werte der letzten vier Testungen liegen oberhalb der Ziellinie (gestrichelte Linie), die Schülerleistung wurde offensichtlich unterschätzt, dies ergibt auch die Betrachtung der Trendlinie (graue Linie). Das zuvor festgelegte Lernziel des Schülers sollte nach oben korrigiert werden (L.S. Fuchs & D. Fuchs, 2005).

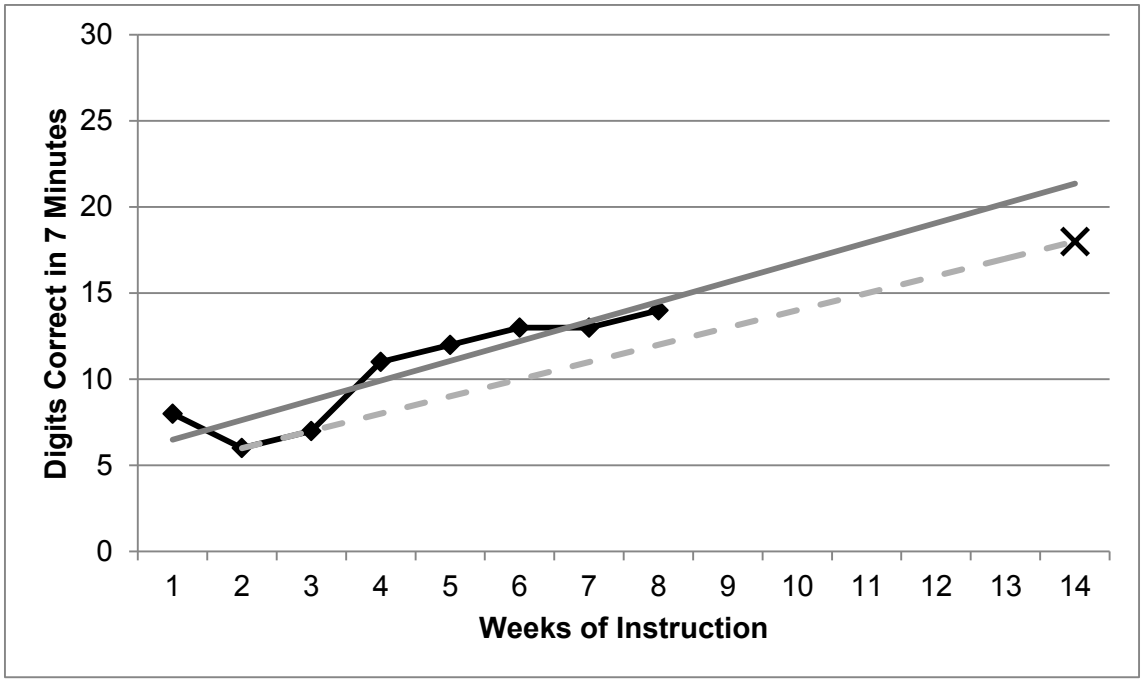


Abbildung 16: Hypothetisches Beispiel für den Lernverlauf eines Schülers (schwarze Linie) erhoben mittels CBM sowie der dazugehörigen Trendlinie (graue Linie) im Vergleich zu einer festgelegten Ziellinie (gestrichelte Linie) und zu einem zuvor definierten Ziel (schwarzes Kreuz) (entnommen und modifiziert aus: L.S. Fuchs & D. Fuchs, 2005, S. 74)

In Abbildung 17 ist ein weiteres hypothetisches Beispiel für den mittels CBM erhobenen Leistungsverlauf eines Schülers (schwarze Linie) dargestellt. Es zeigt sich, dass die Ergebnisse des Schülers ab der fünften Woche deutlich unter der Ziellinie (gestrichelte Linie) angesiedelt sind. Auch die Trendlinie zeigt einen nur geringen Leistungszuwachs an. Ohne entsprechende Maßnahmen kann der Schüler das zuvor festgelegte Lernziel nicht erreichen, eine Unterrichtsanpassung ist angezeigt (L.S. Fuchs & D. Fuchs, 2005).

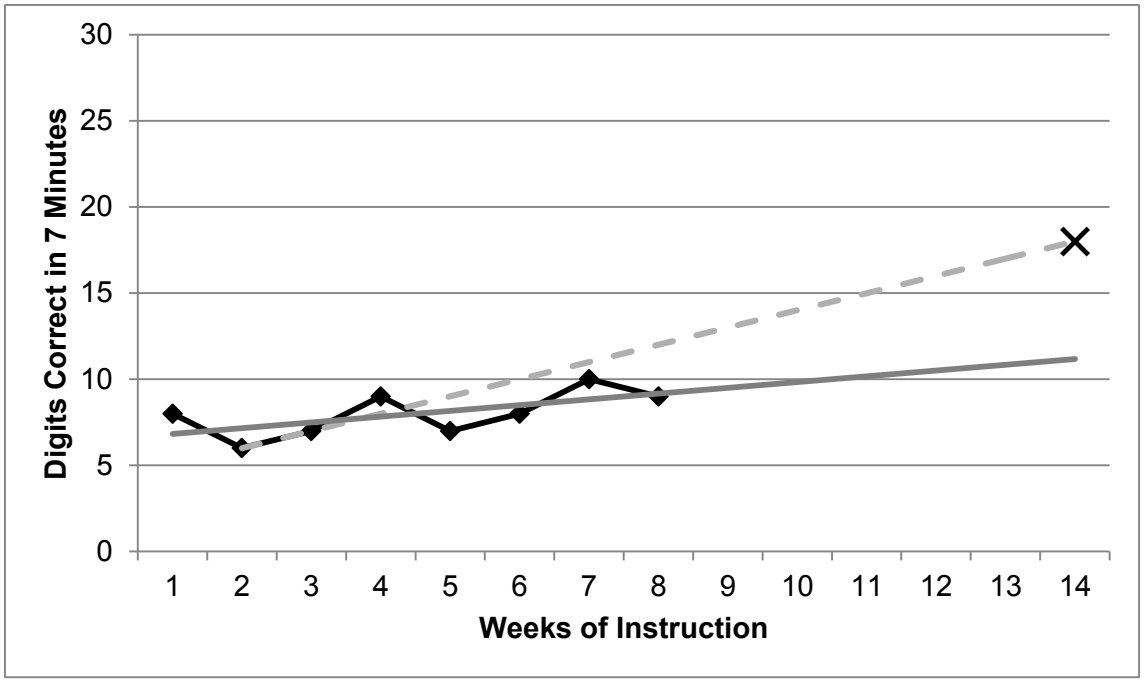


Abbildung 17: Hypothetisches Beispiel für den Lernverlauf eines Schülers (schwarze Linie) erhoben mittels CBM sowie der dazugehörigen Trendlinie (graue Linie) im Vergleich zu einer festgelegten Ziellinie (gestrichelte Linie) und zu einem zuvor definierten Ziel (schwarzes Kreuz) (entnommen und modifiziert aus: L.S. Fuchs & D. Fuchs, 2005, S. 75)

Grafische Auswertungen dieser Art können durch den Einsatz computer- oder internetbasierter Programme deutlich vereinfacht werden. In Amerika ist die computer- bzw. internetgestützte Lernfortschrittsdokumentation schon weiträumig etabliert, so gibt es verschiedene Anbieter, welche einen solchen Service gegen entsprechendes Entgelt anbieten. In Abschnitt 3.2.4 werden weitere Informationen zu diesen Internetseiten gegeben.

Insgesamt lässt sich festhalten, dass CBM-Verfahren, aufgrund ihrer Einfachheit in Durchführung und Auswertung, gut in den schulischen Alltag zu integrieren sind. Zur Anwendung der Verfahren ist dabei kein umfassendes Wissen nötig, sodass eine Einarbeitung nicht sehr kompliziert erscheint. Durch einen regelmäßig wiederholten Einsatz können Lernfortschritte der Schüler eingeschätzt werden, vor allem Leistungsstagnationen oder -abfälle werden gut beobachtbar. Bei der Interpretation der Daten sind jedoch klare Richtlinien notwendig, um

zu eindeutigen Leistungseinschätzungen zu kommen. Der Einsatz von computer- bzw. internetgestützten Programmen kann die Lehrkraft hierbei unterstützen. Insgesamt erscheinen CBM als ein gutes Hilfsmittel zur Lernverlaufsdokumentation, gerade auch im Bereich der Grundschulmathematik.

3.2.4 Verfügbare CBM-Verfahren für den Bereich Mathematik im US-amerikanischen Raum

Seit Entwicklung der CBM in den 80er Jahren sind diverse Forschungsbemühungen zur Untersuchung curriculumbasierter Messverfahren getätigt worden. Im Bereich der Mathematik wurde eine Reihe von Verfahren entwickelt und auf ihre Güte und Einsetzbarkeit evaluiert. Dabei lassen sich in Abhängigkeit von zu testenden mathematischen Inhalten und Konzepten drei verschiedene Typen von CBM unterscheiden: *early numeracy*, *coumputation* und *concepts & applications* (Hosp et al., 2007).

Early numeracy CBM fragen die Verfügbarkeit mathematischer Basiskompetenzen bei Schülern im Kindergarten bzw. zum Schuleintritt ab. In der Literatur sind Verfahren beschrieben, die die Kenntnis der Zahlen und des Zahlenstrahls, das numerische Verständnis von Zahlen sowie Zählkompetenzen erfassen (Chard, Clarke, Baker, Otterstedt, Braun & Katz, 2005; Clarke & Shinn, 2004; Lembke & Foegen, 2005; VanDerHeyden, Broussard, Fabre, Stanley, Legendre & Creppell, 2004; VanDerHeyden, Witt, Naquin & Noell, 2001). Tabelle 4 gibt einen Überblick über die bekanntesten in den USA verfügbaren Aufgabenformate von early numeracy CBM. Dabei gibt es Verfahren, welche jeweils nur einzelne Aufgabenformate abfragen (mastery measures) sowie solche, bei denen verschiedene Aufgabenformate gemischt zu einem Test zusammengefasst werden (skills-based measures).

Tabelle 4: Überblick über die bekanntesten in den USA verfügbare Early Numeracy CBM (Hosp et al., 2007)

Name	Beschreibung	Beispielaufgabe
quantity array	Innerhalb einer Bearbeitungszeit von einer Minute soll der Schüler die Mächtigkeit der jeweils vorgegebenen Punktmengen benennen. Das Verfahren wird in einer Einzelsituation durchgeführt.	●●●●●● ●●●●●●
number identification	In einer Minute soll der Schüler dem Testleiter vorgegebene Zahlen benennen. Der Test findet in einer Einzelsitzung statt.	5
oral counting	Der Schüler zählt in einer Einzelsituation mit dem Testleiter für eine Minute beginnend mit 1.	-
missing number	Dem Schüler werden verschiedene Abschnitte (drei oder vier Elemente) der Zahlenreihe präsentiert, bei welchen jeweils eine Zahl fehlt. Innerhalb einer Minute soll der Schüler die fehlenden Zahlen benennen.	10, 11, __
quantity discrimination	Von gegebenen Zahlenpaaren muss der Schüler innerhalb einer Minute jeweils die größere Zahl benennen.	13 7

Während die Forschungsgruppe um Clarke (Chard et al., 2005; Clarke & Shinn, 2004) CBM des Typs missing number und quantity discrimination erforscht hat, beschäftigten sich VanDerHayden und ihre Kollegen (VanDerHeyden et al., 2004; VanDerHeyden et al., 2001) darüber hinaus mit Verfahren, welche das Einkreisen bestimmter vorgegebener Zahlen (circling numbers) oder das Aufschreiben von genannten Zahlen (drawing numbers) abfragten. Weiterhin arbeiteten beide Forschungsgruppen an CBM des Typs number identification sowie oral counting. Foegen et al. (2007) kommen in ihrer Litaraturanalyse zu dem Schluss, dass die CBM number identification, quantity discrimination und missing number als die am besten empirisch abgesicherten Verfahren gelten.

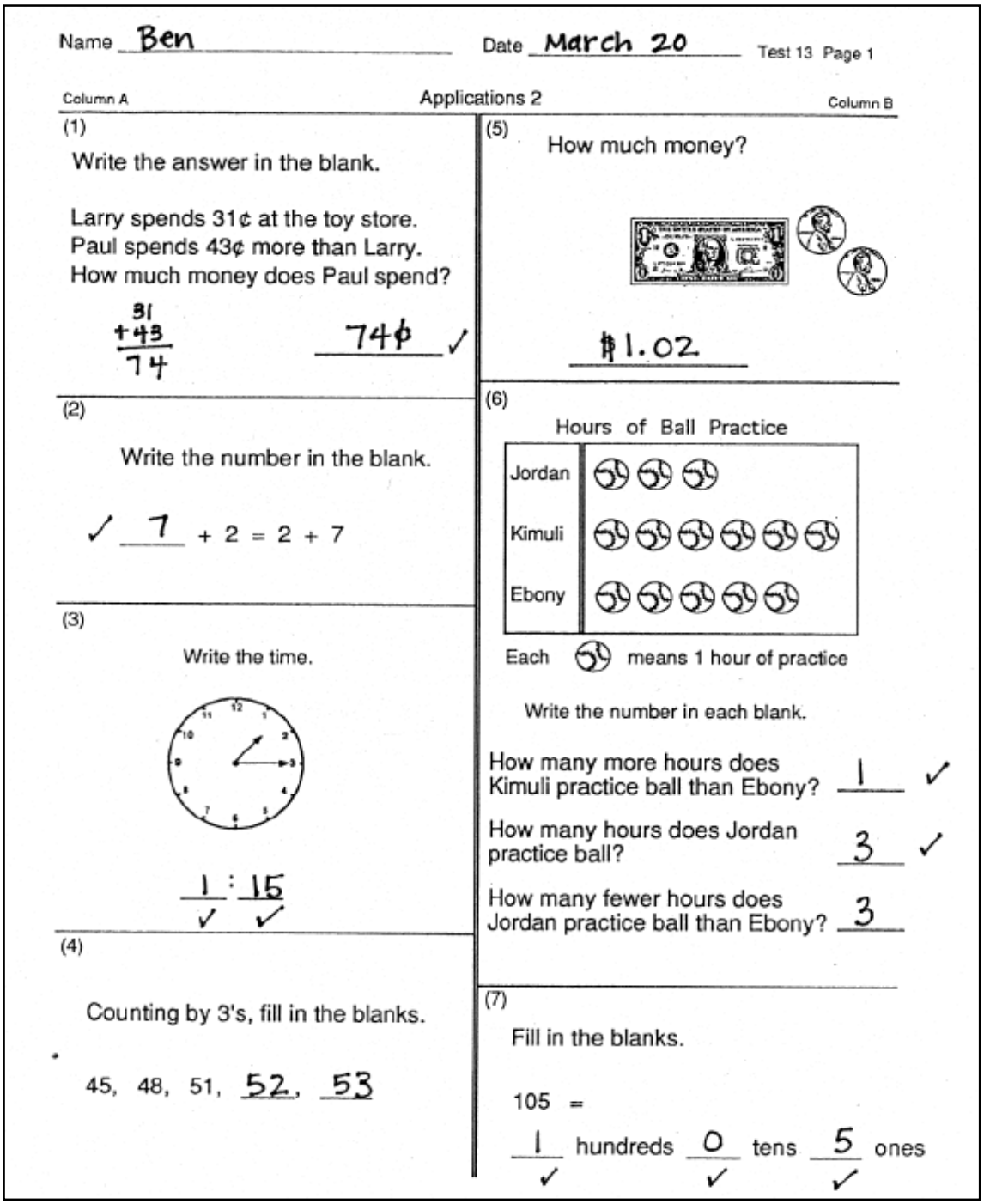
Als eine weitere Form von CBM im mathematischen Bereich gelten sog. CBM des Typs computation, also Verfahren zur Erfassung der Rechenfähigkeiten von Schülern. Abhängig von der Klassenstufe umfassen solche Tests Aufgaben auf unterschiedlichen

Anforderungsniveaus. So können sich Verfahren aus Aufgaben der verschiedenen Operationen Addition, Subtraktion, Multiplikation und Division sowie aus verschiedenen Zahlenräumen zusammensetzen (L.S. Fuchs & D. Fuchs 1992; Tindal, Germann & Deno, 1983). Durch die zeitliche Beschränkung der Bearbeitungszeit auf wenige Minuten wird dabei die Geschwindigkeit und Genauigkeit des Rechnens als Maß für mathematische Kompetenzen betrachtet und in Analogie zur Leseflüssigkeit als Indikator für Lesekompetenzen (Foegen et al., 2007). Nach VanDerHeyden und Burns (2008) weisen Kinder in einer Aneignungsphase mathematischer Aspekte zunächst eine hohe Genauigkeit, jedoch eine geringe Geschwindigkeit bei der Bearbeitung von Aufgaben hinsichtlich des neu erworbenen Inhalts auf. Im Laufe der Zeit, mit steigender Sicherheit, nimmt dann auch die Geschwindigkeit bei der Bearbeitung der Aufgaben zu. Ein angemessener Lernfortschritt im Bereich des Rechnens zeigt sich demnach in einer schnellen und genauen Bearbeitung der Rechen-CBM. Auch das National Mathematics Advisory Panel (2008) nimmt die Fähigkeit des mathematischen Faktenabrufs als einen wesentlichen Aspekt für das erfolgreiche mathematische Lernen an. Rechen-CBM gewinnen demnach neben ihrer Funktion zur Dokumentation des Lernverlaufs von Schülerleistungen auch durch die Möglichkeit des Einsatzes als Screening-Werkzeuge an Bedeutung. Speziell für diesen Zweck haben sich derartige CBM auch aus testtheoretischer Perspektive mehrfach bewährt. Wiederholt konnten mittlere Korrelationen mit standardisierten mathematischen Leistungtests nachgewiesen werden (Foegen & Deno, 2001; Thurber, Shinn & Smolkowski, 2002; Tindal, Marston et al., 1983).

Die Mathematik beschränkt sich nicht auf das bloße Rechnen von Aufgaben, dies gilt speziell auch für höhere Klassenstufen (D. Fuchs et al., 2003; Thurber et al., 2002). Neben der reinen Arithmetik spielen unter anderem auch das Sachrechnen, der Umgang mit Größen wie

dem Geld und der Uhrzeit aber auch stochastische sowie geometrische Inhalte eine wesentliche Rolle. Auch der weiterführende Aufbau von Einsichten und Konzeptvorstellungen (vgl. auch Abschnitt 2) stellt ein zentrales Thema dar. Daher erscheint es sinnvoll, verfügbare Rechen-CBM um solche Verfahren zu erweitern, welche derartigen curricular geforderten Inhalte abfragen. In diesem Zusammenhang wurden CBM des Typs concepts & applications konzipiert. Diese CBM bestehen aus Aufgaben, welche verschiedene mathematische Aspekte abfragen, wie Schätzaufgaben, Sachaufgaben, aber auch Items, die eher das Vorhandensein bestimmter Konzepte prüfen, wie bspw. das Verständnis für die Stellenbedeutung von Ziffern im Stellenwertsystem (Foegen & Deno, 2001; L.S. Fuchs, D. Fuchs, Hamlett, Phillips & Bentz, 1994; L.S. Fuchs, D. Fuchs, Karns, Hamlett & Katzaroff, 1999; Helwig, Anderson & Tindal, 2002; Hosp et al., 2007). CBM des Typs concepts & application unterscheiden sich nicht nur inhaltlich von den zuvor benannten CBM zur Erhebung von Rechenkompetenzen, auch hinsichtlich des Testaufbaus und der Testdurchführung gibt es Unterschiede (Hosp et al., 2007). So gibt es Verfahren, welche jeweils nur einzelne Aufgabenformate abfragen (mastery measures), jedoch auch solche, bei denen innerhalb eines Tests verschiedene Aufgabenformate zusammengefasst präsentiert werden (skills-based measures). Ebenso gibt es verschiedene Antwortformate. Je nach Aufgabe müssen jeweils Zahlen oder Operatoren ergänzt werden, manchmal müssen Zuordnungen von vorgegebenen Lösungen und Aufgaben vorgenommen werden. Für einige Aufgaben sind mehrere Antwortmöglichkeiten vorgegeben, aus welchen der Schüler die richtige auswählen muss (multiple choice), dabei sind neben der richtigen Antwort weitere, jedoch falsche Antwortmöglichkeiten (Distraktoren) abgebildet. Insgesamt sind die Tests sprachlastiger, da die Problemstellungen an kurzen Texten erläutert werden. In der Regel haben die Kinder sechs bis acht Minuten zur Bearbeitung eines Tests zur Verfügung (Hosp et al.,

2007). Ein Beispiel für ein solches CBM ist in Abbildung 18, Abbildung 19 und Abbildung 20 dargestellt.



Name Ben Date March 20 Test 13 Page 1

Column A — Applications 2 — Column B

(1)
Write the answer in the blank.

Larry spends 31¢ at the toy store.
Paul spends 43¢ more than Larry.
How much money does Paul spend?

31
+43
74

74¢ ✓

(2)
Write the number in the blank.

✓ 7 + 2 = 2 + 7

(3)
Write the time.

1 : 15
✓ ✓

(4)
Counting by 3's, fill in the blanks.

45, 48, 51, 52, 53

(5)
How much money?

$1.02

(6)
Hours of Ball Practice

Jordan	3
Kimuli	6
Ebony	5

Each 1 means 1 hour of practice

Write the number in each blank.

How many more hours does Kimuli practice ball than Ebony? 1 ✓

How many hours does Jordan practice ball? 3 ✓

How many fewer hours does Jordan practice ball than Ebony? 3

(7)
Fill in the blanks.

105 =
1 ✓ hundreds 0 ✓ tens 5 ✓ ones

Abbildung 18: Beispielseite 1 eines CBM des Typs concepts & application nach L.S. Fuchs und D. Fuchs (2007) (entnommen aus: L.S. Fuchs & D. Fuchs, 2007, S. 25)

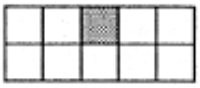
Name Ben Date _____ Test 13 Page 2

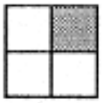
Column C — Applications 2 — Column D

(8)

Write the letter of the matching fraction in each blank.

✓ B

✓ D 

A 

(A) $\frac{1}{3}$

(B) $\frac{1}{2}$

(C) $\frac{1}{4}$

(D) $\frac{1}{10}$

(9)

Write + or - in the blank.

9 + 6 = 15 ✓

(10)

Write the number in the blank.

Of these numbers,

79 73 64

✓ 64 is the smallest.

✓ 79 is the largest.

(11)

Counting by 2's, fill in the blanks.

8, 10, 12, 14 ✓, 15

(12)

Write the number in the blank.

10 + 1 = 1 + 12

(13)

Write the answer in the blank.

There are 13 white mice in the pet store and 14 gray mice. How many mice are there in all?

13 + 14 = 26 26

(14)

Look at this group of numbers.

1 2 3 4 5 6 7

8 9 10 11 12 13 14

15 16 17 18 19 20

Write the third number. 3 ✓

Write the fifteenth number. 15 ✓

Write the twentieth number. 20 ✓

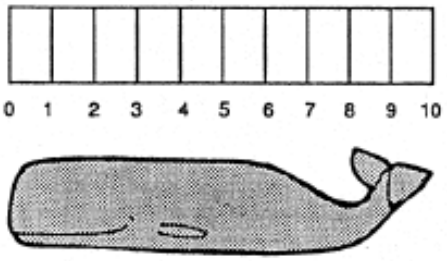
Abbildung 19: Beispielseite 2 eines CBM des Typs concepts & application nach L.S. Fuchs und D. Fuchs (2007) (entnommen aus: L.S. Fuchs & D. Fuchs, 2007, S. 25)

Name _____ Date _____ Test 13 Page 3

Column E — Applications 2 — Column F

(15)

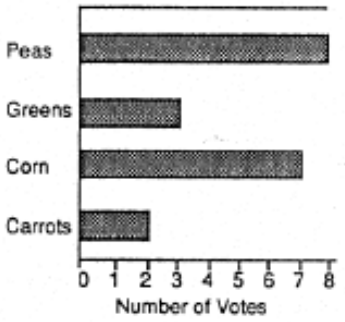
How long is the whale?



10 units ✓

(16)

Favorite Vegetables



Write the number in each blank.

How many more votes did corn get than carrots? 2

How many votes did greens get? 3 ✓

How many votes did greens and peas get together? 11 ✓

+3 8 / 11

(17)

Write + or - in the blank.

10 ___ 7 = 3

(18)

December

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	(9)	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

Holiday play

Write the letter in the blank.

The holiday play falls on which day of the week?

(A) Monday
(B) Wednesday
(C) Tuesday

___

Abbildung 20: Beispielseite 3 eines CBM des Typs concepts & application nach L.S. Fuchs und D. Fuchs (2007) (entnommen aus: L.S. Fuchs & D. Fuchs, 2007, S. 25)

3.2.5 Verfügbare CBM-Verfahren im deutschsprachigen Raum

Die Idee der wiederholten Leistungsmessung zur Abbildung des Lernverlaufs von Schülern im Sinne der formativen Diagnostik ist in der deutschsprachigen Literatur kein Novum (z. B. Klauer, 1978). Schon in den 1990er Jahren veröffentlichte bspw. Klauer seine Diagnose- und Förderblätter (Klauer, 1994a, 1994b, 1994c). Mit ihnen sollte eine an testtheoretischen Gütekriterien orientierte Lernverlaufsmessung für den Bereich Mathematik der zweiten bis vierten Klasse ermöglicht werden. Aber auch Leistungsmessungen im Sinne der curriculum-based measurements sind in den vergangenen Jahren in deutschsprachigen Raum vorgedrungen. So war ein Artikel von Klauer (2006) die erste Veröffentlichung zu diesem Thema in Deutschland. Darauf folgte eine Welle von Forschungsarbeiten, welche sich unabhängig voneinander entwickelten (Diehl & Hartke, 2007; Diehl, Hartke & Knopp, 2009; Klauer, 2011; Müller & Hartmann, 2009, 2010; Souvignier & Förster, 2011; Souvignier, Förster & Salaschek, 2010; Strathmann & Klauer, 2008, 2010; Strathmann, Klauer & Greisbach, 2010; Walter, 2008a, 2008b, 2009a, 2009b, 2010a, 2010b, 2011a, 2011b). In der deutschsprachigen Literatur etablierten sich die Begrifflichkeiten curriculumnahe Lernfortschrittsmessungen (Diehl & Hartke, 2007), Lernverlaufsdiagnostik (Strathmann & Klauer, 2008, 2010) oder auch curriculumbasiertes Messen bzw. Lernprozessbegleitende Diagnostik (Walter, 2008b, 2009a, 2009b), welche im Prinzip synonym behandelt werden können.

Auch in Deutschland wurden erste Versuche gestartet, CBM in Anlehnung an US-amerikanische Vorbilder zu entwickeln. Im Zuge diverser Studien entstand eine Reihe von Verfahren, welche zur Lernfortschrittsdokumentation in verschiedenen schulischen Domänen konzipiert und zum großen Teil auch systematisch auf ihre Güte und Wirksamkeit hin geprüft wurden. So veröffentlichte Walter

(2010a) die Lernfortschrittsdiagnostik Lesen (LDL), ein Verfahren zur wiederholten Erhebung der Leseflüssigkeit im Sinne curriculumbasierter Messungen von Schülern im Grund-, Haupt- und Förderschulbereich. Der Lernverlauf wird dabei über die jeweils innerhalb einer Minute Bearbeitungszeit korrekt gelesenen Wörter in einem von 28 verschiedenen parallelen Lesetexten erhoben. Dabei konnte Walter hohe Reliabilitäten (Paralleltest-Reliabilität bei Grundschülern: r = .90, bei Hauptschülern: r = .80) nachweisen. Auch die Gültigkeit konnte durch die Analyse der Höhe korrelativer Zusammenhänge zwischen den entwickelten CBM und weiteren Kriteriumsmaßen nachgewiesen werden (Walter, 2008b). Darüber hinaus erforschte Walter (2009b) ein weiteres CBM zur Erfassung der Lesekompetenz hinsichtlich seiner Testgüte. In dem Verfahren sollen die Kinder innerhalb einer Minute Auslassungen in einem Lückentext durch Auswahl des passenden von jeweils drei vorgegebenen Wörtern ergänzen (die sog. MAZE-Technik). Die Anzahl der richtig ergänzten Auslassungen werden bepunktet. Es wird dabei nicht die Leseflüssigkeit, sondern das Leseverständnis als Indikator für die Lesekompetenz erfasst. Dies erscheint gerade in höheren Klassenstufen relevant, wo neben der Lesetechnik die inhaltliche Erfassung des im Text gegebenen Kontextes eine wesentliche Rolle spielt. Im Vergleich zu den oben genannten Verfahren zur Leseflüssigkeit konnten für die MAZE-CBM in der Untersuchung keine so hohen Kennwerte zur Testgüte erzielt werden. „Ähnlich wie in den anglo-amerikanischen Untersuchungen lieferte die MAZE-Technik Werte mit zufriedenstellender Validität sowie Änderungssensibilität (Walter, 2009b). Die Paralleltest-Reliabilität der verwendeten Texte von .52 bis .75 liegt aber deutlich unter den Erwartungen“ (Walter, 2009b, S. 62). So ist weitere Forschung in diesem Bereich indiziert (Walter, 2009b).

Strathmann und Klauer (2008, 2010) arbeiten an einem Verfahren zur Diagnostik des Lernverlaufs der Rechtschreibkompetenzen im

Grundschulbereich (Klasse zwei bis vier). Die Besonderheit dabei ist, dass dazu nicht zuvor erstellte Diktattexte verwendet werden. Hingegen stellt ein Zufallsgenerator die Tests aus einer zuvor definierten Grundmenge von insgesamt 480 Items zusammen. Diese Grundgesamtheit der Wörter gestaltet sich als Schnittmenge dreier in der Literatur erwähnten Grundwortschatz-Listen (Augst, 1989; Naumann, 1999; Plickat, 1987). Nach diesem Prinzip werden jeweils 20 zufällig gewählte Wörter zu einem Diktat zusammengefasst. Erste Studien haben jedoch gezeigt, dass der erstellte Itempool für das CBM zur Erfassung der Rechtschreibkompetenzen nicht homogen genug ist. So konnten nicht in jedem Fall die erwarteten Gütekriterien nachgewiesen werden (Strathmann & Klauer, 2008, 2010).

Von Strathmann und Klauer (2012) wurden überdies Verfahren zur Erhebung der Rechenkompetenzen von Grundschülern der 2. bis 4. Klasse konzipiert und auf ihre Güte hin überprüft: die Lernverlaufsdiagnostik – Mathematik für zweite bis vierte Klassen (LVD-M 2-4). Die Autoren nutzten dabei eine modifizierte Version des in Zusammenhang mit den CBM zur Rechtschreibung beschriebenen Itemsamplings, um im Vergleich zu den Rechtschreibverfahren eine höhere Testgüte zu erzielen. Die Testitems wurden den zuvor erwähnten Diagnose- und Förderblättern von Klauer (1994a, 1994b, 1994c) entnommen, welche schuljahresspezifische Aufgaben bezüglich der vier Grundoperationen Addition, Subtraktion, Multiplikation und Division enthalten. Die Aufgaben wurden nach vorgegebenen Kriterien in einzelne nach Schwierigkeitsgrad getrennte Teilmengen zerlegt, aus welchen zufällig jeweils in bestimmten Verhältnissen Items ausgewählt und zu einem Test zusammengefasst wurden. Dieser Prozess wird mittels Computer bewerkstelligt. Für die somit entstandenen Testverfahren zur Erfassung der Rechenkompetenzen konnten in der Studie von Strathmann und Klauer (2010) akzeptable Reliabilitätskoeffizienten von $r = .77$ (zwischen $r = .66$ und $r = .91$) bzw. $r = .73$ (zwischen

r = .56 und r = .88), abhängig von der Art der Reliabilitätsbestimmung (Paralleltest-Reliabilität bzw. Splithalf-Reliabilität) ermittelt werden. Auch Analysen hinsichtlich der Homogenität der Testschwierigkeiten sowie zur Validität der LVDM 2-4 liefern befriedigende Ergebnisse (Klauer, 2011; Strathmann & Klauer, 2012).

Parallel erarbeiteten Müller & Hartmann (2010) die „Freiburger Aufgaben zur Lernfortschrittsdiagnostik (CBM) Mathematik“, welche ebenso der Lernverlaufsdiagnostik der Rechenfähigkeiten von Grundschülern der Klassenstufe 1 bis 4 ermöglichen sollen. Sie fokussieren auf Aufgaben zur Erfassung der Fertigkeiten bezüglich Addition, Subtraktion, Multiplikation und Division. Die Verfahren befinden sich derzeit noch in einer Evaluationsphase und sind noch nicht veröffentlicht.

Die Forschungsgruppe um Souvignier an der Universität Münster konzipiert seit 2009 Verfahren zur Dokumentation des Lernverlaufs im Sinne von CBM. Um den Einsatz dieser Tests weitgehend einfach zu gestalten, werden sie internetbasiert durchgeführt und ausgewertet. Mittlerweile ist ein beachtliches Inventar an Tests für den Bereich des Lesens (Klassenstufe 3 bis 6), der Mathematik (Klassenstufe 1 bis 3) und den Englischunterricht (Klassenstufe 5) erstellt und evaluiert worden. In einem ersten Forschungsbericht zur Güte der Testverfahren und zur Effektivität des Systems („quop“) kommen Souvignier et al. (2010) zu folgendem Schluss:

> *„Die Testkonzepte für das Lesen und für die Mathematik weisen eine hohe Testgüte auf, es resultieren höhere Lernzuwächse in den Klassen, die das quop-System nutzen und die Rückmeldungen der Lehrkräfte weisen auf eine hohe Akzeptanz und eine aktive Nutzung der diagnostischen Information hin“* (S. 4).

Speziell für die Mathematiktests der Klassenstufe 1, welche neben Mengen- und Zahlenwissen sowie Rechenfähigkeit auch Arbeitsgedächtnisleistungen abfragen, ergaben sich zufriedenstellende Reliabilitäts- und Validitätskennwerte. Auch die Parallelität sowie die Sensibilität der Verfahren gegenüber Leistungsentwicklungen konnten nachgewiesen werden (Souvignier et al., 2010). Bei den Tests müssen die Kinder Mengen und Größen in Bildern bestimmen, Zahlen identifizieren, Zahlenfolgen vervollständigen sowie einfache Rechenaufgaben lösen. Die Aufgabenstellungen werden aufgrund noch nicht genügend ausgeprägter Lesekompetenzen von Erstklässlern auditiv via Kopfhörer dargeboten. Die Ergebnisse werden von den Kindern direkt in den Computer eingegeben (Souvignier et al., 2010).

In den letzten Jahren ist auch in Deutschland ein erhöhtes Forschungsinteresse hinsichtlich Testverfahren in Anlehnung an die US-amerikanischen CBM zu beobachten. So konnten zeitgleich zur Erarbeitung der vorliegenden Arbeit für die schulischen Fachbereiche Deutsch und Mathematik verschiedene CBM-Verfahren vorgestellt werden (Souvignier et al., 2010; Strathmann & Klauer, 2012; Walter, 2010a). Dennoch befindet sich die systematische Erforschung vorhandener sowie auch weiterer Verfahren aktuell noch in den Kinderschuhen. Wünschenswert ist der Ausbau von Verfahren speziell im mathematischen Bereich, da vorhandene CBM einige Einschränkungen in ihren Einsatzmöglichkeiten aufweisen. So gibt die LVD-M 2-4 (Strathmann & Klauer, 2012) ausschließlich Auskunft über die Rechenleistungen der Kinder, nicht aber über ihre ihnen verfügbaren basalen mathematischen Kompetenzen. Darüber hinaus ist ein Einsatz erst ab Klassenstufe 2 möglich. Erste Erfahrungen zum praktischen Einsatz der LVD-M 2-4 im Rahmen des Rügener Inklusionsmodells (vgl. Abschnitt 3.1.4.4) deuten auf eine nur bedingte Einsatzmöglichkeit des Verfahrens zur Lernverlaufsdokumentation in relativ kurzen (hier: monatlichen)

zeitlichen Abständen. Die von Souvignier et al. (2010) zeitlich parallel entwickelten zur vorliegenden Studie CBM hingegen fragen auch grundlegende mathematische Kompetenzen ab und sind somit ebenso für den Einsatz in Klassenstufe 1 geeignet. Allerdings können diese ausschließlich am Computer bearbeitet werden. Zwar wird dadurch der Aufwand hinsichtlich der Auswertung und Interpretation der Ergebnisse minimiert, jedoch tendieren Lehrkräfte oft eher zur Papierversion eines Tests (Shapiro, Edwards & Zigmond, 2005). Zudem ist nicht abschließend geklärt, ob die Durchführung der Tests am Computer Auswirkungen auf die erzielten Ergebnisse der Erstklässler hat. Hierzu sollten die Ergebnisse der Online-Durchführung mit denen einer Durchführung als Paper-and-Pencil-Form vergleichend analysiert werden. Derartige Analysen hinsichtlichd er von Souvignier et al. (2010) entwickelten CBM-Verfahren sind derzeit nicht bekannt. Die innerhalb der hier vorliegenden Arbeit konzipierten CBM-Verfahren sollen diese hier aufgezeigten Einschränkungen in den Einsatzmöglichkeiten derzeit verfügbarer Verfahren im deutschsprachigen Raum umgehen. Anhand der Verfahren soll eine Dokumentation von Lernfortschritten basaler arithmetischer Kompetenzen sowie erster Rechenkompetenzen ab Beginn der ersten Klasse ermöglicht werden. Dazu werden die Kurztests als Papierversion konzipiert, um Verzerrungen der Schülerergebnisse durch unterschiedlich ausgeprägte Kompetenzen der Kinder im Umgang mit dem Computer auszuschließen. Schließlich sollen die CBM-Verfahren möglichst sensibel gestaltet sein, um auch kurzfristige Rückmeldung zur mathematischen Leistungsentwicklung der Schüler, bspw. innerhalb eines Zeitraumes von vier Schulwochen zu ermöglichen.

3.2.6 Bedeutung von CBM-Verfahren

CBM bieten eine Reihe an Vorzügen, welche für die Unterrichtspraxis relevant sind. Diese wurden an verschiedenen Stellen bereits benannt und sollen im folgenden Abschnitt zusammengefasst werden. Vor allem durch die einfache Durchführung und Auswertung sind CBM gut in den schulischen Alltag zu integrieren (Deno, 1985), was in Abschnitt 3.2.3 gezeigt wurde. Der wiederholte Einsatz der Messungen ermöglicht es, den Lernverlauf eines Schülers nachzuvollziehen. Da die CBM sensitiv auf Veränderungen reagieren, kann Lernfortschritt bzw. -stagnation schnell festgestellt werden und ein zeitnahes Reagieren auf unzureichende Lernentwicklungen wird ermöglicht (Deno, 1985; L.S. Fuchs & D. Fuchs, 2005; Hosp et al., 2007). Entgegen des schlichten unsystematischen Beobachtens der Schülerleistungen lassen standardisierte Verfahren wie CBM einen objektiven Blick auf die Lernentwicklung zu und die Wahrnehmung der Leistungen der Schüler durch die Lehrkraft wird präziser (Deno, 1985). Dabei ist es möglich den Fortschritt des Einzelschülers aber auch der gesamten Klasse zu analysieren. Die Lehrkraft erhält Informationen zum Handlungserfolg des Unterrichts bzw. einer Förderung und kann pädagogische Maßnahmen hinsichtlich ihrer Wirksamkeit evaluieren. Darüber hinaus liefern CBM Hinweise, welches die Problembereiche eines Schülers sind (Hosp et al., 2007). Solche Informationen können genutzt werden, um den Unterricht durch gezielte Maßnahmen zu optimieren, sodass ungenügend ausgeprägte Kompetenzen gezielt gefördert werden können (Deno, 1985).

Jedoch nicht nur die Lehrkraft erhält Rückmeldungen über den Unterricht. Die grafische Aufbereitung der Lernverlaufsdaten ermöglicht einen anschaulichen Überblick über die Leistungsentwicklung des Schülers und bietet eine gute Kommunikationsgrundlage (Deno, 1985; L.S. Fuchs & D. Fuchs, 2005). Der einzelne Schüler kann anhand der grafischen

Aufbereitung den eigenen Leistungsfortschritt nachvollziehen. Er erhält Einblick in Ursache-Wirkungszusammenhänge beim Lernprozess, bspw. eine gesteigerte Punktzahl durch Üben. Das solche selbstevaluierenden Maßnahmen durch die Schüler positive Effekte auf deren schulische Leistung haben, ist in Abschnitt 3.1.1 bereits näher erläutert worden (Black & Wiliam, 2009; CCSSO, 2008a; Harris, 2007). In Gesprächen mit bspw. Eltern, Sonderpädagogen oder weiteren Fachleuten können solche Diagramme, ähnlich wie sie in Abbildung 16 oder Abbildung 17 dargestellt sind, genutzt werden, um den Entwicklungsverlauf eines Kindes schnell und einfach sichtbar sowie verständlich zu machen (Deno, 1985). Darüber hinaus ist es möglich, durch weitere Vergleichsdaten wie bspw. den Klassendurchschnitt (auch ein Jahrgangsdurchschnitt bzw. Durchschnitt einer Normstichprobe sind denkbar) eine solche Übersicht zu erweitern. Die CBM-Daten gestatten somit eine präzise Einschätzung der vorliegenden Schülerleistungen. Dadurch ist eine Beurteilung auf Grundlage einer sozialen (durch den Vergleich mit den Leistungen anderer Kinder), einer kriterialen (durch das Überprüfen, ob ein zuvor festgelegtes Lernziel erreicht wurde) aber auch einer individuellen Bezugsnorm (durch Analyse der Leistungsentwicklung des Schülers über die Zeit) möglich. Eine computergestützte Auswertung kann darüber hinaus die Arbeit mit CBM noch weiter vereinfachen (L.S. Fuchs & D. Fuchs, 2005).

3.2.7 Grenzen von CBM-Verfahren

So vielversprechend der Einsatz von CBM auch erscheint, es gibt auch einige limitierende Aspekte, welche nachfolgend beleuchtet werden sollen. Diese beziehen sich im Wesentlichen auf die Kompetenzen der Lehrkräfte, die Einstellungen der Lehrkräfte und die Verfügbarkeit von entsprechenden Messverfahren.

Obwohl CBM für einen einfachen Einsatz in der Schule konzipiert sind, ist ein grundlegendes Theoriewissen über diagnostische Maßnahmen im Allgemeinen Voraussetzung für die richtige Anwendung der Verfahren. So muss dem Testleiter zunächst die Bedeutung einer standardisierten Durchführung und Auswertung von Testverfahren zur Sicherung der Objektivität und Gewährleistung reliabler und valider Daten bewusst sein. Diese Sensibilisierung weisen viele Lehrkräfte jedoch noch eher wenig auf. Dies zeigte L.S. Fuchs (1982) in einer Studie mit 147 Sonderpädagogen. Diese wurden hinsichtlich ihrer Vorgehensweise zur Ermittlung von Schülerleistungen befragt. Dabei stellte sich heraus, dass 80 % der Probanden informelle Beobachtungen, 20 % selbst erstellte Übungen (z. B. Spiele oder Lernkarteien) und lediglich 5 % Testverfahren zur Leistungseinschätzung der Schüler nutzten. Hingegen waren sich 95 % von ihnen sicher, dass sie dadurch zu validen Beurteilungen der Schüler gekommen sind. L.S. Fuchs (1982) konnte allerdings nachweisen, dass ein sehr hoher Anteil der untersuchten Personen einer Fehleinschätzung unterlag.

Auch Sandall, Schwartz und LaCroix (2004) berichten in ihrer Studie zur Einstellung von Förderlehrern hinsichtlich der Datenerhebung als Grundlage für pädagogische Entscheidungsfindungen über Hinweise, dass die meisten Lehrkräfte unsystematische Beobachtungen zur Schülerbeurteilung bevorzugt einsetzen. In ihren Entscheidungen stützen sie sich in der Regel nicht auf standardisiert erhobene Daten. Dieses Vorgehen wird von den Autoren vor allem auf die Einstellung der Lehrkräfte zurückgeführt, dass erhobene Daten als nicht so bedeutsam angesehen werden, da diese ihrer Meinung nach nicht repräsentativ die Schülerfähigkeiten widerspiegeln können. Weiterhin berichten Sandell et al. (2004), dass vor allem die Datenverwaltung sowie eingeschränkte zeitliche Ressourcen den Lehrkräften Probleme bereiten. Viel bedeutsamer erscheint laut den Autoren an dieser Stelle jedoch die begrenzte

Fähigkeit zur Interpretation der erhobenen Daten und deren Verwendung als Grundlage zur Entscheidungsfindung für weitere Fördermaßnahmen (Sandell et al., 2004). Auch Hojnoski, Gischlar und Missall (2009) resümieren: „However, research suggests that lack of skill in interpreting data is a barrier to frequent and systematic use of data, despite the general belief in its importance“ (S. 3). Eine Sensibilisierung testtheoretischer Grundlagen von allgemeinbildenden und sonderpädagogischen Lehrkräften erscheint demnach insgesamt notwendig, um eine breitere Akzeptanz sowie ein tieferes Verständnis für den Umgang mit und die Interpretation von Daten zu ermöglichen. Diese Problematik muss auch beim regelmäßigen Einsatz von CBM-Verfahren im Unterrichtsalltag beachtet werden.

Das bloße Sammeln und Organisieren von Daten führt jedoch noch nicht zur Verbesserung von Schülerleistungen (L.S. Fuchs & D. Fuchs, 1986). Der Einsatz von CBM ersetzt nicht den Unterricht oder die Förderung, sondern bildet nur die Grundlage für pädagogische Entscheidungen (Hosp et al., 2007). Die Auswahl und Durchführung fördernder Maßnahmen obliegt allein der Lehrkraft. Die Frage ist demnach, welche Hinweise diesen Daten, hinsichtlich der Leistungen und Bedürfnisse einzelner Schüler, entnommen werden können und wie sie zu interpretieren sind. Lehrkräfte müssen über Kompetenzen zur Ableitung „adaptiver“ Lerngelegenheiten auf Basis erhobener Schülerleistungen verfügen. In den Daten von Bamonto, Graney und Shinn (2005) spiegelt sich jedoch wider, dass hier noch weiterer Schulungsbedarf besteht. In ihrer Studie haben die Autoren die Effekte verschiedener mittels CBM erhobenen Arten des Feedbacks für Lehrkräfte (N = 44) auf die Leseleistungen der unterrichteten Schüler (N = 184) untersucht. In einem Kontrollgruppendesign bekamen die Grundschullehrkräfte einer zweiten Klasse entweder ein aggregiertes Feedback über die Leistungen der gesamten Klasse, ein individuelles Feedback für einen zufällig ausgewählten Schüler der

Klasse oder gar kein Feedback. Die Leseleistungen der Zweitklässler wurden dabei wöchentlich mittels CBM erhoben. Die Ergebnisse der Studie bestätigen die Annahme nicht, dass Schülerleistungen durch eine bloße Rückmeldung an Grundschullehrkräfte in Form eines Lernverlaufsdiagramms unter Angabe des Leistungsanstiegs – gleich, ob für die Gruppe oder für individuelle Schüler – gesteigert werden können, hingegen kommen Bamonto et al. (2005) zu dem Schluss:

> *„The principal implication of this study's results for practice is that, to use student progress data effectively, general education teachers need more than feedback in the form of a graph and a numerical summary of slope of improvement accompanied by general recommendations"* (S. 198).

Ebenso sind Stecker, L.S. Fuchs und D. Fuchs (2005) der Auffassung, dass die Lehrkräfte sehr viel und vor allem auch fortlaufende Unterstützung brauchen, um effektiv mit CBM arbeiten zu können. Auch L.S. Fuchs et al. (1994) kommen mit ihrer Studie zu dem Ergebnis, dass der Einsatz von CBM in der Schule dann am effektivsten ist, wenn die Lehrkräfte nicht nur zusammengefasste Berichte über die Schülerergebnisse erhalten, sondern darüber hinaus auch konkrete Hinweise für die Ausgestaltung von Unterrichts- und Fördereinheiten vorgeschlagen bekommen. Auch klar definierte systematische Regeln bezüglich der Auswertung der Daten und in Hinblick sich anschließender Maßnahmen tragen zur Steigerung der Schülerleistungen bei (Farlow & Snell, 1989; L.S. Fuchs, Deno & Mirkin, 1984; Haring, Liberty & White, 1980). Ganz essentiell erscheint es dabei, dass Unterstützungsmaßnahmen auf die individuellen Bedürfnisse der Schüler zugeschnitten werden, speziell bei Kindern mit Problemen beim Lernen (Stecker & L.S. Fuchs, 2000). Baker und Zigmond (1990) stellten fest, dass, während Sonderpädagogen sich für die individuellen Bedürfnisse der Kinder verantwortlich zu zeigen haben, der Fokus von

Grundschulpädagogen eher auf der Klasse als Gesamtheit liegt und weniger auf dem einzelnen Schüler. In Verbindung mit dem Einsatz von CBM zur Steigerung der Schülerleistungen muss demnach ein Perspektivwechsel stattfinden. Als Grundlage für einen sinnvollen Einsatz von CBM gilt es dem Anwender, ganz gleich ob Sonderpädagoge oder Regelschullehrer, Verantwortung für die individuellen Bedürfnisse jedes einzelnen Kindes, auch derer mit Lernschwächen, zu tragen.

Ein weiteres Problemfeld zeigt sich in den technischen Kompetenzen der Lehrkräfte. So nutzen laut Angaben von Shapiro et al. (2005) nur wenige Lehrer computerbasierte Verfahren zur Generierung von Lernverlaufsdiagrammen. Die meisten erstellen diese selbst mittels Stift und Papier. Auch wenn dies durchaus eine Möglichkeit zur Veranschaulichung der Daten ist, so kann durch den Einsatz computer- bzw. internetbasierter Technologien große Zeitersparnis erzielt werden (Shapiro et al., 2005). In diesem Zusammenhang konstatieren auch Stecker et al. (2005):

> „*Computer applications facilitated the use of decision rules and included a goal-raising feature that also stimulated student growth.* […] *Moreover, computer programs that incorporated student data collection as well as data management reduced teacher time devoted to CBM procedures and enhanced feasibility and utility of the system"* (S. 816).

L.S. Fuchs (2004) sieht in der Konzeption und Erforschung computergestützter Systeme zur Verarbeitung von CBM-Daten noch deutlichen Handlungsbedarf. Ziel sollte es sein, Programme zu entwickeln, die eine benutzerfreundliche Datensammlung und -verwaltung und darüber hinaus Auswertungen in Form von grafischen Ausgaben und Schülerprofilen ermöglichen, aus welchen man ohne große Schwierigkeiten Hinweise zur Unterrichts- und Fördergestaltung entnehmen kann.

Ein weiterer Problemaspekt ergibt sich in der Auswahl an verfügbaren CBM-Verfahren für den Mathematikunterricht. Während für den Bereich Lesen ein umfangreiches Netzwerk an Studien und Verfahren zum Thema CBM zu finden ist, befindet sich die Konzeption und Erforschung von CBM-Verfahren für den mathematischen Bereich noch am Anfang (Foegen et al., 2007; Jiban & Deno, 2007). Speziell für die höheren Klassenstufen der US-amerikanischer Schulen existieren keine oder nur wenige CBM für den Bereich Mathematik. L.S. Fuchs (2004) gibt zudem zu bedenken, dass sich die Erforschung von CBM-Verfahren oftmals ausschließlich auf die Bestimmung der wissenschaftlichen Gütekriterien Objektivität, Reliabilität und Validität beschränken. Damit sind Aussagen hinsichtlich der Anwendbarkeit der Tests als statische Messungen im Sinne eines Screenings möglich, es können jedoch keine Hinweise entnommen werden, wie sensibel die Verfahren die Leistungsentwicklung der Schüler im Laufe der Zeit erfassen können. Gerade letzteres ist jedoch der entscheidende Aspekt, wenn CBM als Lernverlaufsdiagnostikum eingesetzt werden soll. Die Forschung muss sich also, so L.S. Fuchs (2004), auch mit Fragen in Hinblick auf die Sensibilität, aber auch Effektivität ihres Einsatzes befassen. Drei Jahre nach L.S. Fuchs Postulat können Foegen et al. (2007) jedoch noch immer keine wesentliche Verbesserung der Forschungssituation erkennen. „Only a single mathematics measure (MBSP computation) has evidence supporting its use to improve student achievement for students with disabilities" (Foegen et al., 2007, S. 138). Dies mag an der Besonderheit des Mathematiklernens in Relation zum Lesenlernen liegen. Lesen ergibt sich im Wesentlichen aus dem Dekodieren von Texten in Sprache. Lesens einmal verstanden, kann man es auf weitere Lesevorlagen anwenden. Der Lernfortschritt zeigt sich dann in der Leseflüssigkeit bzw. -geschwindigkeit (Foegen et al., 2007; Walter, 2008b). Der Bereich der Mathematik hingegen besteht aus verschiedenen, immer komplexer werdenden Inhalten, wodurch das Erlernen der

Mathematik sich bis in hohe Klassenstufen vollzieht und nicht einfach eine Kompetenz ist, welche als Indikator für die Mathematikleistung insgesamt steht, extrahiert werden kann (Foegen et al., 2007). Gerade aus diesem Grund sollte man sich in Zukunft verstärkt darum bemühen, CBM für den Bereich Mathematik zu entwickeln und zu evaluieren.

> *„As instructional practices become more integrated with data and permit more sophisticated adult responses to accelerate child learning, it will be important to develop conceptually sound and technically sound systems of assessment that efficiently inform instructional decision making. Fluency in computation and operations may play an important role as an efficient screening within such a system“* (VanDerHeyden & Burns, 2008, S. 223).

Im nachfolgenden Abschnitt sollen aktuelle Forschungsergebnisse hinsichtlich der Konzeption und des Einsatzes von CBM-Verfahren näher beleuchtet werden soll.

3.2.8 Forschungslage und -ergebnisse hinsichtlich der Konzeption und des Einsatzes von CBM-Verfahren im Bereich Mathematik

Trotz einer nun über 30jährigen Forschungstradition sowie der in Abschnitt 3.1.3 und 3.2.6 geschilderten Vorteile des regelmäßigen Einsatzes von Tests als Evaluationswerkzeug zur Optimierung von Unterricht und Förderung, sind noch nicht alle Fragen bezüglich der Verwendung curriculumbasierter Messverfahren im schulischen Alltag zufriedenstellend beantwortet. In Abschnitt 3.2.7 wurden die Grenzen bei der Konzeption und beim Einsatz von CBM dargelegt, welche die Forschung, insbesondere bezogen auf CBM-Verfahren im mathematischen Bereich, beschäftigen. Während hinsichtlich der Entwicklung von Lesekompetenzen schon viele Studien zur Güte

curriculumbasierter Messungen sowie zur Effektivität ihres Einsatzes, inklusive mehrerer Metaanalysen, vorliegen (Colón & Kranzler, 2006; L.S. Fuchs et al., 1984; L.S. Fuchs, D. Fuchs & Hamlett, 1989b; Jones & Krause, 1988; Shapiro, Keller, Lutz, Edwards Santoro & Hintze, 2006; Wayman, Wallace, Wiley, Ticha & Espin, 2007; Wesson, 1991; Wiley & Deno, 2005; Yeo, 2010), fällt die Forschungslage im Bereich der Mathematik deutlich geringer aus (Christ, Scullin, Tolbize & Jiban, 2008; Foegen et al., 2007; Lembke & Stecker, 2007). Der aktuelle Forschungsstand zur Entwicklung von CBM-Verfahren für diese schulische Domäne wurde in jüngster Vergangenheit von den Forschungsgruppen Foegen et al. (2007) sowie Christ et al. (2008) zusammengefasst. Die beiden Artikel thematisieren die wissenschaftliche Güte der CBM speziell im Bereich der Mathematik. Darüber hinaus befassen sich Foegen et al. (2007) mit Angaben zur Effektivität des Einsatzes von Rechen-CBM im schulischen Alltag.

Foegen et al. (2007) konnten im Rahmen ihrer Literaturanalyse 32 relevante Artikel extrahieren, welche von 25 unterschiedlichen Studien unter verschiedenen Testbedingungen berichten. So streut die Klassenstufe, der in den Studien untersuchten Probanden vom Kindergarten bis hin zur achten Klasse. Zudem umfassen die Untersuchungen Probanden verschiedener Schultypen. Darüber hinaus finden sich Studien zu verschiedenen Aspekten der Mathematik, wie erste mathematische Kompetenzen und Einsichten (early numeracy), Rechenkompetenzen (computation) als auch weiterführende mathematische Konzepte und Problemlösekompetenzen (concepts & application). Der Großteil der analysierten Studien beschäftige sich mit der Erforschung von CBM im Grundschulbereich (N = 17). Eine deutlich kleinere Anzahl an Untersuchungen lag für das Gebiet der Mathematik im Vorschulbereich (N = 4) sowie im Bereich der weiterführenden Schule (N = 4) vor.

Christ et al. (2008) beziehen ihre Untersuchung ausschließlich auf CBM-Verfahren des Typs computation. Dabei beschränken sie sich auf Forschungsergebnisse, welche peer-reviewed nach dem Jahr 1989 veröffentlicht wurden. Die ausgewählten Studien unterscheiden sich hinsichtlich der Klassenstufe der untersuchten Kinder sowie des Schultyps. So umfassen die Untersuchungen Ergebnisse von Regelschul- und Sonderschulkindern der Klassenstufe 1 bis 9. Im Folgenden sollen ausgewählte Ergebnisse näher beleuchtet werden. Dabei wird eine Einschränkung auf Verfahren für den Vorschul- und Grundschulbereich vorgenommen, da die Forschungsintention dieser Arbeit sich auf diesen Entwicklungsabschnitt konzentriert. Zunächst werden die Ergebnisse hinsichtlich der Reliabilität der in den Studien untersuchten CBM-Verfahren diskutiert und anschließend hinsichtlich ihrer Validität. Danach sollen ausgewählte Studien zur Effektivität des Einsatzes von CBM im Unterricht zur Steigerung von Schülerleistungen vorgestellt werden.

3.2.8.1 Forschungsergebnisse hinsichtlich der Reliabilität verfügbarer CBM-Verfahren im Bereich Mathematik

Die von ihnen analysierten Studien zusammenfassend, kommen Christ et al. (2008) zu dem Ergebnis: „It is often written within the professional literature that CBM is a valid and reliable set of measurement procedures or that CBM is a sensitive measurement procedure. This type of statement is incomplete and potentially misleading" (S. 204). Zu diesem Resultat führt unter anderem die dürftige Forschungslage bezüglich der Reliabilität der untersuchten CBM-Verfahren. In keiner der von den Autoren analysierten Studien wurden Angaben hinsichtlich der Test-Retest-Reliabilität der untersuchten Tests angeführt. Lediglich zwei Studien bezogen sich auf die Paralleltest-Reliabilität (dargestellt in Tabelle 5).

Tabelle 5: Übersicht über Studien zur Paralleltest-Reliabilität von CBM-Verfahren im Bereich Mathematik (entnommen aus: Christ et al., 2008, S. 202)

	Probanden			Paralleltest-reliabilität	
Studie	Stichprobe	Klassenstufe	N	Min.	Max.
Allinder, L.S. Fuchs, D. Fuchs & Hamlett (1992)	Regelschule	2 – 5	95	.83	.93
Thurber et al. (2002)	Regelschule	4	207	.90	.92

Zwei weitere Untersuchungen beschäftigten sich mit der internen Konsistenz der Verfahren (dargestellt in Tabelle 6), wobei diese jedoch nicht unter begrenzter Bearbeitungszeit erfasst wurden, sodass ihre Bedeutung unklar bleibt (Christ et al., 2008). Darüber hinaus kritisieren die Autoren die Vorgehensweise bei der Konstruktion von CBM-Verfahren im Bereich des Rechnens, welche maßgeblichen Einfluss auf die wissenschaftliche Güte der Tests hat. Ursprünglich wurde dabei für den Bereich Mathematik derart vorgegangen, dass zunächst eine zufällig gezogene, für das Schuljahr bzw. den Testzeitraum repräsentative Auswahl von Rechenaufgaben gezogen wurde, welche dann jeweils in zufälliger Ordnung auf einem Testblatt positioniert wurden (Shapiro, 1996; Shinn, 2002; Shinn, Bamonto, Cornachione, Parker, Peterson, Thurber et al., 1997). Für mastery measures (Verfahren, welche einzelne Aspekte isoliert abfragen, bspw. Addition im Zahlenraum bis zehn) erscheint eine solche Vorgehensweise angebracht, da von einer geringeren Variabilität der Itemschwierigkeiten ausgegangen werden kann (Christ et al., 2008; Christ & Vining, 2006). Jedoch könnte eine derartige Konstruktionsweise bei general outcome measures (Verfahren, welche gleichzeitig verschiedene Aspekte abprüfen, bspw. verschiedene mathematische Konzepte) dazu führen, dass die verschiedenen Testinhalte nicht gleichmäßig durch Items repräsentiert sowie darüber hinaus auch nicht gleichmäßig auf das Verfahren verteilt sind. Die erzielte Schülerleistung ist dann eher durch die Begrenzung der Zeit und nicht durch die tatsächliche Fähigkeitsausprägung zu erklären. Um diesen Missstand zu überwinden, schlagen Christ und Vining (2006) die Methode der

stratifizierten Itemanordnung vor, welche in Abschnitt 3.3 detailliert dargestellt wird.

Tabelle 6: Übersicht über Studien zur internen Konsistenz von CBM-Verfahren im Bereich Mathematik (entnommen aus: Christ et al., 2008, S. 202)

	Probanden			mittlere interne
Studie	Stichprobe	Klassenstufe	N	Konsistenz
L.S. Fuchs, D. Fuchs & Hamlett (1989a)	Sonderschule	2 – 9	30	.98
Phillips, Hamlett, L.S. Fuchs & D. Fuchs (1993)	Regelschule	3	20	.85

Ergänzungen zu diesen Angaben liefern Foegen et al. (2007). Sie analysierten auch weitere Studien, welche bei Christ et al. (2008) durch die von den Autoren angelegten Ausschlusskriterien nicht berücksichtigt wurden. Die meisten dieser Studien sind vor bzw. im Jahr 1989 publiziert worden, wodurch sie in der Analyse von Christ et al. (2008) keine Beachtung fanden. Foegen et al. (2007) berichten von hohen Reliabilitätskoeffizienten über die Studien hinweg, von welchen die meisten über r = .80 liegen. Einzig erste Untersuchungen zu Beginn der 80er Jahre von Tindal, Germann und Deno (1983) sowie Tindal, Marston und Deno (1983) konnten nur geringere Reliabilitäten der Rechen-CBM nachweisen, speziell für den Bereich der Multiplikation (r = .61, Tindal, Germann et al., 1983) und Division (r = .48, Tindal, Germann et al., 1983).

Den Forschungsstand hinsichtlich der Testgüte von Rechen-CBM erweiternd führen Foegen et al. (2007) Reliabilitätsdaten von CBM zur Erfassung erster und weiterführender mathematischer Einsichten und Konzepte (early numeracy, concepts & application) an. Auch hier zeigen sich durchweg hohe Reliabilitätskoeffizienten von r > .80. Ausnahme dabei stellen die Verfahren *free count* (r = .71, VanDerHeyden et al., 2004), *choose shape* (r = .40, VanDerHeyden et al., 2004) sowie *draw circles* (r = 70, VanDerHeyden et al., 2001) dar.

3.2.8.2 Forschungsergebnisse hinsichtlich der Validität verfügbarer CBM-Verfahren im Bereich Mathematik

Neben Angaben zur Reliabilität analysierten Christ et al. (2008) die in den Untersuchungen verwendeten CBM-Verfahren zur Erfassung von Rechenkompetenzen auch hinsichtlich ihrer Validität. Bezüglich der Inhaltsvalidität konnten die Autoren lediglich eine Studie (Allinder et al., 1992) ausmachen, welche anhand von Befragungen der involvierten Lehrkräfte sowie Lehrplanbeauftragten konkrete Angaben zur inhaltlichen Validität der eingesetzten CBM liefern konnte. Für Studien, die das Monitoring Basic Skills Progress (MBSP) von L.S. Fuchs et al. (1998, 1999) einsetzten, kann, so Christ et al. (2008), die Inhaltsvalidität indirekt über das beim MBSP verwendete Itemsampling angenommen werden, welches sich am jeweils gültigen Lehrplan der entsprechenden Klassenstufe für den Bereich des Rechnens ausrichtet. Zur Einschätzung der Konstrukt- und Kriteriumsvalidität liegen deutlich mehr Informationen in den untersuchten Studien vor. Je nach verwendetem Kriterium streuen die ermittelten Korrelationskoeffizienten im geringen bis hohen Bereich (Christ et al., 2008). Dabei wird deutlich, dass die höchsten Korrelationen dann erzielt werden, wenn als Kriterium ein Test eingesetzt wird, der ausschließlich die Rechenkompetenzen der Schüler erfasst. Korrelationen mit Verfahren, welche weitere mathematische Aspekte abfragen, fallen deutlich geringer aus. Ähnliche Ergebnisse hinsichtlich der Kriteriumsvalidität für Rechen-CBM zeigen sich bei Foegen et al. (2007). Diese Ergebnisse sind nicht sonderlich verwunderlich. So wird durch Rechen-CBM lediglich ein Aspekt mathematischer Inhalte abgefragt: das Rechnen. Ein mathematisches Verständnis zeigt sich jedoch nicht im bloßen Anwenden von Lösungsalgorithmen auf gestellte Rechenprobleme, sondern zeichnet sich, wie in Kapitel 2 erläutert, durch das Vorhandensein von mathematischen Einsichten und Konzeptvorstellungen aus. Darüber hinaus umfassen mathematische

Kompetenzen, neben dem Rechnen von Aufgaben, auch weitere Gesichtspunkte, wie bspw. das Lösen von Sachaufgaben, den Umgang mit Größen wie Geld und Uhrzeiten sowie auch stochastische und geometrische Inhalte (siehe auch Abschnitt 3.2.4). Somit geben Rechen-CBM lediglich einen engen Fokus auf mathematische Inhalte. Auch Christ et al. (2008) kommen zu dem Schluss: „Our review of the instrumentation, procedures, and psychometric evidence supports the conclusion that CBM-M has limited construct representation, which limits its construct validity as a generalized mathematics assessment" (S. 200).

Ähnliches berichten Foegen et al (2007) für CBM-Verfahren, welche erste mathematische Einsichten und Kompetenzen abfragen (early numeracy). Auch bei den untersuchten Studien, welche sich auf diese CBM konzentrieren wurde eine weite Spannbreite von Koeffizienten zur Beschreibung der Kriteriumsvalidität ermittelt, die auf mittlere bis hohe Zusammenhänge mit den Kriteriumsmaßen hinweisen. Die höchsten Werte ergaben sich bei Korrelation mit Verfahren, welche sich ausschließlich auf das Zahlenwissen der untersuchten Kinder bezogen. Geringere Koeffizienten wurden errechnet bei Korrelation mit Tests, die umfassendere mathematische Kompetenzen erfassen. Auch diese Ergebnisse sind zu erwarten, da sich doch die CBM des Typs early numeracy auf abgegrenzte Kompetenzbereiche (number identification, quantity discrimination, missing number etc., siehe dazu 3.2.4) beziehen, welche eher Aufgaben zum Zahlenwissen fokussieren und somit lediglich einen Teilaspekt der Mathematik abbilden.

Bei CBM des Typs concepts & application (Monitoring Basic Skills Progress measures, MBSP, L.S. Fuchs et al., 1998, 1999) sieht die Forschungslage etwas anders aus. Neben den oben genannten hohen Werten hinsichtlich der Reliabilität dieser Verfahren, zeigen sich ähnlich hohe Werte bezüglich der Validität (Klasse zwei: r = .81;

Klasse drei: r = 74; Klasse vier: r = 79). Insgesamt, so stellen Foegen et al. (2007) fest, seien die von L.S. Fuchs et al. (1998, 1999) entwickelten Monitoring Basic Skills Progress Measures die empirisch am besten bewährtesten. „The most well established measures at the elementary level are the MBSP measures (L.S. Fuchs et al., 1998, 1999). These measures have empirical data supporting their reliability, criterion validity, and sensitivity in detecting student growth" (Foegen et al., 2007, S. 132).

Die Untersuchungen zur Güte von CBM des Typs computation zusammenfassend, kommen Christ et al. (2008) zu dem Schluss , dass der Einsatz dieser Verfahren in der Schule, trotz der zuvor genannten Einschränkungen, sinnvoll ist, unter der Voraussetzung, dass sie nicht als Messverfahren zur Bestimmung allgemeiner mathematischer Fähigkeiten aufgefasst, sondern als Tests verstanden werden, welche nur einen limitierten Fähigkeitsausschnitt abfragen und somit eher Screeningcharakter besitzen. Dennoch besteht weiterer Bedarf bei der Erforschung von Rechen-CBM sowie weiteren Verfahren, welche einen weiteren Blick auf mathematische Inhalte abbilden können (Christ et al., 2008). Hinsichtlich CBM-Verfahren der Typen early numeracy und concepts & application gelangen Foegen et al. (2007) zu einem ähnlichen Ergebnis.

Einen weiteren, regelmäßig aktualisierten Überblick über verfügbare CBM-Verfahren für den schulischen Einsatz gibt der Internetauftritt des National Center on Response to Intervention (2012). Dieser soll insbesondere Lehrkräften Hilfestellung bei der Auswahl entsprechender CBM geben. Die dort gelisteten Verfahren sind durch ein Komitee verschiedener Forscher (u. a. Hosp, Christ, Hintze, Shapiro, Stecker) hinsichtlich ihrer Güte inspiziert worden. In dem regelmäßig durchgeführten Review-Prozess werden angemeldete CBM-Verfahren auf folgende Kriterien hin untersucht:

- Höhe der Reliabilität des Verfahrens
- Höhe der Reliabilität des mit dem Verfahren ermittelten Anstiegs
- Höhe der Validität des Verfahrens
- Höhe der prognostischen Validität des mit dem Verfahren ermittelten Anstiegs
- verfügbare Anzahl Parallelversionen
- Höhe der Sensitivität des Verfahrens
- Belege für Effektivität
- Verfügbarkeit von Normdaten getrennt nach Gruppen (soziale Schicht, Herkunft etc.)
- Verfügbarkeit von Analyseergebnissen der Gütekriterien getrennt nach Gruppen (soziale Schicht, Herkunft etc.) (National Center on Response to Intervention, 2012)

Eine Auswahl der auf der Internetseite verfügbaren Verfahren ist in Abbildung 21 dargestellt. Dabei wurden alle CBM ausgewählt, welche den mathematischen Bereich der Grundschule adressieren. Ein ausgefüllter Kreis deutet an, dass für das jeweilige CBM ausreichende Belege zur Erfüllung des entsprechenden Kriteriums geliefert wurden. Ein halb gefüllter Kreis zeigt mäßige Belege, ein leerer Kreis nicht überzeugende Belege an. Ein Strich gibt zu erkennen, dass für das jeweilige Kriterium keine Daten ermittelt werden konnten.

Tools	Area	Reliability of the Performance Level Score	Reliability of the Slope	Validity of the Performance Level Score	Predictive Validity of the Slope of Improvement	Alternate Forms	Sensitive to Student Improvement	End-of-Year Benchmarks	Rates of Improvement Specified	Norms Disaggregated for Diverse Populations	Disaggregated Reliability and Validity Data	COMPARE / RESET
AIMSweb	[i] Math	●	●	●	●	◐	○	●	●	No	●	☐
AIMSweb	[i] Test of Early Numeracy - Missing Number	●	●	●	●	●	○	●	●	No	●	☐
AIMSweb	[i] Test of Early Numeracy - Number ID	●	●	●	●	●	○	●	●	No	●	☐
AIMSweb	[i] Test of Early Numeracy - Oral Counting	●	●	●	●	●	○	●	●	No	●	☐
AIMSweb	[i] Test of Early Numeracy - Quantity Discrimination	●	●	●	●	●	○	●	●	No	●	☐
easyCBM	[i] Math	●	●	●	●	●	—	●	○	No	●	☐
mClass Math	Computation	●	○	●	○	—	—	●	—	No	—	☐
mClass Math	Concepts	●	○	●	○	●	—	●	—	No	—	☐
mClass Math	Oral Counting	●	○	●	—	—	—	●	—	No	—	☐
mClass Math	Missing Number	●	○	●	○	◐	—	●	—	No	—	☐
mClass Math	Next Number	●	○	●	—	◐	—	●	—	No	—	☐
mClass Math	Number Facts	●	○	●	○	●	—	●	—	No	—	☐
mClass Math	Number Identification	●	○	●	—	●	—	●	—	No	—	☐
mClass Math	Quantity Discrimination	●	○	●	○	●	—	●	—	No	—	☐
Monitoring Basic Skills Progress (MBSP)	Basic Math Computation	●	●	●	●	●	●	●	●	No	—	☐
Monitoring Basic Skills Progress (MBSP)	Basic Math Concepts / Applications	●	●	●	●	●	●	●	●	No	—	☐
Orchard Software	* Math	○	—	○	—	○	○	—	—	No	—	☐
STAR	[i] Math	●	●	●	●	●	●	●	●	No	●	☐
Vanderbilt RTI Monitor (Hasselbring et al.)	* Math	●	—	●	—	◐	—	○	—	No	—	☐
Yearly ProgressPro	Math	●	●	●	●	●	●	●	●	No	—	☐

Legend: ● Convincing evidence ◐ Partially convincing evidence ○ Unconvincing evidence — Data unavailable or inadequate

* Added in the 2010 review [i] Updated in the 2010 review

Abbildung 21: Ausschnitt der Internetseite des NCRTI zum Überblick über CBM-Verfahren (Stand 12.01.2013) für den Bereich der Grundschulmathematik, welche durch das Review-Komitee des National Center on Response to Intervention begutachtet wurden (National Center on Response to Intervention, 2012)

Die meisten der in Abbildung 21 aufgeführten Verfahren werden in den USA kommerziell vertrieben. Einzig die Internetplattform easyCBM.com (2012a) bietet neben einem kommerziellen auch einen eingeschränkten, dafür jedoch kostenfreien Zugang zu CBM-Verfahren für einzelne Lehrkräfte an. Ein Zugang für einen gesamten Distrikt muss dort käuflich erworben werden. Die Internetplattform

wurde von Mitarbeitern der Universität Oregon entwickelt und ist seit dem Jahr 2006 in Betrieb. Es werden CBM- als auch Screeningverfahren für den Bereich Lesen sowie Mathematik vom Kindergarten bis zur Klassenstufe 8 angeboten. Neben der Onlinedurchführung der Tests können alle Verfahren als Paper-Pencil-Version heruntergeladen werden. Die auf der Internetseite angebotenen Verfahren sind hinsichtlich ihrer Testgüte wissenschaftlich analysiert worden. Die Ergebnisse der Untersuchungen sind als Berichte auf der Homepage verfügbar (easyCBM.com, 2012b). Die Konzeption von Testverfahren für den Bereich Mathematik vollzog sich in mehreren Schritten (Alonzo & Tindal, 2009). Zunächst wurde durch acht ausgewählte Personen ein Itempool erstellt. Diese waren an konkreten Richtlinien orientiert. Besonders wurde darauf geachtet, dass die Items gültigen Bildungsstandards entsprechen und dass sie sich vor allem für besonders schwache Kinder eignen. Die konzipierten Items wurden anschließend einem Review-Prozess von sechs Forschern unterzogen. Es resultierte ein Itempool aus 6600 Aufgaben getrennt nach wesentlichen mathematischen Themen für den Kindergarten bis zur Klassenstufe 8. Anhand einer Stichprobe von etwa 2800 Kindern wurden die erarbeiteten Items pilotiert. Dabei wurden die Items auf Rasch-Konformität geprüft. Weitere Analysekriterien waren die errechneten Schwierigkeiten sowie die ermittelten Standardfehler der Items. Auf Grundlage dieser Analysen wurden anschließend jeweils für wesentliche mathematische Bereiche jeder Klassenstufe (z. B. Zahlen und Operationen, Geometrie, Größen und Messen etc.) jeweils 16 Items zu zehn parallelen CBM-Verfahren zusammengefasst. Zusätzlich wurden jeweils weitere 3 Verfahren mit je 16 Items erstellt. Diese werden als Screening für jede Klassenstufe und jeden mathematischen Aspekt mehrfach im Schuljahr eingesetzt. Die Parallelität der so entstandenen Verfahren wurde anhand der mittleren Schwierigkeiten bereichs- sowie klassenstufenweise geprüft (Alonzo & Tindal, 2009). In weiteren Untersuchungen konnten hohe

Reliabilitätskoeffizienten (r > .80) der Verfahren für die Klassenstufen 1 und 2 sowie für den Kindergartenbereich nachgewiesen werden. Auch Analysen hinsichtlich der Validität (Kriteriums-, Konstrukt- sowie prognostische Validität) wurden durchgeführt. Die CBM der Internetplattform www.easyCBM.com stellen somit eine weitere vielversprechende sowie wissenschaftlich fundierte Möglichkeit der Lernverlaufsdiagnostik in der Schule dar (Anderson, Lai, Nese, Park, Sáez, Jamgochian, Alonzo & Tindal, 2010).

3.2.8.3 Forschungsergebnisse hinsichtlich der Effektivität verfügbarer CBM-Verfahren im Bereich Mathematik

CBM als wiederholt einsetzbare Verfahren zur Evaluation von Unterrichts- und Fördermaßnahmen gelten als Spezialfall des formative assessment (siehe dazu 3.1.1). Die in Abschnitt 3.1.3 aufgeführten Studien hinsichtlich der Effektivität des Einsatzes formativer Leistungsmessung in der Schule geben somit Hinweise darauf, dass auch regelmäßig wiederholte Lernverlaufserhebungen mittels CBM zur Steigerung von Schülerleistungen beitragen. Neben den genannten Effekten des formative assessment gibt es darüber hinaus auch eine Reihe an Untersuchungen, welche sich konkret auf die Effektivität des Einsatzes curriculumbasierter Messungen beziehen. Die Forschungsergebnisse zeigen auf, dass regelmäßige CBM-Erhebungen zur Leistungssteigerung der Schüler führen können, sofern die erfassten Ergebnisse zur Anpassung des Unterrichts an individuelle Bedürfnisse der Schüler genutzt werden (Colón & Kranzler, 2006; L.S. Fuchs et al., 1984; L.S. Fuchs, D. Fuchs & Hamlett, 1989b; L.S. Fuchs, D. Fuchs, Hamlett & Allinder, 1991; L.S. Fuchs, D. Fuchs, Hamlett & Stecker, 1991; Jones & Krause, 1988; Shapiro et al., 2006; Stecker & L.S. Fuchs, 2000; Wayman et al., 2007; Wesson, 1991; Wiley & Deno, 2005; Yeo, 2010). Die

meisten Studien beziehen sich jedoch auf CBM im Bereich des Lesens. Welche Effekte der Einsatz von CBM für den Bereich der Mathematik erzielt, soll nachfolgend an ausgewählten Beispielen dargestellt werden.

Hinweise auf die Wirksamkeit von CBM im Bereich der Mathematik in Verbindung mit darauf abgestimmten Interventionsmaßnahmen im Sinne des peer tutoring („peer-assisted learning strategies“, PALS) liefern L.S. Fuchs, D. Fuchs, Phillips, Hamlett und Karns (1995). In die Untersuchung der Autoren flossen die Daten von 20 Klassen, in denen klassenweit das PALS-Programm einsetzt und die Effektivität dieser Maßnahme mittels regelmäßigem Einsatz von CBM erhoben wurde. Weitere 20 Klassen nahmen als unbehandelte Kontrollgruppe teil. Dabei wurde mittels wöchentlicher CBM-Testungen die Leistungsentwicklung dreier verschiedener Gruppen von Kindern beobachtet: Schüler mit durchschnittlicher Leistung, Schüler mit unterdurchschnittlicher Leistung und Schüler mit unterdurchschnittlicher Leistung, welche eine diagnostizierte Lernschwäche hatten. Die Autoren berichten von Effektstärken der Intervention (PALS zusammen mit CBM) von d = 0.30 bis 0.95 für die Kinder mit unterdurchschnittlichen Leistungen sowie Effektstärken von d = 0.32 bis 0.34 für Kinder mit durchschnittlichen Leistungen (L.S. Fuchs et al., 1995). Zwar kann aufgrund der Kopplung von peer-tutoring mit curriculumbasierten Messen nicht präzise Auskunft gegeben werden, welche Methode die beobachteten Effekte tatsächlich bedingt hat, dennoch reihen sich diese Ergebnisse in den Kanon der Resultate weiterer Studien zur Effektivität von CBM ein.

In der Literaturanalyse von Foegen et al. (2007) konnten die Autoren sieben Studien identifizieren, welche der Frage der Effektivität des Einsatzes von Mathematik-CBM in der Schule nachgingen. Dabei wurden nur solche Untersuchungen ausgewählt, in denen CBM nicht als ein Aspekt einer breiten Interventionsmaßnahme genutzt wurden,

sondern die Studien, in welchen ausschließlich der Gebrauch von curriculumbasierten Messverfahren untersucht wurde. Alle diese Studien wurden von der Forschungsgruppe um Lynn S. Fuchs an der Vanderbilt Universität in Nashville, Tennessee durchgeführt. Sechs der sieben Studien bezogen sich auf das von L.S. Fuchs et al. (1998, 1999) entwickelte Rechen-CBM „Monitoring Basic Skills Progress“ (MBSP) (Allinder & BeckBest, 1995; Allinder, Bolling, Oats & Gagnon, 2000; L.S. Fuchs, D. Fuchs & Hamlett, 1989b; L.S. Fuchs et al., 1990; L.S. Fuchs, D. Fuchs, Hamlett & Stecker, 1991; Stecker & L.S. Fuchs, 2000). Alle aufgeführten Untersuchungen deuten darauf hin, dass der regelmäßige Einsatz von CBM (hier konkret MBSP computation) zu einer Leistungssteigerung der Schüler führen kann. Durchweg weisen die Schülergruppen, welche regelmäßig mittels curriculumbasierter Tests geprüft wurden, höhere Leistungen im Vergleich zu einer Kontrollgruppe auf (Foegen et al., 2007). Die hohe Variabilität der ermittelten Effektstärken, von gering zu hoch, zeigt jedoch an, dass nicht alle gefundenen Ergebnisse in der Praxis relevant sind. Besonders hoch scheinen die Effekte, wenn beteiligte Lehrkräfte zu den regelmäßigen CBM-Ergebnissen ihrer Schüler weitere Unterstützung, im Sinne zusätzlicher Beratungssysteme erhalten. Dies belegen u. a. Studien von L.S. Fuchs et al. (1991) sowie L.S. Fuchs et al. (1994). In beiden Studien wurde die Leistungsentwicklung dreier Schülergruppen mit unterschiedlichen Versuchsbedingungen analysiert. Dabei handelte es sich in beiden Untersuchungen stets um eine Gruppe, die an regelmäßigen CBM-Testungen teilnahm (Gruppe CBM). Anschließend erhielt die unterrichtende Lehrkraft jeweils Rückmeldung über die erzielten Schülerleistungen. In der zweiten Gruppe erhielt die unterrichtende Lehrkraft zusätzlich zu den ermittelten Schülerergebnissen darauf abgestimmte Handlungsempfehlungen für die Unterrichtsgestaltung (Gruppe CBM+E). Die jeweils dritte Schülerkohorte erhielt als Kontrollgruppe keinerlei Intervention. Bei den CBM-Verfahren handelte es sich um das Programm MBSP (L.S. Fuchs et al., 1998,

1999). Als Kriterium diente ein standardisierter Test, welcher mathematische Problemstellungen der Klassenstufen 1 bis 6 abfragt.

In der Studie von L.S. Fuchs et al. (1991) wurde die Leistungsentwicklung von 33 Klassen (Klassenstufe 1 bis 6) in 15 Schulen analysiert. Die Zuteilung der Klassen zu einer der drei oben genannten Versuchsoptionen wurde zufällig vorgenommen. Es fanden zweimal wöchentlich CBM-Testungen zum Rechnen und Sachrechnen statt. Es zeigte sich, dass die Schüler der Klassen, deren unterrichtende Lehrkraft neben den CBM-Ergebnissen zusätzlich weitere Hinweise zur Unterrichtsanpassung erhielt, durchschnittlich höhere Leistungszuwächse als die Kontrollgruppenkinder erzielten (mittlere Effektstärke: d = 0.84 bei Bewertung richtiger Aufgaben bzw. d = 0.94 bei Bewertung richtiger Ziffern). Auch im Vergleich zur Schülergruppe CBM konnten die Kinder der Gruppe CBM+E höhere Leistungszuwächse verzeichnen (mittlere Effektstärke: d = 0.64 bei Bewertung richtiger Aufgaben bzw. d = 0.81 bei Bewertung richtiger Ziffern) (L.S. Fuchs et al., 1991).

L.S. Fuchs et al. (1994) hingegen untersuchten die mittels wöchentlich durchgeführter CBM-Tests ermittelten Lernverläufe der Schüler von insgesamt 40 Grundschulklassen an 11 Schulen, welche zufällig auf die beschrieben Gruppen verteilt wurden. Dabei konnte eine höhere Leistungsentwicklung der Schülergruppe CBM+E gegenüber der Kontrollgruppe (mittlere Effektstärke: d = 0.43) sowie gegenüber der Schülergruppe CBM (eine konkrete Effektstärke konnte nicht ermittelt werden, trotzdem zeigte sich eine bessere Tendenz der Gruppe CBM+E) verzeichnet werden (L.S. Fuchs et al., 1994). Damit zeigt auch diese Studie eine Überlegenheit der Kinder, dessen Lehrkräfte CBM-Daten der Schüler zusammen mit individuell abgestimmten Handlungsempfehlungen für die Unterrichtsgestaltung erhalten. Zusammenfassend kommen L.S. Fuchs et al. (1994) zu folgendem Schluss:

> „*The main effects documented in this study (along with the lack of statistically significant treatment by student type interactions) indicate that providing general educators with CBM information, along with specific instructional recommendations for acting on those assessment profiles, results in superior achievement regardless of whether students are classified as learning disabled, low achieving, or average achieving"* (S. 535).

Die positiven Auswirkungen über alle Leistungsgruppen einer heterogenen Schülerschaft hinweg, geben Anlass zu der Vermutung, dass die regelmäßige Leistungserfassung mittels CBM (zumindest in der Grundschule) eine vielversprechende Maßnahme in einer gemeinsamen Beschulung von Kindern verschiedener Leistungsniveaus darstellt. Dies setzt jedoch Unterstützungssysteme zur Beratung von Lehrkräften voraus, welche Hinweise zur Gestaltung von individuell auf die Schüler angepassten Unterrichts- und Fördermaßnahmen liefern (L.S. Fuchs et al., 1994).

Trotz der genannten Untersuchungen hinsichtlich der Güte sowie Effektivität von CBM-Verfahren im Bereich der Mathematik, weisen Foegen et al. (2007) auf Grenzen der vorhandenen Forschungslage hin. So zeigt die vorgenommene Literaturanalyse, dass der größte Teil der Forschungsbemühungen sich auf das Feld der Grundschule bezieht. Hingegen gibt es nur wenige bis keine Forschungsaktivitäten im Bereich erster sowie weiterführender Mathematik in der Vor- bzw. der Sekundarschule. Darüber hinaus weist von den untersuchten Verfahren lediglich das MBSP (L.S. Fuchs et al., 1998, 1999) umfassende Forschungsbelege auf. Nur für Verfahren des MBSP computation konnte die wissenschaftliche Güte der Verfahren nachgewiesen und außerdem Hinweise für die Einsetzbarkeit der Verfahren als Werkzeug zur Lernverlaufsdokumentation sowie für die Effektivität ihres Einsatzes zur Steigerung der Schülerleistungen erbracht werden. Laut der von L.S. Fuchs (2004) postulierten Forschungskriterien für CBM lässt erst eine umfassende

Untersuchung eines CBM hinsichtlich dieser drei Aspekte eine abschließende Bewertung des Verfahrens zu (genauer in Kapitel 3.3.2, 3.3.3 und 3.3.4). So konstatieren Foegen et al. (2007) resümierend, dass die Erarbeitung und Evaluation von CBM-Verfahren im Bereich der Mathematik auf Grundlage dieser Kriterien ein zentrales Forschungsdesiderat darstellt.

3.3 Schlussfolgerungen für die Konzeption und Evaluation von CBM für den mathematischen Erstunterricht

„Je eher, desto besser." – dieses Motto gilt vor allem auch für den erfolgreichen Erwerb erster mathematischer Kompetenzen im Unterricht der ersten Klasse. Dabei steht die Lehrkraft vor der durchaus nicht immer einfachen Aufgabe, Stärken als auch Problembereiche eines jeden Kindes der Klasse zu identifizieren und zu dokumentieren, da das zeitnahe und adäquate Reagieren auf Schwierigkeiten beim Erlernen des Rechnens dem Entstehen großer sich manifestierender Wissenslücken vorbeugen kann (Aunola et al., 2004; Helmke & Weinert, 1997). Die Identifikation von Risikokindern und die Erfassung des Lernverlaufs aller Kinder zu ermöglichen, ist Aufgabe einer umfassenden Diagnostik. Zudem muss sie förderrelevante Hinweise zur inhaltlichen Ausgestaltung von Unterricht und Förderung ermöglichen. Dabei muss angemerkt werden, dass es nicht das Diagnoseverfahren gibt, welches alle diese Anforderungen abdeckt (Heller & Hany, 2002). Es muss vielmehr auf ein vielschichtiges diagnostisches Vorgehen zurückgegriffen werden. Wie schon in Kapitel 3 und Abschnitt 3.1.4.2 beschrieben, schlägt bspw. das National Center on Response to Intervention (2010) eine Gliederung von Diagnostik in drei wesentliche Aspekte vor: den Einsatz von Screeningverfahren, eine fortlaufend begleitende Überwachung des Lernprozesses (progress monitoring) sowie den

Einsatz von Methoden und Verfahren zur Beschreibung von Stärken und Schwächen bei zuvor erkannten Risikokindern zwecks Ableitung von Verursachungshypothesen und anschließender Fördermaßnahmen.

Auch Hartke (2005) beschreibt regelmäßige Evaluationsschleifen des Unterrichts im Sinne eines progress monitorings als einen wesentlichen Aspekt der schulischen Prävention. Ausgehend von dem daraus gezogenem Feedback sind Anpassungen der Unterrichtsmaßnahmen an den jeweiligen Wissens- und Könnensstand der Schüler möglich, so wie es auch das Konzept des formative assessment vorsieht. Die Effekte des formative assessment als eine Maßnahme der formativen Evaluation zur Steigerung der Schülerleistungen sind in Abschnitt 3.1.3 aufgeführt worden. Der erfolgreiche Einsatz in der Schule ist wissenschaftlich belegbar. Auch in der deutschen Fachliteratur ist der regelmäßige Einsatz von Verfahren zur Leistungsstandbestimmung kein Novum, ein solches Vorgehen wurde bspw. schon von Klauer (1978) beschrieben. Nun wurde zwar bereits aus verschiedenen Richtungen der Forschung auf die Vorteile formativer Evaluationsmaßnahmen im Unterricht hingewiesen, aber eine flächendeckende Umsetzung von regelmäßigen Leistungstestungen im Sinne curriculumbasierter Messungen ist in Deutschland noch nicht bekannt. Dabei stellen curriculum-based measurements als eine Spezialform des formative assessment mit besonderer Akzentuierung der wissenschaftlichen Güte der Verfahren eine besonders vielversprechende Methode dar. Mittels ihres regelmäßigen Einsatzes sind objektive, reliable und valide Einschätzungen der Schülerleistungen möglich. Dadurch werden substanzielle Informationen für die Lehrkräfte gewonnen. Bezogen auf die mathematische Domäne der Grundschule scheinen curriculumbasierte Messungen ein nützliches Hilfsmittel für Lehrkräfte zu sein, um sie bei der Identifikation rechenschwacher Schüler zu unterstützen und Hinweise für die Unterrichts- bzw.

Förderarbeit zu geben. In Hinblick auf Entwicklung und Anwendung von CBM-Verfahren kann von einer mittlerweile etwa 30 Jahre andauernden Forschungstradition in den USA profitiert werden. In dieser Zeit haben sich dort Forschungskonventionen entwickelt, welche konkrete Hinweise für die Konzeption, Konstruktion und Evaluation von CBM-Verfahren geben.

3.3.1 Notwendige Analysen bei der CBM-Entwicklung nach Fuchs (2004)

Da sich CBM vor allem durch ihre psychometrische Betonung von anderen Maßnahmen des formative assessments abheben, spielt die Orientierung an zentralen wissenschaftlichen Gütekriterien eine wesentliche Rolle bei der Konzeption und anschließenden Evaluation der Verfahren. L.S. Fuchs (2004) beschreibt für die systematische Evaluation von CBM-Verfahren drei Stufen der wissenschaftlichen Überprüfung. Mit jeder Stufe sind bestimmte Kriterien verbunden, deren Erfüllung als Nachweis für die Güte, Anwendbarkeit bzw. Effektivität eines Tests zur Dokumentation des Lernverlaufs gilt:

> *„Investigation of the technical features of the static score (i.e., performance level at one point in time) constitutes Stage 1. In Stage 2, technical features of slope are assessed to determine whether increasing CBM scores (or slope) are in fact associated with improvement in overall competence in the academic domain (e.g., Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). Stage 3 concerns instructional utility. This is where studies are conducted to determine whether practitioners can use the CBM information to improve instructional decisions and student achievement (e.g., Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, Hamlett, & Stecker, 1991)“* (L.S. Fuchs, 2004, S. 189).

In einem ersten Schritt sind demnach Maßnahmen der Überprüfung allgemein anerkannter Gütekriterien erforderlich. Belege für die Reliabilität und Validität der Verfahren sollen hier sicherstellen, dass

das zu messen beabsichtigte Konstrukt auch tatsächlich durch die Tests erfragt wird (Validität) und das möglichst ohne Verzerrung durch Messfehler (Reliabilität). Von besonderem Interesse sind dabei die Übereinstimmungsvalidität, welche die Höhe eines Zusammenhangs mit einem Außenkriterium beschreibt, sowie die Güte des Verfahrens zur Vorhersage zukünftiger Leistungsentwicklung (prognostische Validität) (Foegen et al., 2007). Diesbezügliche Informationen sind ein wesentlicher Schritt bei der Evaluation von CBM-Verfahren, geben zunächst jedoch nur Auskunft darüber, inwieweit sich der Test eignet, den Leistungsstand der Kinder zu einem festen Zeitpunkt, im Sinne eines statischen Indikators (Foegen et al., 2007), zu erfassen. Damit ist der Nachweis erbracht, dass das Verfahren zum Einsatz als Screening im Unterricht geeignet ist, aber Belege für den zuverlässigen Gebrauch als Hilfsmittel zur Lernverlaufsdokumentation sind dabei nicht gegeben (L.S. Fuchs, 2004). Hierfür sind weitere Analysen entsprechend der von L.S. Fuchs (2004) beschriebenen zweiten Stufe notwendig. Dabei geht es um die Frage, wie reliabel und sensibel CBM-Verfahren den durch regelmäßig wiederholte Messungen erhobenen Lernverlauf der Schüler abbilden können. Auf der Stufe 3 wird schließlich die Einsetzbarkeit der Verfahren im schulischen Alltag geprüft. Letzteres betreffend existieren Studien, die untersuchen, ob Lehrer die entwickelten CBM-Verfahren zur Dokumentation von Lernverläufen ihrer Schüler nutzen. Zentraler Fokus solcher Durchführbarkeitsuntersuchungen liegt auf der Zeiteffizienz sowie der einfachen Schulung zur Durchführung der wiederholten Testungen (Deno, 2003). Neben dem Aspekt der Realisierbarkeit spielt jedoch auch die Effektivität des Einsatzes von CBM-Verfahren auf Stufe 3 eine zentrale Rolle, d. h. der durch den Einsatz von CBM im Unterricht hervorgerufene Leistungsanstieg der Schüler sollte anhand einer empirisch gewonnenen Effektstärke quantifiziert werden. Die von Fuchs (2004) beschriebenen drei Stufen der wissenschaftlichen Prüfung curriculumbasierter Messungen wurden in der Fachliteratur

als sinnvolle Kriterien aufgenommen (Foegen et al., 2007) und sollen der vorliegenden Arbeit als Leitfaden der Untersuchung dienen.

3.3.2 Notwendige Analysen hinsichtlich der von L.S. Fuchs (2004) vorgeschlagenen Stufe 1

Fragen hinsichtlich der von L.S. Fuchs (2004) vorgeschlagenen Forschungsstufe 1 zielen vor allem auf Analysen bezüglich der Reliabilität und Validität ab. Die Reliabilität eines Testverfahrens bezieht sich auf dessen Messgenauigkeit. „Ein Test ist dann reliabel (zuverlässig), wenn er das Merkmal, das er misst, exakt, d. h. ohne Messfehler, misst“ (Moosbrugger & Kelava, 2012, S. 11). Bei wiederholten Messungen mit gleichen bzw. ähnlichen Testverfahren spielen vor allem die Retest-Reliabilität bzw. die Paralleltest-Reliabilität eine zentrale Rolle (Deno, 2003). Die Retest-Reliabilität gilt als das Maß der Übereinstimmung bei wiederholt eingesetzten Tests (Moosbrugger & Kevala, 2012; Schermelleh-Engel & Werner, 2012). Da CBM zeitlich vergleichsweise hochfrequentiert eingesetzt werden sollen, kann jedoch nicht jeweils ein und dasselbe Verfahren verwendet werden, da ein Lerneffekt die Ergebnisse verzerren könnte. Entsprechend müssen für ein CBM verschiedene Aufgabenblätter von gleicher Schwierigkeit (Paralleltests) vorliegen. Anhand der Paralleltest-Reliabilität soll dann eingeschätzt werden, inwieweit die Testvorlagen Aufgaben mit ähnlicher Schwierigkeit umfassen, d. h. inwieweit sie parallel sind (Moosbrugger & Kevala, 2012; Schermelleh-Engel & Werner, 2012). Dies stellt bei wiederholten Leistungserhebungen zur Dokumentation des Lernverlaufs einen besonders wichtigen Aspekt dar. Ist keine Parallelität der eingesetzten Testverfahren gegeben, kann der sich durch die Messungen abzeichnende Lernverlauf nicht eindeutig auf die Fähigkeiten der Kinder zurückgeführt werden, da die Messungen durch unterschiedliche Schwierigkeitsgrade der Verfahren beeinflusst werden können.

Während die Reliabilität die Genauigkeit einer Messung fokussiert, bezieht sich die Validität auf die inhaltliche Güte eines Testverfahrens. Dabei wird sie folgendermaßen definiert: „Ein Test gilt dann als valide (»gültig«), wenn er das Merkmal, das er messen soll, auch wirklich misst und nicht irgendein anderes“ (Moosbrugger & Kevala, 2012, S. 13). Bezogen auf curriculumbasierte Messverfahren sind vor allem die Bestimmung der Inhalts- sowie Konstruktvalidität von entscheidender Bedeutung (Deno, 2003; Foegen et al., 2007; Fuchs, 2004). Im Rahmen von Prävention schulischer Minderleistungen, in dem auch CBM-Verfahren angesiedelt sind, ist darüber hinaus auch die Kriteriumsvalidität ein wesentlicher Aspekt. In folgenden Abschnitten werden demnach die auch schon zuvor von Deno und Fuchs (1987) aufgeworfenen Fragen thematisiert, nämlich: 1. Was soll gemessen werden? 2. Wie soll gemessen werden? Dabei bezieht sich die erste Frage auf Aspekte der Validität, die zweite auf Aspekte der Reliabilität.

3.3.2.1 Was soll gemessen werden?

Die Antwort auf die Frage was ein CBM messen soll, führt zur Operationalisierung der Testitems. Dabei, so Deno (2003a), stehen Gesichtspunkte der Validität im Vordergrund: „*The first part of the strategy – initial task selection – is based on research using criterion validity paradigm to select tasks that would seem to be the best candidates for repeated performance measurement”* (S. 7). Zur Auswahl von Testitems beschreibt Lynn S. Fuchs in einem Artikel im Jahr 2004 zwei Wege: das *curriculum sampling* und die Auswahl sog. *robust indicators*. Beide Vorgehensweisen werden nachfolgend näher beschrieben.

Beim curriculum sampling werden systematische Stichproben von Aufgaben gezogen, welche alle fachspezifischen, bis zum Ende des Schuljahres zu erlernenden Fähigkeiten abfragen. Diese Aufgaben

werden so auf die verschiedenen parallelen CBM-Verfahren verteilt, dass bei jeder der regelmäßigen Messungen Tests eingesetzt werden, die zum einen die Fähigkeiten des gesamten Schuljahres abfragen und zum anderen ähnliche Aufgabenstellungen und -schwierigkeiten umfassen (L.S. Fuchs, 2004). Die dadurch gegebene Nähe zum Curriculum stellt einen Vorteil einer solchen Vorgehensweise dar, jedoch beziehen sich so erstellte Verfahren dann auch nur auf dieses spezifisch ausgewählte Curriculum. Der Einsatz über Klassenstufen hinweg ist somit bspw. nicht möglich (Foegen et al., 2007). Ein solches Vorgehen liegt z. B. den in Abschnitt 3.2.2 beschriebenen skills-based measures (SBM) zugrunde und wurde vor allem bei der Konzeption von CBM-Verfahren des Typs computation sowie concepts & application verwendet (Foegen et al., 2007).

Bei der Methode der robust indicators müssen Aufgabenformate entwickelt werden, welche in hohem Maß mit den verschiedenen für den jeweiligen Lernbereich bedeutsamen Kompetenzen korrelieren (L.S. Fuchs, 2004). Dieser Weg wird für die Erarbeitung von GOM bzw. DIBS verfolgt. Deno, Mirkin und Chiang (1982) haben u.a. herausgefunden, dass die Leseflüssigkeit, gemessen über die Anzahl der in einer bestimmten Zeitspanne richtig gelesenen Wörter, in beträchtlichem Maße mit den allgemeinen Leseleistungen von Kindern zusammenhängt, die somit als ein solcher robuster Indikator für die Lesekompetenz angesehen werden kann. Auf diese Weise erstellte CBM-Verfahren haben den Vorteil, dass sie nicht an spezifische Curricula gebunden sind, jedoch bieten sie keine ausdifferenzierten Informationen über die Leistungen der Kinder, da die Indikatoraufgaben mehrere Kompetenzen aggregiert abfragen. Weiterhin stellt sich hier die Frage, ob für jeden Lernbereich auch solche Indikatoritems gefunden werden können und wie diese konkret aussehen (Foegen et al., 2007).

Laut Kelley, Hosp und Howell (2008) ist es für den Bereich der Mathematik nur schwer möglich, GOM auf Basis prägnanter Indikatoraufgaben zu konstruieren, welche alle wesentlichen inhaltlichen Aspekte dieses Lernbereichs in sich vereinen, da diese Domäne durch viele aufeinander aufbauende Fähigkeitsbereiche zu komplex ist.

> *„Because math does not include a capstone task, developing GOMs for math CBM has proved to be a challenge. No single interactive measure clearly and accurately communicates generalized math skills because math is a multitopic area (e.g., application, computation, and problem solving); therefore, developing one GOM for math competence has proven difficult"* (Kelley et al., 2008, S. 253).

Auch Foegen et al. (2007) werfen die Frage auf, ob die Aneignung von Kompetenzen zur Bewältigung verschiedener Anforderungsbereiche des Mathematikunterrichts nur mittels robuster Indikatoren gewonnener CBM-Verfahren Rechnung getragen werden kann: „Perhaps it is the case that learning mathematics, unlike reading, requires the acquisition of increasingly differentiated types of knowledge as students move to higher levels of development" (Foegen et al., 2007, S. 137).

Geht es jedoch darum, Verfahren zur Identifikation von Kindern mit Schwierigkeiten beim Erlernen mathematischer Kompetenzen zu entwickeln, erscheint die zeitlich begrenzte Bearbeitung grundlegender mathematischer Operationen oder Wissenselementen im Sinne von curriculumbasierten Messverfahren ein robuster Indikator zu sein, hängt doch die Fähigkeit des schnellen und präzisen Faktenabrufs mathematischer Inhalte wesentlich mit dem erfolgreichen Aneignen allgemeiner mathematischer Kompetenzen zusammen. So wurde an vielen Stellen in der Literatur beschrieben, dass eine geringe Geschwindigkeit sowie Präzision beim Faktenabruf

mit Rechen- bzw. Lernschwächen assoziiert sind (Garnett & Fleischner, 1983; Geary, 2004; Gersten, Jordan & Flojo, 2005; Jordan et al., 2003). Die Schnelligkeit und Genauigkeit, mit welcher eine Person auf einen Stimulus, sei er allgemeiner oder mathematischer Natur reagiert, wird dabei als Flüssigkeit (*fluency*) bezeichnet (Binder, 1996). Analog zur Leseflüssigkeit kann die Rechenflüssigkeit somit Aufschluss über den Leistungsstand eines Kindes hinsichtlich seiner mathematischen Kompetenzen geben: „So for example, the degree to which a child can rapidly and correctly respond to a particular math problem reflects how well that student has mastered the skill" (VanDerHeyden & Burns, 2008, S. 215). Befindet sich ein Kind gerade bei der Aneignung einer Kompetenz wie bspw. der Addition, so ist davon auszugehen, dass es anfangs noch recht langsam bei der Bearbeitung von Additionsaufgaben vorgeht und dabei durchaus Fehler macht. Ist diese Kompetenz erst einmal eingeübt, so wird das Kind weniger Fehler machen als zuvor, jedoch bei langsamer Geschwindigkeit. Erst durch weitere Übungsphasen kann darüber hinaus eine Geschwindigkeitssteigerung erreicht werden (VanDerHeyden & Burns, 2008). Das Verhältnis von Genauigkeit und Schnelligkeit bei der Bearbeitung einer Aufgabe kann somit Hinweis darauf geben, auf welcher Stufe der Kompetenzaneignung sich ein Kind aktuell befindet (Binder, 1996; Burns, VanDerHeyden & Jiban, 2006; VanDerHeyden & Burns, 2008). Auf der Suche nach analogen Indikatoren zur Leseflüssigkeit für den Bereich der Mathematik wurde versucht, verschiedene Aufgabenformate zu konzipieren, welche als Hinweisgeber für die allgemeine Mathematikleistung fungieren sollen. Speziell im Bereich früher mathematischer Kompetenzen sind alle CBM-Verfahren nach dem Prinzip der robust indicators entwickelt (Foegen et al., 2007). Die bis dato wichtigsten early numeracy CBM sind die in Abschnitt 3.2.4 beschriebenen Verfahren number identification, quantity discrimination und missing number (Chard et al., 2005; Clarke & Shinn, 2004; Foegen et al., 2007). Jedoch bleibt fraglich, inwieweit diese Verfahren, einzeln betrachtet, allgemeine

mathematische Kompetenzen widerspiegeln. Weiterhin erinnert das Abarbeiten von seitenweisen Aufgabenblöcken, wie es bei diesen skills-based measures der Fall ist, an traditionelle Türmchenaufgaben, welche von Praktikern oft als ein „drill and kill" Ansatz verstanden und eher abgelehnt werden (Jiban & Deno, 2007, S. 79f). Einen weiteren großen Kritikpunkt stellt die Diskrepanz zwischen den in CBM abgefragten Kompetenzen und denen, die in den geltenden Bildungsstandards gefordert werden, wodurch Zweifel an der inhaltlichen Validität der Verfahren aufkommen.

Die gleiche Problematik zeigt sich auch bezüglich der in Deutschland bundesweit geltenden Bildungsstandards (Kultusministerkonferenz, 2004b), welche in Abschnitt 2.9 etwas näher beschrieben wurden. CBM im Mathematikunterricht des Schuleingangsbereichs fokussieren eher inhaltsbezogene und weniger allgemeine mathematische Kompetenzen (Foegen et al., 2007). Aber stehen diese Verfahren damit tatsächlich im Widerspruch zu den Forderungen der Bildungsstandards? Die in den CBM abgefragten Aspekte, vor allem im Schuleingangsbereich, stellen grundlegende arithmetische Kompetenzen dar. Damit eng verbunden sing mathematische Einsichten und Basiswissen. Das Automatisieren von Basiswissen und grundlegenden mathematischen Kompetenzen wird als Fundament für die Bearbeitung höherer mathematischer Aufgaben, speziell bei Kindern mit Schwierigkeiten beim Erwerb des Rechnens, angenommen (Cumming & Elkins, 1999; Garnett, 1992; Pellegrino & Goldman, 1987; VanDerHeyden & Burns, 2008). So schreiben auch Jiban, Deno und Foegen resümierend (2009): „Fluency or automaticity with basic computational skills may support one's ability to succeed at higher level tasks such as problem-solving (Gagne, 1983; Hasselbring, Goin, & Bransford, 1988)" (S. 3). Somit stellen CBM-Verfahren, welche einzelne im Bereich der Mathematik relevanten Kompetenzen prüfen, ein sinnvolles Hilfsmittel zur Identifikation von Kindern mit Schwierigkeiten beim Rechnen dar,

auch wenn sie nicht das gesamte Kontinuum mathematischer Kompetenzen erfragen (VanDerHeyden & Burns, 2008). Eine kombinierte Betrachtung der Schülerergebnisse in verschiedenen CBM kann an dieser Stelle hilfreich sein, um einen differenzierteren Einblick in den Könnensstand der Kinder hinsichtlich verschiedener mathematischer Aspekte zu erhalten. Speziell für den Mathematikunterricht in der ersten Klasse erscheint es essentiell, eine Komposition von verschiedenen CBM-Verfahren zu konzipieren, um sich ein umfassendes Bild über die Fähigkeiten eines Kindes zu verschaffen (Hosp et al., 2007). Dabei sollte neben den konkreten Rechenkompetenzen auch das Vorhandensein zentraler mathematischer Einsichten und Konzeptvorstellungen im Sinne der early numeracy measures geprüft werden.

3.3.2.2 Wie soll gemessen werden?

Bei der Konzeption von CBM-Verfahren spielen neben Aspekten der Validität auch Fragen der Reliabilität eine tragende Rolle, auf welche entsprechend nachfolgend eingegangen werden soll.

Im Abschnitt zuvor wurden Möglichkeiten der Gewinnung inhaltlich sinnvoller Aufgaben geschildert. Diese Grundmenge an repräsentativen Items muss möglichst homogen gehalten werden, damit die daraus generierten Aufgabensets von ähnlichem Schwierigkeitsgrad sind. Da die CBM-Tests zeitlich eng begrenzt sind, so Christ et al. (2008), spielt die Homogenität der Itemanordnung innerhalb der einzelnen Testverfahren eine entscheidende Rolle. Strathmann und Klauer (2008) merken diesbezüglich an:

> *„Häufig liegt es nahe, die so definierte Grundmenge in pädagogisch relevante Teilmengen zu zerlegen. Beispielsweise ist es üblich und sinnvoll, für die*

Grundschulmathematik eine Differenzierung nach Addition, Subtraktion, Multiplikation und Division vorzunehmen“ (S. 7).

Darüber hinaus sollten jedoch auch innerhalb solcher von Strathmann und Klauer (2008) beschriebenen groben Teilmengen noch weitere Kategorien ausdifferenziert werden, um eine möglichst hohe Homogenität des Itempools zu gewährleisten. Eine derartige Verfahrensweise geht auf Gagné (1962) zurück und wird als sog. *Lernhierarchie* bezeichnet. Das Ergebnis eines solchen Vorgehens ist „eine hierarchisch geordnete Menge von Lernvoraussetzungen“ (Klauer, 1978, S. 861, vgl. auch Abbildung 22). Diese Lernvoraussetzungen müssen inhaltlich begründet operationalisiert werden, indem sie sich an bestehenden (möglichst empirisch belegten) Modellen oder Theorien orientieren, um das Kriterium der Inhaltsvalidität zu erfüllen.

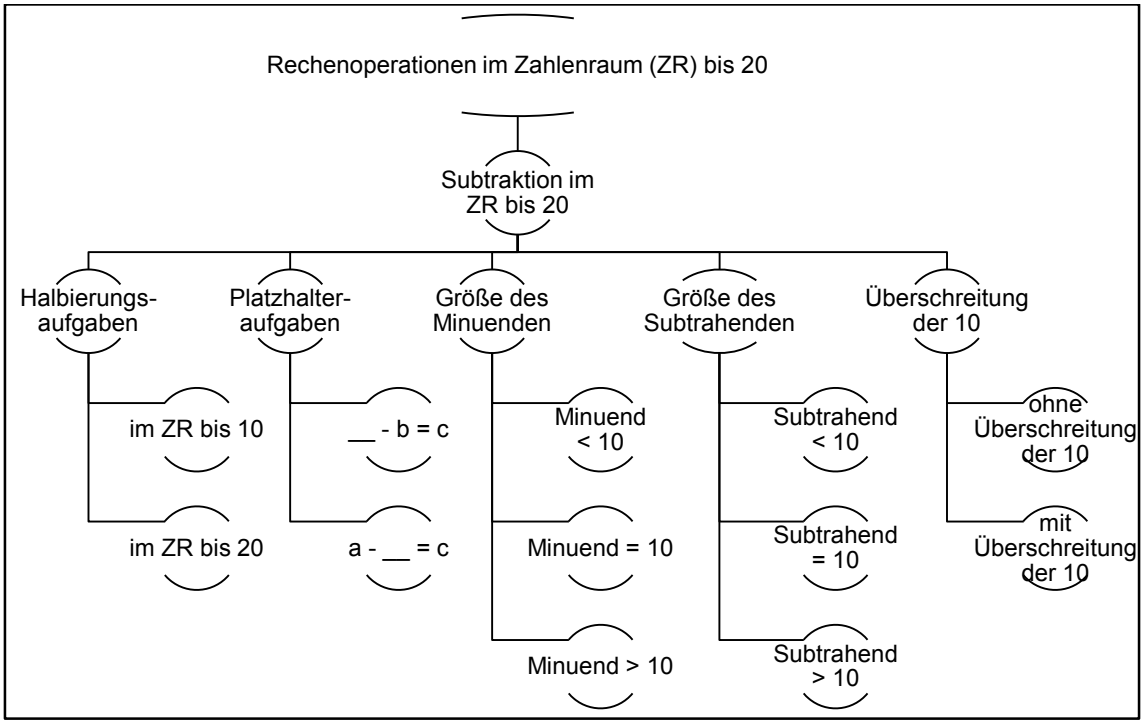


Abbildung 22: Ausschnitt einer möglichen Ordnung von Teilfertigkeiten innerhalb einer Lernhierarchie am Beispiel der Subtraktion im Zahlenraum (ZR) bis 20 (in Anlehnung an Klauer, 1978)

In einer ersten Phase der Konzeption von curriculumbasierten Messverfahren müssten demnach Itempools für die einzelnen zu erfassenden Kompetenzbereiche generiert werden. Aufgrund der Ausdifferenzierung der erforderlichen Kompetenzbereiche in Teilfertigkeiten, ist davon auszugehen, dass die Items innerhalb einer solchen Teilkategorie als gleich schwer vermutet werden können. In Abbildung 22 wird eine solche Struktur ausschnittsweise exemplarisch am Beispiel der Subtraktion im Zahlenraum bis 20 dargestellt. So lässt sich unterstellen, dass die Aufgaben zur Subtraktion im Zahlenraum bis 20 sich, bspw. hinsichtlich der Größe des Subtrahenden, in ihrer Schwierigkeit unterscheiden. Vermutlich werden Aufgaben mit einem Subtrahenden größer Zehn dabei schwieriger sein als solche mit einem Subtrahenden kleiner bzw. gleich Zehn. In einer zweiten Phase sollten dann mittels der erstellten Itempools verschiedene CBM-Verfahren aus den ermittelten Aufgabenmengen generiert werden. Erst durch die Sicherstellung einer gleichen bzw. ähnlichen Aufgabenstruktur der einzelnen Verfahren kann die Homogenität der Tests als gegeben angenommen werden.

Ein solches Vorgehen zur Absicherung gleicher Schwierigkeitsgrade der Verfahren ist jedoch nicht bei allen verfügbaren US-amerikanischen CBM für den Bereich der Mathematik gewährleistet (Christ et al., 2008). Die Autoren kritisieren die bis dato übliche Vorgehensweise bei der Konstruktion von CBM-Verfahren im Bereich des Rechnens, bei der die Testitems jeweils zufällig auf einem Testblatt angeordnet wurden (Shapiro, 1996; Shinn, 2002; Shinn et al., 1997). Testergebnisse können dadurch verzerrt werden, dass die ersten Items des einen Tests sehr leicht sind, in einem weiteren Test hingegen sehr schwer. Dies hat maßgeblichen Einfluss auf die wissenschaftliche Güte der Verfahren. Um diesen Missstand zu überwinden, schlagen Christ und Vining (2006) die Methode der *stratifizierten Itemanordnung* vor. Dabei wird von einer zufälligen

Anordnung der Testitems abgesehen. Die Aufgaben werden dafür systematisch je nach inhaltlichen Aspekten spaltenweise verteilt (Christ & Vining, 2006). In Abbildung 23 sind die Vorgehensweisen veranschaulicht gegenübergestellt. Dabei lässt sich bei Verfahren mit zufälliger Itemplatzierung gegenüber denen, welche mittels der von Christ und Vining (2006) empfohlenen stratifizierten Vorgehensweise konstruiert wurden, eine Inkonsistenz der Struktur innerhalb und über die Parallelformen hinweg erkennen.

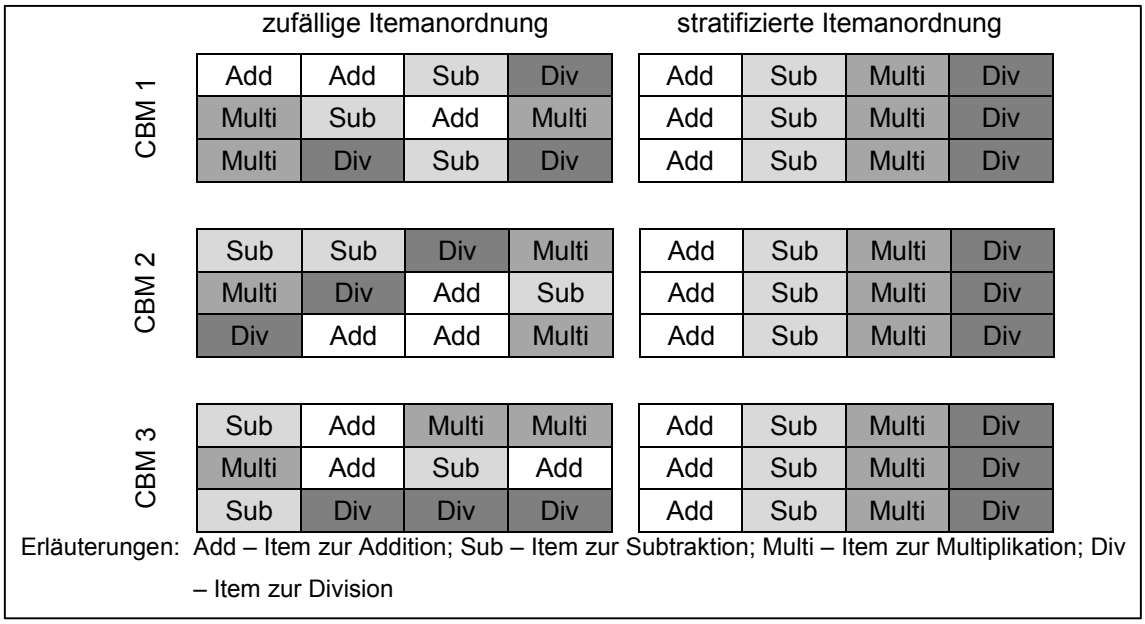


Abbildung 23: Beispiel zur Veranschaulichung der Itemanordnung in Parallelformen von Mathematik-CBM: zufällige Itemanordnung (links) vs. stratifizierte (nach inhaltlichen Aspekten geordnete) Items (rechts) (entnommen und modifiziert nach Christ & Vining, 2006, S. 390)

In einer Studie mit insgesamt 219 Kindern (Klasse1: N = 40; Klasse zwei: N = 41; Klasse drei: N = 44; Klasse vier: N = 44; Klasse 5: N = 50) untersuchten Christ und Vining (2006) mittels varianzanalytischer Verfahren, welche der beiden Vorgehensweisen zur CBM-Konstruktion am geeignetsten erscheint. Dazu wurden die Kinder in zwei Gruppen aufgeteilt, wobei die eine Gruppe Verfahren mit zufälliger Itemanordnung bearbeitete und die andere Verfahren mit stratifizierter Itemanordnung bekam. Die Autoren verglichen anschließend den Anteil der Varianz, welcher jeweils auf die

Leistungen des Probanden zurückzuführen ist. Während in den Klassenstufen 1 und 2 nur marginale Unterschiede in der Varianzaufklärung zwischen den beiden Verfahrensweisen ermittelt wurden (Klasse eins: 65 % vs. 67 %; Klasse zwei: 67 % vs. 75 %), zeigten sich in höheren Klassenstufen deutlich höhere zu Gunsten der Verfahren mit stratifizierter Itemanordnung (Klasse drei: 7 % vs. 72 %; Klasse vier: 46 % vs. 61 %; Klasse 5: 37 % vs. 60 %). Trotz der Anomalie in Klasse drei, welche nur vorsichtig interpretiert werden darf, geben die Ergebnisse Hinweise darauf, dass durch eine stratifizierte Itemanordnung konsistentere Verfahren erstellt werden können, ein Fakt der speziell bei Speedtests wie CBM eine tragende Rolle spielt (Christ & Vining, 2006).

Die von L.S. Fuchs (2004) beschriebene Stufe eins zur systematischen Evaluation von CBM-Verfahren umfasst Analysen zur wissenschaftlichen Güte der Verfahren. Dies umfasst vor allem Aspekte der Objektivität, Reliabilität und Validität, die auch im Rahmen der psychologischen und pädagogischen Diagnostik an Testverfahren gestellt werden. Diese Analysen für sich genommen geben jedoch lediglich Aufschluss über die Einsatzfähigkeit eines Verfahrens als statisches Messinstrument im Sinne einer Statusdiagnostik. Damit ist jedoch noch nicht nachgewiesen, dass sich der Test zur Dokumentation von Lernverläufen eignet. Diesem Aspekt soll auf der von L.S. Fuchs (2004) beschriebenen zweiten Stufe nachgegangen werden.

3.3.3 Notwendige Analysen hinsichtlich der von L.S. Fuchs (2004) vorgeschlagenen Stufe 2

Ziel der curriculumbasierten Messverfahren ist es, den Lernverlauf von Kindern zu erfassen, um daraus Schlüsse hinsichtlich des Unterrichts bzw. der Förderung zu ziehen. Um dies zu gewährleisten, müssen eingesetzte CBM-Verfahren nachweislich fähig sein,

Leistungsveränderungen über die Zeit reliabel und sensibel zu messen (Shin, Espin, Deno & McConnell, 2004). Dieses Kriterium soll laut L.S. Fuchs (2004) innerhalb der zweiten Stufe der Erforschung von CBM-Verfahren nachgewiesen werden. Als einen Ansatz dazu schlägt Klauer (2011) eine Untersuchung mittels eines Zwei-Gruppen-Versuchsplans mit einer Treatment- und einer Kontrollgruppe sowie eines eingänglichen Prätests und eines abschließenden Posttests, in welchen das jeweilige Kriterium (hier die Rechenkompetenzen) erfasst wird, vor:

> *„Die hierzu übliche und vielfach auch angewandte Versuchsanordnung besteht darin, dass der Leistungsstand erhoben wird, danach ein Treatment erfolgt, um anschließend den Leistungsstand erneut zu testen. Dabei ist es zweckmäßig, eine Kontrollgruppe ohne Treatment einzusetzen, alleine schon um den Einfluss der Testwiederholung berücksichtigen zu können“* (S. 212).

Eine Mittelwertverschiebung in der Treatmentgruppe von Prä- zu Posttest kann dann als Hinweis auf die Änderungssensibilität interpretiert werden. Jedoch merkt Klauer (2011) an, dass diese Vorgehensweise sehr unspezifisch erscheint. Klauer weist darüber hinaus auf ein generelles Dilemma hinsichtlich der Reliabilität von Verfahren zur Lernverlaufsmessung hin. Sind hohe Retest- bzw. Paralleltest-Reliabilitäten von CBM-Verfahren wünschenswert, um Hinweise auf die Homogenität der Verfahren zu erhalten, so sind diese jedoch nicht vereinbar mit der Erfassung unterschiedlicher Leistungsentwicklungen. Werden hohe Reliabilitätskoeffizienten ermittelt, kann dies auf die gleiche Entwicklung aller Probanden schließen lassen, was bezogen auf die schulische Leistungsentwicklung verschiedener Kinder eher als unwahrscheinlich angenommen werden kann. Entwickeln sich die Probanden unterschiedlich hinsichtlich des zu erfassenden Merkmals, so schlägt sich dies in niedrigen Reliabilitätskoeffizienten

wider. Klauer (2011) zieht daraus drei wesentliche Schlüsse hinsichtlich der Bestimmung der Änderungssensibilität von Verfahren zur Erfassung des Lernverlaufs:

1. Ein änderungssensibler Test weist nur mittlere Retest- bzw. Paralleltest-Reliabilitäten auf.
2. Die Korrelationen zwischen änderungssensiblen Tests werden kleiner, je größer der Zeitraum zwischen den Erhebungszeitpunkten ist.
3. Die Splithalf-Reliabilität eines änderungssensiblen Tests ist höher als die Retest- bzw. Paralleltest-Reliabilitäten (Klauer, 2011).

Ausgehend von den Ausführungen in Abschnitt 2.10 zur Entwicklung von Kindern mit Schwierigkeiten beim Erwerb des Rechnens sollten sich bei einem änderungssensiblen Verfahren zudem zwischen Kindergruppen mit verschieden ausgeprägtem mathematischen Vorwissen unterschiedliche Lernverläufe abbilden lassen. D. h, Kinder mit einem lediglich gering ausgeprägtem mathematischen Vorwissen zu Beginn der Klasse eins sollten theoriekonform geringere Lernzuwächse erzielen als Kinder mit einem breiten Vorwissen. Shin et al. (2004) geben zusätzlich zu bedenken, dass die steigende Varianz der Testergebnisse über die Zeit weitere Anhaltspunkte für die Zuverlässigkeit der gemessenen Leistungsänderungen gibt:

> „*Increased variances in test scores across time lead to increased variability in individual growth rates across time. Subsequently, such increased heterogeneity of growth rates among individuals is expected to result in higher reliability of growth-rate estimates (Bryk & Raudenbush, 1987, 1992; Willet, 1989b; Zimmerman & Williams, 1982)”* (S. 138).

Die von Klauer genannten Hinweise, ergänzt durch die Überlegungen von Shin et al. (2004) stellen eine solide Basis zur Analyse der Änderungssensibilität von CBM-Verfahren dar. Wie dem Nachweis der Änderungssensibilität durch entsprechende statistische Methoden Rechnung getragen werden kann, wird in Abschnitt 6.4 beschrieben.

3.3.4 Notwendige Analysen hinsichtlich der von L.S. Fuchs (2004) vorgeschlagenen Stufe 3

Zusätzlich zu den Analysen auf den Forschungsstufen eins und zwei postuliert L.S. Fuchs (2004) schließlich die Untersuchung der Einsetzbarkeit von CBM-Verfahren im schulischen Alltag innerhalb der dritten Forschungsstufe. „This is where studies are conducted to determine whether practitioners can use the CBM information to improve instructional decisions and student achievement (e.g., Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, Hamlett, & Stecker, 1991)" (L.S. Fuchs, 2004, S. 189). Auf der dritten Stufe soll im Wesentlichen der Frage nachgegangen werden, ob Lehrkräfte die regelmäßigen Lernstandserhebungen mittels CBM im Unterricht realisieren können und ob sie anhand der Ergebnisse ihrer Schüler in den CBM-Verfahren tatsächlich Hinweise für die Planung von Unterrichts- bzw. Fördermaßnahmen ableiten und Leistungssteigerungen erzielen können. Aspekte für eine solche Untersuchung repräsentieren dabei einerseits die erlebte Effizienz bezüglich des zeitlichen Aufwands zur Durchführung und Auswertung der CBM durch die Lehrkräfte (Deno, 2003b) und andererseits die Bedeutsamkeit der Ergebnisse für sie. Darüber hinaus spielt der empirische Nachweis der Effektivität regelmäßig eingesetzter CBM-Verfahren eine zentrale Rolle. Analysen entsprechend der Stufe 1 stellen zwar einen wesentlichen Schritt in der Erforschung von CBM-Verfahren dar, belegen jedoch nur die Einsetzbarkeit der Tests als Screening-Werkzeug. Dies allein ist sicherlich schon ein nützlicher

Aspekt, jedoch werden die den CBM inne liegenden Vorteile, wie sie Deno (2003b) beschreibt, erst durch weiterführende Analysen auf Stufe 2 und 3 nachgewiesen. Die Notwendigkeit der Analyse aller von L.S. Fuchs (2004) geforderten Evaluationskriterien der drei Forschungsstufen wird dadurch deutlich. Foegen et al. (2007) sehen hier – wie in Abschnitt 3.2.8 erläutert wird – noch Forschungsbedarf, da lediglich für das MBSP computation von L.S. Fuchs et al. (1998) Studien vorliegen, welche die Erfüllung aller dieser Kriterien belegen (Foegen et al., 2007). Die Autoren erklären die Erforschung von CBM-Verfahren für den Bereich der Mathematik somit als noch nicht zufriedenstellend abgeschlossen: „The search for technically and theoretically appropriate measures, largely resolved in CBM of reading, remains active in mathematics" (Foegen et al., 2007, S. 137). In der vorliegenden Studie sollen die CBM-Verfahren daher hinsichtlich der von L.S. Fuchs (2004) gegebenen Kriterien aller drei Forschungsstufen untersucht werden.

Neben den Richtungsweisenden Vorschlägen zur Konzeption und Evaluation von CBM-Verfahren für den Einsatz im schulischen Setting, fasst L.S. Fuchs (2004) den Forschungsstand hinsichtlich der Entwicklung von curriculumbasierten Messungen bis zum Jahr 2004 zusammen. Auf die von Deno (1985) aufgeworfene Frage hin, ob sich CBM als alternative Methode zu den bis dahin verbreiteten Maßnahmen des curriculum-based assessment zur Lernverlaufsdokumentation in der Schule eignen, kommt sie zu folgendem Schluss:

> „*In 1985, Stan Deno wrote the now classic article, 'Curriculum-based measurement: The emerging alternative.' Thirty years after Deno's conceptualization of CBM in the early 1970s and hundreds of studies later, it may be time to redesignate CBM progress monitoring as the 'validated alternative'"* (S. 192).

Hat L.S. Fuchs‘ Aussage, welche sich an eine US-amerikanische Forschungstradition anlehnt, jedoch auch allgemeine Gültigkeit? Der Frage, inwieweit sich curriculumbasierte Messungen als Verfahren zur Dokumentation von Lernfortschritten im Bereich des mathematischen Erstunterrichts als eine Maßnahme zur Prävention von Rechenschwäche auch im deutschen Schulsystem konzipieren und anwenden lassen, soll in den folgenden Abschnitten nachgegangen werden.

4 CBM für den mathematischen Erstunterricht

4.1 Allgemeine Aspekte zur Konzeption und Evaluation von CBM für den mathematischen Erstunterricht

Die in den Kapiteln 2 und 3 ausgearbeiteten Erkenntnisse und Schlüsse zugrunde legend, werden in nachfolgenden Abschnitten wesentliche Aspekte bei der Erarbeitung von CBM für den Anfangsunterricht Mathematik zusammengetragen. Dazu erfolgt zunächst eine eher allgemeine und übergreifende Beschreibung struktureller sowie inhaltlicher Überlegungen. Im Anschluss werden die einzelnen, in der vorliegenden Studie untersuchten Verfahren im Detail dargestellt.

Die in Kapitel 1 beschriebenen Studien von Aunola et al. (2004), Krajewski (2003), Mazzocco und Thompson (2005) und Stern (2003) haben gezeigt, dass ein erfolgreiches Aneignen erster mathematischer Kompetenzen im Anfangsunterricht entscheidend für den späteren Lernerfolg im Schulfach Mathematik ist. Dass Kinder jedoch mit sehr unterschiedlichen Lernvoraussetzungen in die Schule kommen, zeigen die in Kapitel 2 beschriebenen Untersuchungen (Grassmann et al., 1995; Grassmann et al., 2002; Hengartner & Röthlisberger, 1995; van Heuvel-Panhuizen, 1995; Kesting, 2005; Knapstein & Spiegel, 1995; Rinkens, 1997; Rinkens & Eilerts, 2001; Schipper, 1996, 1998; Schmidt, 1982; Selter, 1995). Um eine erfolgreiche schulische Prävention von Lernschwierigkeiten in Mathematik bereits in Klassenstufe 1 zu ermöglichen, müssen die Kinder, welche Schwierigkeiten beim Erwerb erster mathematischer Kompetenzen aufweisen, möglichst frühzeitig identifiziert werden. Der Einsatz der im Rahmen dieser Studie konzipierten CBM soll dies leisten. Dabei sollen sie keine lediglich einmalig erhobenen, punktuellen Informationen über die Schüler liefern, sondern die

Lernentwicklung aller Kinder über den Verlauf des ersten Schuljahres abbilden. Um dies zu erreichen, muss jedes erarbeitete CBM in einer ausreichenden Anzahl an parallelen Versionen vorliegen. Dabei sollten die innerhalb der CBM verwendeten Items ein möglichst breites Spektrum an Schwierigkeitsgraden abdecken. Außerdem muss sichergestellt sein, dass die einzelnen Aufgabenformate ohne umfassende zusätzliche, schriftliche oder mündliche Durchführungsanweisungen auskommen, sodass sie den Kindern einer ersten Klasse leicht verständlich sind.

Inhaltlich muss sichergestellt sein, dass die CBM die wesentlichen mathematischen Schwerpunkte der Klassenstufe 1 bzw. basale mathematische Kompetenzen, welche im vorschulischen Rahmen bereits bei vielen Kindern ausgeprägt sein könnten, abbilden. Eine gute Basis für die Gestaltung solcher Testverfahren stellt die Vielfalt der in den USA für den Kindergarten bzw. die Klassenstufe 1 verfügbaren und meist auch evaluierten CBM dar. Für einen Rückgriff auf diese Verfahren spricht die mittlerweile fast 30 Jahre andauernde Forschungstradition in den Vereinigten Staaten (Deno, 1985). In Abschnitt 3.2.4 wurden die in den USA zur Verfügung stehenden Instrumente ermittelt und analysiert. Dort gibt es für den Anfangsunterricht Mathematik viele verschiedene CBM, welche unterschiedliche Aspekte des Erwerbs arithmetischer Kenntnisse abfragen. Wie in Abschnitt 3.2.4 bereits beschrieben, lassen sich in der Regel zwei verschiedene Sorten von CBM im Anfangsbereich Mathematik unterscheiden. So gibt es Verfahren, welche grundlegende mathematische Kompetenzen erfassen, sog. CBM des Typs early numeracy. In der Literatur wird in diesem Zusammenhang von verschiedenen CBM berichtet, die die Kenntnis der Zahlen und des Zahlenstrahls, das numerische Verständnis von Zahlen sowie Zählkompetenzen erfassen (Chard et al., 2005; Clarke & Shinn, 2004; Lembke & Foegen, 2005; VanDerHeyden et. al., 2004; VanDerHeyden et. al., 2001). Foegen et al. (2007) berichten, dass

Verfahren, welche die Fähigkeit zur Identifikation von Zahlen (number identification, durch lautes Vorlesen von vorgegebenen Zahlen), zum Vergleich von Mengen (quantity discrimination, durch Nennen oder Ankreuzen der größeren zweier Zahlen) und zur Seriation von Zahlen (missing number, durch Nennen oder Ergänzen der fehlenden Zahl eines vorgegebenen Ausschnitts der Zahlenreihe) dabei als die am besten empirisch abgesicherten Verfahren zur Messung erster mathematischer Kompetenzen gelten. Verfahren dieses Typs werden in der hier vorliegenden Arbeit hinsichtlich der Passung des Einsatzes im Mathematikunterricht der Grundschulklassenstufe 1 analysiert.

Neben den CBM des Typs early numeracy zum Überprüfen allgemeinerer sowie basaler mathematischer Einsichten werden in der US-amerikanischen Literatur weitere CBM geschildert, welche eher die Rechenfertigkeiten der Schüler fokussieren (computation). Diese CBM erfassen die Fähigkeit zur Anwendung der grundlegenden arithmetischen Operationen Addition und Subtraktion sowie später auch Multiplikation und Division in einem zeitlich begrenzten Rahmen (L.S. Fuchs & D. Fuchs 1992; Tindal, Marston et al., 1983). Vor allem die CBM des Programms Monitoring Basic Skills Progress measures (MBSP, L.S. Fuchs et al., 1998, 1999) haben sich im Kreis der CBM des Typs computation empirisch am meisten bewährt (Christ et al., 2008; Foegen et al., 2007).

Die in dieser Studie erarbeiteten CBM greifen die oben genannten Forschungsergebnisse aus den USA auf. Es wurde eine Kombination verschiedener Verfahren zur Lernverlaufsdiagnostik für die Klassenstufe 1 entwickelt. Zunächst wurden Verfahren erarbeitet, welche wesentliche grundlegende mathematische Konzepte, im Sinne der CBM des Typs early numeracy, abfragen. Dazu zählen die Folgenden:

1. Mengenvergleich (in Anlehnung an CBM des Typs quantity discrimination)
2. Zahlenreihe (in Anlehnung an CBM des Typs missing number)
3. Zahlenlesen (in Anlehnung an CBM des Typs number identification)
4. arithmetische Basiskompetenzen (in Anlehnung an CBM des Typs early numeracy)
5. Zahlzerlegung (hierzu wurde keine Entsprechung in US-amerikanischen CBM gefunden)

Darüber hinaus wurden in der vorliegenden Studie Verfahren konzipiert, welche die Rechenfertigkeiten erster, in Klassenstufe 1 thematisierter mathematischer Operationen prüfen. Dazu zählen die folgenden Verfahren:

6. Addition im Zwanzigerraum (in Anlehnung an CBM des Typs computation)
7. Subtraktion im Zwanzigerraum (in Anlehnung an CBM des Typs computation)

In Anlehnung an diese inhaltliche Einteilung wird für die im Rahmen der vorliegenden Studie erarbeiteten CBM eine Gliederung in zwei Niveaustufen vorgenommen (vgl. Tabelle 7). Dabei gehören CBM, welche grundlegende mathematische Kompetenzen abfragen (CBM 1 – 4), der Niveaustufe 1 an und finden in der ersten Hälfte des ersten Schuljahres Anwendung. Zur Niveaustufe 2, welche innerhalb der zweiten Hälfte des ersten Schuljahres eingesetzt werden, zählen die CBM 5 – 7. Diese Einteilung nach Schulhalbjahren soll auftretende Boden- oder Deckeneffekte verhindern bzw. minimieren. Zu leichte Items führen dazu, dass sehr viele Kinder schnell alle Aufgaben eines Verfahrens lösen können und über die Zeit die maximal erreichbare Punktzahl erhalten. Sind die Aufgaben zu schwer, wird der Großteil der Kinder über die Zeit nur geringe bzw. keine Lernzuwächse

erzielen können. In beiden Fällen ist die Dokumentation von Lernfortschritten nicht möglich.

Es ist zu erwarten, dass ein Großteil der Kinder entweder bereits über ein ausgeprägtes mathematisches Vorwissen zu Schulbeginn verfügt (entsprechend der Studien in Kapitel 2) bzw. sich dieses Wissen innerhalb der ersten Schuljahreshälfte aneignen wird. Der Lernverlauf der Rechenfähigkeiten kann erst dokumentiert werden, wenn ein Lernzuwachs in ausreichendem Maß gegeben ist (Clarke & Shinn, 2004). “For many students, this means that M-CBM can be used in mid to late first grade at best. Prior to this point, nearly all students will have initial scores of 0 on M-CBM probes and thus early identification is hampered by floor effects” (Clarke & Shinn, 2004, S. 235 f.).

Das CBM „Zahlzerlegung im Zwanzigerraum“ wird hier der Niveaustufe 2 zugeordnet, da es zur Lösung der Aufgaben schon eine weitreichende Einsicht in das Teile-Ganzes-Konzept bedarf und eine relativ große Nähe zu den Anforderungen der CBM „Addition im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ aufweist. Auch das CBM „Arithmetische Basiskompetenzen“ nimmt eine Sonderstellung innerhalb der Niveaustufe 2 ein. Es wird über das ganze Schuljahr hinweg eingesetzt, um differenzierte Aussagen speziell über den Lernverlauf schwächerer Kinder zu ermöglichen. Es gehört somit der Niveaustufe 1 und 2 an.

Tabelle 7: Übersicht über die curriculumbasierten Messungen im ersten Schuljahr (Niveaustufe 1 sowie 2)

Niveaustufe 1	Niveaustufe 2
„Arithmetische Basiskompetenzen“	„Arithmetische Basiskompetenzen“
„Zahlenlesen im Zwanzigerraum“	„Zahlzerlegung im Zwanzigerraum“
„Mengenvergleich im Zwanzigerraum“	„Addition im Zwanzigerraum“
„Zahlenreihe im Zwanzigerraum“	„Subtraktion im Zwanzigerraum“

Die Schuljahreshälfte der ersten Klasse ist in der Konzeption der hier vorgestellten CBM als Grenze bestimmt worden, da wesentliche mathematische Inhalte (z. B. Zahlenkenntnis bis 20, Orientierung im Zwanzigerraum) bis zu diesem Zeitpunkt grundlegend erarbeitet worden sein sollten. In der zweiten Hälfte der ersten Klasse wird in der Regel darauf aufbauend die Erarbeitung und das Einüben der Rechenoperationen Addition und Subtraktion fokussiert. Für jede Niveaustufe gibt es demnach vier verschiedene CBM, welche in den nachfolgenden Abschnitten im Einzelnen genauer beschrieben werden.

Grundsätzlich können alle Verfahren bis auf das CBM „Zahlenlesen im Zwanzigerraum“ als Gruppenverfahren eingesetzt werden. Das CBM „Zahlenlesen im Zwanzigerraum“ erfordert eine Testung in einer Eins-zu-Eins-Situation, da die innerhalb des Tests dargebotenen Zahlen laut vorgelesen werden sollen. Für die Bearbeitung aller hier aufgeführten CBM wird eine Bearbeitungszeit von einer Minute gegeben, innerhalb welcher die Kinder die Aufgabe haben, möglichst viele Items, möglichst korrekt zu bearbeiten. Wie in Abschnitt 3.2 beschrieben, spiegelt sich ein anhand von CBM gemessener Lernfortschritt der Kinder in der schnellen und sicheren Bearbeitung von Aufgaben wider. Um auch die Messung von Fortschritten über den gesamten Einsatzzeitraum eines CBM zu gewährleisten, werden den Kindern innerhalb der Bearbeitungszeit zum einen zu viele und teilweise auch zu schwere Aufgaben dargeboten. Dies zeigt sich vor allem zu Beginn einer Niveaustufe. Hier wird erwartet, dass die Kinder nur relativ wenige Aufgaben der eingesetzten CBM lösen, da die Sicherheit bei der Bearbeitung noch fehlt. Um Frustrationen zu vermeiden, sollte dies den Kindern vor jeder Testung bewusst gemacht werden. Dabei sollte darauf hingewiesen werden, dass es nicht schlimm sei, wenn Aufgaben noch nicht gelöst werden können bzw. die Bearbeitung zeitlich nicht geschafft wird.

Fraglich ist nun, inwieweit die im Rahmen der vorliegenden Studie erarbeiteten CBM in deutschen Schulen eingesetzt werden können. Dazu muss zunächst geprüft werden, inwieweit die mit den CBM gestellten Anforderungsbereiche sich in den Curricula und Rahmenlehrplänen der deutschen Bundesländer widerspiegeln. In diesem Zusammenhang wurde eine Analyse der Curricula und Rahmenpläne der einzelnen Bundesländer vorgenommen, deren Ergebnisse in Tabelle 8 zusammengefasst sind. Da sich die Art der Beschreibung der Unterrichtsziele in den verschiedenen Lehrplänen teilweise sehr in Umfang und Differenziertheit unterscheidet, sind nicht alle mit den CBM verknüpften inhaltlichen Anforderungen in jedem Lehrplan explizit wiederzufinden, was jedoch nicht bedeutet, dass diese Aspekte keine Relevanz besitzen. Insgesamt sind alle inhaltlichen Anforderungen der in der vorliegenden Arbeit entwickelten CBM in verschiedenen Rahmenplänen explizit wiederzufinden, was die curriculare Relevanz der Verfahren unterstreicht.

Tabelle 8: Einordnung der CBM in die Curricula und Rahmenlehrpläne der deutschen Bundesländer

	ZL	ZR	MV	AB	ZZ	AD	SUB
Baden-Württemberg	✓	✓	✓	✓		✓	✓
Bayern	✓	✓	✓	✓	✓	✓	✓
Hansestadt Hamburg	✓	✓	✓	✓	✓	✓	✓
Hessen	✓	✓	✓	✓		✓	✓
Mecklenburg-Vorpommern, Berlin, Brandenburg, Hansestadt Bremen	✓	✓	✓	✓		✓	✓
Niedersachsen	✓	✓	✓	✓	✓	✓	✓
Nordrhein-Westfalen		✓	✓	✓		✓	✓
Rheinland-Pfalz		✓	✓	✓		✓	✓
Saarland	✓	✓	✓	✓	✓	✓	✓
Sachsen		✓	✓	✓		✓	✓
Sachsen-Anhalt	✓	✓	✓	✓		✓	✓
Schleswig-Holstein		✓	✓	✓		✓	✓
Thüringen		✓	✓	✓		✓	✓

Erläuterungen: ZL – CBM „Zahlenlesen im Zwanzigerraum“; ZR – CBM „Zahlenreihe im Zwanzigerraum“; MV – CBM „Mengenvergleich im Zwanzigerraum“; AB – CBM „Arithmetische Basiskompetenzen“; ZZ – CBM „Zahlzerlegung im Zwanzigerraum“; AD – CBM „Addition im Zwanzigerraum“; SUB – CBM „Subtraktion im Zwanzigerraum“; BW – Baden-Württemberg (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2004); BY – Bayern (Staatsinstitut für Schulqualität und Bildungsforschung München, 2000); HH – Hansestadt Hamburg (Behörde für Schule und Berufsbildung Hamburg, 2011); HE – Hessen (Hessisches Kultusministerium, 1995); MV, BE, BB, HB – Mecklenburg-Vorpommern, Berlin, Brandenburg, Hansestadt Bremen (Ministerium für Bildung, Jugend und Sport des Landes Brandenburg, Senatsverwaltung für Bildung, Jugend und Sport Berlin, Senator für Bildung und Wissenschaft Bremen & Ministerium für Bildung, Wissenschaft und Kultur des Landes Mecklenburg Vorpommern, 2004); NI – Niedersachsen (Niedersächsisches Kultusministerium, 2006); NW – Nordrhein-Westfalen (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2009); RP – Rheinland-Pfalz (Ministerium für Bildung, Frauen und Jugend, 2002); SL – Saarland (Ministerium für Bildung, Familie, Frauen und Kultur Saarland, 2009); SN – Sachsen (Sächsisches Staatsministerium für Kultus, 2009); ST – Sachsen-Anhalt (Kultusministerium Sachsen-Anhalt, 2007); SH – Schleswig-Holstein (Ministerium für Bildung, Wissenschaft Forschung und Kultur des Landes Schleswig-Holstein, 2005); TH - Thüringen (Thüringer Ministerium für Bildung, Wissenschaft und Kultur, 2010); ✓ – CBM inhaltlich explizit passend zum jeweiligen Rahmenplan

Neben dieser allgemeinen curricularen Einordnung der CBM wird eine differenzierte Einbettung der mit den Verfahren verbundenen Anforderungen in theoretische Modelle und empirische Befunde vorgenommen. Auf diese Weise wird Aufschluss über die inhaltliche Validität der Verfahren gewonnen. Inwieweit wesentliche Konzepte, die mit der Entwicklung mathematischer Kompetenzen verknüpft sind, durch die vorhandenen Verfahren abgefragt werden, soll separat für die einzelnen CBM in den nachfolgenden Abschnitten analysiert werden.

4.2 CBM „Zahlenlesen im Zwanzigerraum“ (Niveaustufe 1)

Das CBM „Zahlenlesen im Zwanzigerraum“ wurde in Anlehnung an bei Hosp et al. (2007) beschriebene Verfahren des Typs number identification (siehe Abschnitt 3.2.4) konzipiert. Dabei sollen die Kinder innerhalb einer Bearbeitungszeit von einer Minute möglichst viele der Zahlen des Zahlenraumes bis 20 laut vorlesen. Jede Parallelform dieses CBM umfasst 64 Items. Dazu müssen sie die Zahlsymbole identifizieren und eine korrekte Zuordnung von entsprechenden Zahlworten zu den abgebildeten Ziffern vornehmen. Somit ist die Kenntnis der Zahlen eine Voraussetzung zur Lösung dieser Aufgaben. Dieses CBM ist ausschließlich in einer Einzelsituation mit dem Kind durchführbar. Damit nimmt das CBM „Zahlenlesen im Zwanzigerraum“ eine Sonderstellung in dem hier beschriebenen Set an CBM ein. Die Durchführung erscheint, im Vergleich zu den anderen in dieser Arbeit beschriebenen Verfahren, etwas aufwändiger, da deutlich mehr Zeit gebraucht wird, den gesamten Klassenverband zu testen und ein organisatorischer Rahmen geschaffen werden muss, der es der Lehrkraft erlaubt, Kinder einzeln zur Testung zu sich zu bestellen. Auch wenn die hier abgeprüfte Zahlenkenntnis eine wichtige Grundlage für weitere mathematische Aspekte darstellt (Clarke & Shinn, 2004; Gaupp et al.,

2004), wird das CBM „Zahlenlesen im Zwanzigerraum“, aufgrund des erhöhten Aufwandes bei der Durchführung, bei Lehrkräften vermutlich eher geringere Akzeptanz finden. Eine alternative Möglichkeit, die Fähigkeit zur Zahlenidentifikation im Klassenverband zu prüfen, wäre ein Zahlendiktat, bei dem der Testleiter Zahlen laut vorliest und die Kinder diese notieren lässt. Zwar bestünde dann das Problem der Zeitintensität aufgrund einer Einzeltestung nicht, bei einem solchen Vorgehen wird jedoch zwangsläufig nicht nur die Identifikation von Zahlsymbolen, sondern auch das Ausmaß, in welchem die Kinder in der Lage sind, Zahlen zu schreiben, erfasst. Deshalb fiel die Entscheidung bei der Konzeption des CBM „Zahlenlesen im Zwanzigerraum“ für die hier vorliegende Verfahrensweise, die tatsächlich die Fähigkeit des Lesens von Zahlen erfasst.

Gaupp et al. (2004) ordnen die Fähigkeit des Lesens von Zahlen, neben dem Schreiben und Ordnen von Zahlen, die Fähigkeit, sich Zahlen als Mengen vorzustellen, die Größe von Mengen abzuschätzen und sicher zählen zu können als numerische Basiskompetenzen ein, welche ein Kind für ein erfolgreiches Erlernen weiterführender mathematischer Kompetenzen benötigt. Auch Clarke und Shinn (2004) ordnen die Fähigkeit, Zahlen zu identifizieren als einen Bestandteil des number sense ein. Darüber hinaus zeigen die in Kapitel 2 erwähnten Studien von Schmidt (1982) und Grassmann et al. (2002), dass der Großteil (etwa 90 %) der Erstklässler zum Schulbeginn schon Zahlen im Zahlenraum bis 10 korrekt benennen kann. Festgestellte Probleme eines Kindes bei der Benennung vorgegebener Zahlen können somit als erster Hinweis auf grundlegende Schwierigkeiten im Bereich der Zahlenkenntnis und damit auch beim Erwerb mathematischer Kompetenzen angesehen werden, welche sich ohne anschließende spezifische Fördermaßnahmen manifestieren können (Fritz et al., 2007; Krajewski & Schneider, 2006). Von Aster et al. (2007) weisen ebenso

daraufhin, dass rechenschwache Kinder nur über wenig ausgeprägtes Wissen über die arabischen Zahlen verfügen. Anhand des CBM „Zahlenlesen im Zwanzigerraum“ soll es ermöglicht werden, diese Kinder frühzeitig zu identifizieren. In Abbildung 24 sind Beispielitems des CBM „Zahlenlesen im Zwanzigerraum“ dargestellt.

18	5	3	4	20	16	8	11

Abbildung 24: Beispielitems des CBM „Zahlenlesen im Zwanzigerraum“

Von dem CBM „Zahlenlesen im Zwanzigerraum“ liegen acht als parallel angenommene Versionen vor, welche in die Analysen innerhalb der vorliegenden Arbeit einbezogen werden können.

4.3 CBM „Zahlenreihe im Zwanzigerraum“ (Niveaustufe 1)

Das innerhalb der vorliegenden Arbeit konzipierte CBM „Zahlenreihe im Zwanzigerraum“ wurde in Anlehnung an das bei Hosp et al. (2007) beschriebene Verfahren des Typs missing number (siehe Abschnitt 3.2.4) entwickelt. Dabei sollen Kinder innerhalb einer Minute möglichst viele der Seriationsaufgaben im Zwanzigerraum lösen. Aus einem Tripel aufeinanderfolgender Zahlen des Zahlenraumes bis 20 ist dazu jeweils eine durch einen Unterstrich als Platzhalter ersetzt. Die Lösung der Items besteht somit in der Bestimmung von jeweils entweder des Vorgängers oder des Nachfolgers einer Zahl. Um eine möglichst breites Spektrum an Aufgabenschwierigkeiten zu erreichen, wurden die CBM dabei so zusammengestellt, dass es sowohl Items mit Zahlenreihen im Zahlenraum bis 10 (bspw.: 2, 3, __) als auch im Zahlenraum zwischen 10 und 20 (bspw.: __, 15, 16) und mit Zehnerüberschreitung (bspw.: 9, __, 11) vertreten sind. Insgesamt umfassen die Parallelformen des CBM jeweils 24 Items. Das

Ergänzen der Zahlenreihen ist – abhängig von den mathematischen Kompetenzen der Kinder – über verschiedene Wege zu lösen. Eine Möglichkeit zur Lösung dieser Aufgaben besteht in einer guten Kenntnis des Zahlenstrahls, d. h. die feste Reihenfolge der Zahlen ist bekannt und somit eine gute Orientierung im Zahlenraum bis Zwanzig möglich. Dabei werden die Zahlen geistig auf einem mentalen Zahlenstrahl, im Sinne von Resnick (1989, 1992), Dehaene (1992) oder von Aster (2005), organisiert, welcher ein schnelles und sicheres Zuordnen der fehlenden Zahlen ermöglicht. Durch gute Zählkompetenzen lässt sich dieses CBM, wenn auch langsamer, ebenfalls lösen. Aufgrund der zeitlich begrenzten Bearbeitungsdauer würde der Einsatz von Zählstrategien demnach zu geringeren Punktzahlen führen. Bei der Bestimmung des Vorgängers der vorgegebenen Zahlen setzt dies darüber hinaus voraus, dass das Kind auch rückwärts im Zahlenraum bis 20 zählen kann, was Kindern mit Rechenschwäche Schwierigkeiten bereiten kann (Aunola et al., 2004). Laut Clarke und Shinn (2004) ist das Aufgabenformat missing number geeignet, um Hinweise über das Ausmaß des Zahlenbewusstseins (number sense) eines Kindes zu erhalten. Somit erscheint das CBM „Zahlenreihe im Zwanzigerraum“ ein geeignetes Werkzeug zur Identifikation von Kindern mit einem gering ausgeprägten mathematischen Vorwissen zu sein. In Abbildung 25 sind Beispielitems des CBM „Zahlenreihe im Zwanzigerraum“ dargestellt.

__, 19, 20	8, 9, __	12, __, 14

Abbildung 25: Beispielitems des CBM „Zahlenreihe im Zwanzigerraum“

Von dem CBM „Zahlenreihe im Zwanzigerraum“ liegen acht als parallel angenommene Versionen vor, von welchen jedoch nur sieben

in die Analysen innerhalb der vorliegenden Arbeit einbezogen werden können.

4.4 CBM „Mengenvergleich im Zwanzigerraum“ (Niveaustufe 1)

Das CBM „Mengenvergleich im Zwanzigerraum“ wurde in Anlehnung an das bei Hosp et al. (2007) beschriebene Verfahren des Typs „quantity discrimination“ (siehe Abschnitt 3.2.4) entwickelt. Die Kinder sollen dabei innerhalb einer Minute von möglichst vielen der präsentierten Zahlenpaare im Zahlenraum bis 20 immer die jeweils größere Zahl angeben. Jede Parallelform des CBM umfasst 32 Items. Um eine Streuung von Aufgabenschwierigkeiten zu ermöglichen, wurde bei der Gestaltung der Items darauf geachtet, dass sowohl Vergleiche zweier Zahlen des Zahlenraumes bis 10 (bspw.: 7 und 8) bzw. des Zahlenraumes zwischen 10 und 20 (bspw.: 16 und 18) als auch Vergleiche zwischen den verschiedenen Zahlenräumen (bspw.: 5 und 14) vorgenommen werden müssen. Da hier die Kompetenz zum Vergleich zweier Zahlen und nicht die richtige Anwendung der Relationszeichen überprüft werden soll, werden die Kinder nicht aufgefordert, die jeweiligen Symbole <, = oder > zu setzen, sondern die jeweils größere der gegebenen Zahlen anzukreuzen. Abhängig vom Vorhandensein erster mathematischer Einsichten können die Kinder verschiedene Strategien zur Lösung der Aufgaben anwenden. Eine sichere Kenntnis der Zahlenreihe bis zwanzig ermöglicht dabei einen Vergleich anhand der Stellung in der Zahlenfolge. Dazu müssen die Kinder über eine mentale Repräsentation des Zahlenstrahls verfügen, wie sie Resnick (1989, 1992), Dehaene (1992) oder von Aster (2005) beschreiben. Dieses Lösungsvorgehen basiert auf einem ordinalen Verständnis der Zahlen. Zur Lösung der hier gegebenen Items ist jedoch auch ein Vergleich anhand der Mächtigkeit der dargestellten Zahlen denkbar. Ein solches Vorgehen bei der Lösungsfindung erfordert hingegen ein kardinales Verständnis

der Zahlen, d. h. die Kinder müssen dafür über ein Anzahlkonzept verfügen (Fritz et al., 2007; Krajewski & Schneider, 2006). Dabei muss zwischen einem zunächst noch unpräzisen Anzahlkonzept, das es ermöglicht, lediglich hinreichend verschieden große Zahlen (bspw.: 5 und 14) und einem präzisen Anzahlkonzept, das es ermöglicht, auch relativ nah beieinander liegende Zahlen (bspw.: 7 und 8 bzw. 16 und 18) hinsichtlich ihrer Mächtigkeit zu vergleichen (Krajewski & Schneider, 2006). Aufgaben des Formats Mengenvergleich, wie es im vorliegenden CBM der Fall ist, werden nach Clarke und Shinn (2004) als ein Aspekt des number sense aufgefasst. Insofern spielt das CBM „Mengenvergleich im Zwanzigerraum“ eine wichtige Rolle bei der Erfassung erster mathematischer Kompetenzen in der ersten Klasse der Grundschule. In Abbildung 26 sind Beispielitems des CBM „Mengenvergleich im Zwanzigerraum“ dargestellt.

16 ___ 8	20 ___ 5	0 ___ 16	9 ___ 15

Abbildung 26: Beispielitems des CBM „Mengenvergleich im Zwanzigerraum“

Von dem CBM „Mengenvergleich im Zwanzigerraum“ liegen acht als parallel angenommene Versionen vor, von welchen jedoch nur sieben in die Analysen innerhalb der vorliegenden Arbeit einbezogen werden können.

4.5 CBM „Arithmetische Basiskompetenzen“ (Niveaustufe 1 & 2)

Das CBM „Arithmetische Basiskompetenzen“ stellt eine Adaption eines Verfahrens des amerikanischen Internetportal www.easycbm.com dar. Wie in Abschnitt 3.2.4 beschrieben, liegen für diese CBM umfangreiche Nachweise hinsichtlich ihrer Testgüte

vor. Der mehrstufige Prozess der Itemgenerierung spricht für die inhaltliche Validität der Verfahren (Alonzo & Tindal, 2009). Anhand einer Pilotierungsstudie mittels einer Stichprobe von etwa 2800 Kindern des Kindergartens bis zur Klassenstufe 8 wurden die erarbeiteten Items auf Rasch-Konformität sowie ihre Schwierigkeiten und ihre Standardfehler untersucht. So entstanden zehn CBM mit jeweils 16 Items zu wesentlichen mathematischen Bereichen einer jeden Klassenstufe (z. B. Zahlen und Operationen, Geometrie, Größen und Messen etc.), deren Parallelität anhand ihrer mittleren Schwierigkeiten bereichs- sowie klassenstufenweise geprüft wurde (Alonzo & Tindal, 2009). Zudem konnten hohe Reliabilitätskoeffizienten von $r > .80$ (ermittelt als interne Konsistenz sowie Splithalf-Reliabilität) für die Verfahren der Klassenstufen 1 und 2 sowie für den Kindergartenbereich nachgewiesen werden. In weiteren Analysen konnten Nachweise für die Kriteriums-, die Konstrukt- sowie die prognostische Validität der CBM ermittelt werden (Anderson et al., 2010).

Die genannten Ergebnisse zusammenfassend kann festgehalten werden, dass die CBM der Internetplattform www.easyCBM.com eine vielversprechende sowie wissenschaftlich fundierte Möglichkeit der Lernverlaufsdiagnostik für den Mathematikunterricht in der ersten Klasse darstellen. Ein Einsatz in deutschen Grundschulen hingegen erweist sich als schwierig, da einige der Items sprachgebunden gestaltet sind. Daher wurde die hier vorgestellte Adaption, im Gegensatz zum amerikanischen Original, sprachfrei gestaltet, um den Einsatz in ersten Klassen deutscher Grundschulen zu ermöglichen. In diesem Zusammenhang wurden einige Aufgabenformate des Originals durch neue Aufgabenformate ersetzt. Dabei wurde darauf geachtet, dass gleiche Aufgabenformate durch gleiche Alternativen ersetzt werden, sodass die Struktur der Aufgabenblätter mit dem Original vergleichbar ist. Einige der verwendeten Mengendarstellungen- bzw. Operationsdarstellungen

sind vom Zwanzigerfeld des Lehrwerks „Das Zahlenbuch“ (Wittmann & Müller, 2004) inspiriert. Das CBM „Arithmetische Basiskompetenzen“ weicht in seinem Aufbau von den anderen in Kapitel 4 dargestellten CBM ab. Innerhalb der Bearbeitungszeit von einer Minute werden nicht allein die Fähigkeiten in einem isolierten Kompetenzbereich abgefragt, sondern verschiedene Aufgabenformate erfassen unterschiedliche mathematische Einsichten und Konzeptvorstellungen. Zur Lösung der Aufgaben ist immer die eine richtige der Antwortalternativen anzukreuzen (multiple choice), wobei neben dieser einen korrekten Lösung jeweils immer zwei Distraktoren abgebildet sind.

Abbildung 27 enthält eine Übersicht über die erfassten Kompetenzbereiche innerhalb der hier vorliegenden CBM-Adaption „Arithmetische Basiskompetenzen“, welche zusätzlich durch Beispielitems visualisiert werden. Es werden Kompetenzen zur Lösung von Seriationsaufgaben der Schüler erfasst. Dabei sollen fehlende Abschnitte von Muster- oder Zahlenfolgen ergänzt werden. Ähnlich wie im CBM „Zahlenreihe im Zwanzigerraum“ müssen zur Lösung derartiger Aufgabenformate Gesetzmäßigkeiten in bildlichen oder arithmetischen Mustern erkannt und fortgesetzt werden. Dies dient als Grundlage zum Zählprozess bzw. zum Verständnis der Zahlenreihe, welche eine Abfolge von Elementen, welche immer jeweils um eins größer werden, widerspiegelt. Weiterhin wird die Kompetenz zur Zuordnung von Mengen und Zahlen geprüft. Dabei sollen auf Grundlage verschiedener Mengendarstellungen abgebildete Anzahlen erfasst werden. Dies kann entweder durch Zählen, durch Simultanerfassung der Elemente oder einer Kombination beider Verfahren vollzogen werden. Die Kompetenz zur Menge-Zahl-Zuordnung stellt eine Grundvoraussetzung für das weitere mathematische Lernen dar, über welche vor allem viele rechenschwache Kinder nicht verfügen (Fritz et al., 2007; Krajewski & Schneider, 2006). Damit verbunden ist ein kardinales Verständnis

der Zahlen, also die Einsicht, dass Zahlen eine gewisse Menge bzw. Anzahl symbolisieren. Erst auf dieser Grundlage kann ein Verständnis für das Zerlegen von Zahlen und somit für erste Rechenoperationen entstehen. Laut den in Kapitel 2 geschilderten Studien verfügt ein relativ großer Teil der Schulanfänger über Kompetenzen zur Menge-Zahl-Zuordnung. Können solche Aufgaben jedoch nicht gelöst werden, ist dies ein guter Indikator für ein nur gering ausgeprägtes mathematisches Vorwissen. Darüber hinaus wird anhand von Zahlenhäusern die Kompetenz zur Lösung von Aufgaben zur Zahlzerlegung erfasst. In Zahlenhäusern entspricht die Summe der beiden Zahlen im Erdgeschoss des Zahlenhauses der Zahl im Dachgeschoss. Bei diesem Aufgabenformat soll jeweils aus einer gegebenen Auswahl jenes Zahlenpaar herausgesucht werden, das zur Zahl im Dachgeschoss passt. Das korrekte Zerlegen von Zahlen deutet auf ein gesichertes Verständnis des Teile-Ganzes-Konzepts hin, welches grundlegend für die wesentlichen mathematischen Rechenoperationen ist. Das Beherrschen des Teile-Ganzes-Konzept hat somit eine zentrale Bedeutung für den weiteren Lernerfolg im Mathematikunterricht (Fritz et al., 2007; Krajewski & Schneider, 2006). In einem weiteren Aufgabenformat wird die Kompetenz zur Zuordnung von einfachen symbolisch dargestellten Additions- bzw. Subtraktionsaufgaben zu entsprechenden ikonischen Darstellungen geprüft. Dabei müssen die Aufgaben innerhalb der verschiedenen Repräsentationsformen transferiert werden. Laut Aebli (1963, 1976) gilt das Übersetzungsvermögen zwischen ikonischen und symbolischen Darstellungsformen als Grundlage für ein umfassendes Operationsverständnis. Wie bereits erwähnt, sind einige der abgebildeten Operationen angelehnt an die Darstellung am Zwanzigerfeld des Lehrwerkes „Das Zahlenbuch“ (Müller & Wittmann, 2004). Letztlich wird die Qualität erster Rechenkompetenzen anhand einfacher Additionsaufgaben abgefragt. Dabei werden die Aufgaben bzw. die Ergebnisse in verschiedenen Repräsentationsebenen dargestellt. Es gibt Items, bei denen

- sowohl Aufgabe als auch Ergebnis ikonisch abgebildet sind,
- die Aufgabe ikonisch, jedoch das Ergebnis symbolisch abgebildet ist,
- die Aufgabe symbolisch, jedoch das Ergebnis ikonisch abgebildet ist,
- sowohl Aufgabe als auch Ergebnis symbolisch abgebildet sind.

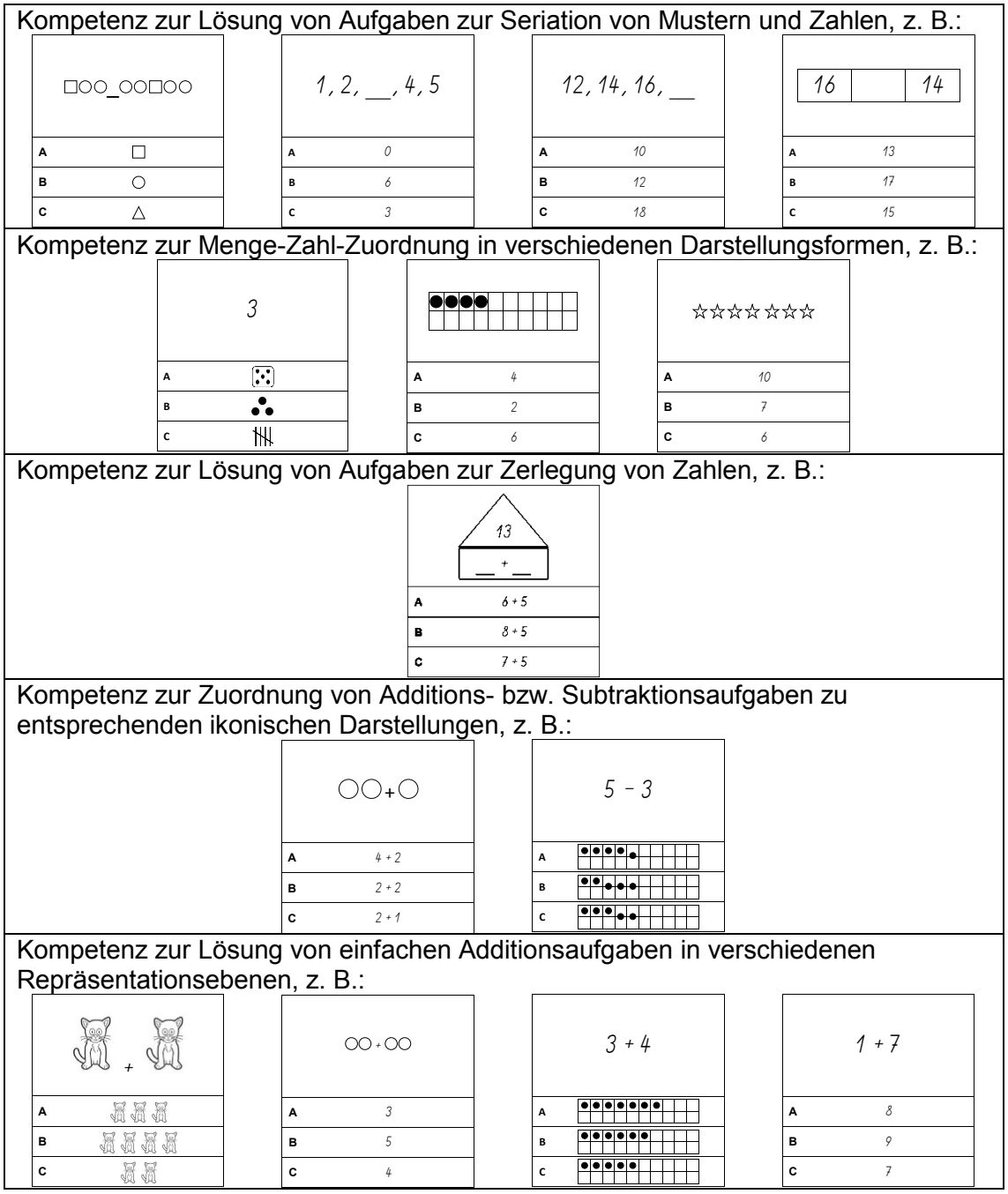


Abbildung 27: In den CBM „Arithmetische Basiskompetenzen“ abgefragte Kompetenzen verdeutlicht anhand von Beispielitems

Da das CBM „Arithmetische Basiskompetenzen“ über beide Schuljahreshälften der Klassenstufe 1 hinweg eingesetzt werden soll, liegen hiervon insgesamt zehn als parallel angenommene Versionen vor, welche in die Analysen innerhalb der vorliegenden Arbeit einbezogen werden.

4.6 CBM „Zahlzerlegung im Zwanzigerraum“ (Niveaustufe 2)

Wie schon in der Beschreibung des CBM „Arithmetische Basiskompetenzen“ erläutert, hat das Verständnis des Teile-Ganzes-Konzeptes eine zentrale Bedeutung für den weiteren Lernerfolg im Mathematikunterricht, da es grundlegend für die wesentlichen mathematischen Rechenoperationen Addition und Subtraktion ist (Fritz et al., 2007; Krajewski & Schneider, 2006). Mittels des CBM „Zahlzerlegung im Zwanzigerraum“ soll die Entwicklung dieses Verständnisses geprüft werden. Dazu sind die Kinder aufgefordert, ähnlich wie bei den Items zur Zahlzerlegung des CBM „Arithmetische Basiskompetenzen“, innerhalb einer Minute möglichst viele der vorgegebenen Zahlenhäuser im Zahlenraum bis 20 zu bearbeiten. Jede Parallelversion des CBM umfasst 20 Items. Beim CBM „Zahlzerlegung im Zwanzigerraum“ fehlt in jedem vorgegebenen Zahlenhaus jeweils eine Zahl im Erdgeschoss (Summand), welche ergänzt werden soll. Um ein möglichst breites Schwierigkeitsspektrum der Aufgaben zu erreichen, sind dabei die Zahlen im Dach- bzw. Erdgeschoss sowohl aus dem Zahlenraum bis 20 als auch aus dem Zahlenraum bis 10. In Abbildung 28 sind Beispielitems des CBM „Zahlzerlegung im Zwanzigerraum“ dargestellt.

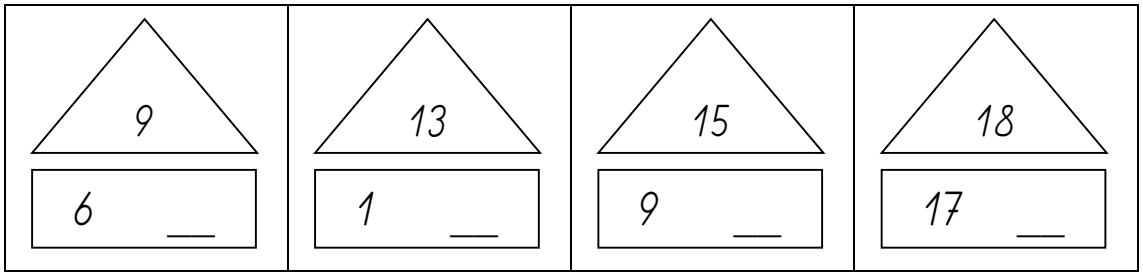


Abbildung 28: Beispielitems des CBM „Zahlzerlegung im Zwanzigerraum“

Von dem CBM „Zahlzerlegung im Zwanzigerraum“ liegen acht als parallel angenommene Versionen vor, welche in die Analysen innerhalb der vorliegenden Arbeit einbezogen werden sollen.

4.7 CBM „Addition im Zwanzigerraum“ (Niveaustufe 2)

Das CBM „Addition im Zwanzigerraum“ wurde in Anlehnung an die in der US-amerikanischen Literatur beschriebenen CBM des Typs computation (siehe Abschnitt 3.2.4) erarbeitet. Die Kinder sollen dabei innerhalb einer Minute möglichst viele der präsentierten Additionsaufgaben im Zahlenraum bis 20 lösen. Jede Parallelversion des CBM umfasst 40 Items. Um einerseits den Entwicklungsverlauf aller Erstklässler, andererseits jedoch auch speziell den von Kindern mit Schwierigkeiten beim Rechnen zu erfassen, müssen die CBM jeweils Items verschiedener Schwierigkeitsgrade enthalten. Ergebnisse der Untersuchungen von Rinkens und Eilerts (2001) geben Informationen über die Schwierigkeitsgrade von Rechenaufgaben. Diese wurden bei der Auswahl der Items berücksichtigt. Die Autoren überprüften im Jahr 2001 die Lösungshäufigkeiten von Rechenaufgaben im Zahlenraum bis 20. Dazu erhielten 284 Erstklässler zum Ende des ersten Schulhalbjahres aus einem Aufgabenpool von 231 Additionsaufgaben im Zahlenraum bis 20 randomisierte Aufgabenblätter, welche durch eine zuvor festgelegte Clusterung der Aufgaben als strukturgleich bzw. -ähnlich angenommen wurden. In Abbildung 29 sind alle gestellten Rechenaufgaben sowie die entsprechenden prozentualen Anteile der Kinder, die die jeweilige Aufgabe richtig gelöst haben, farblich codiert aufgeführt.

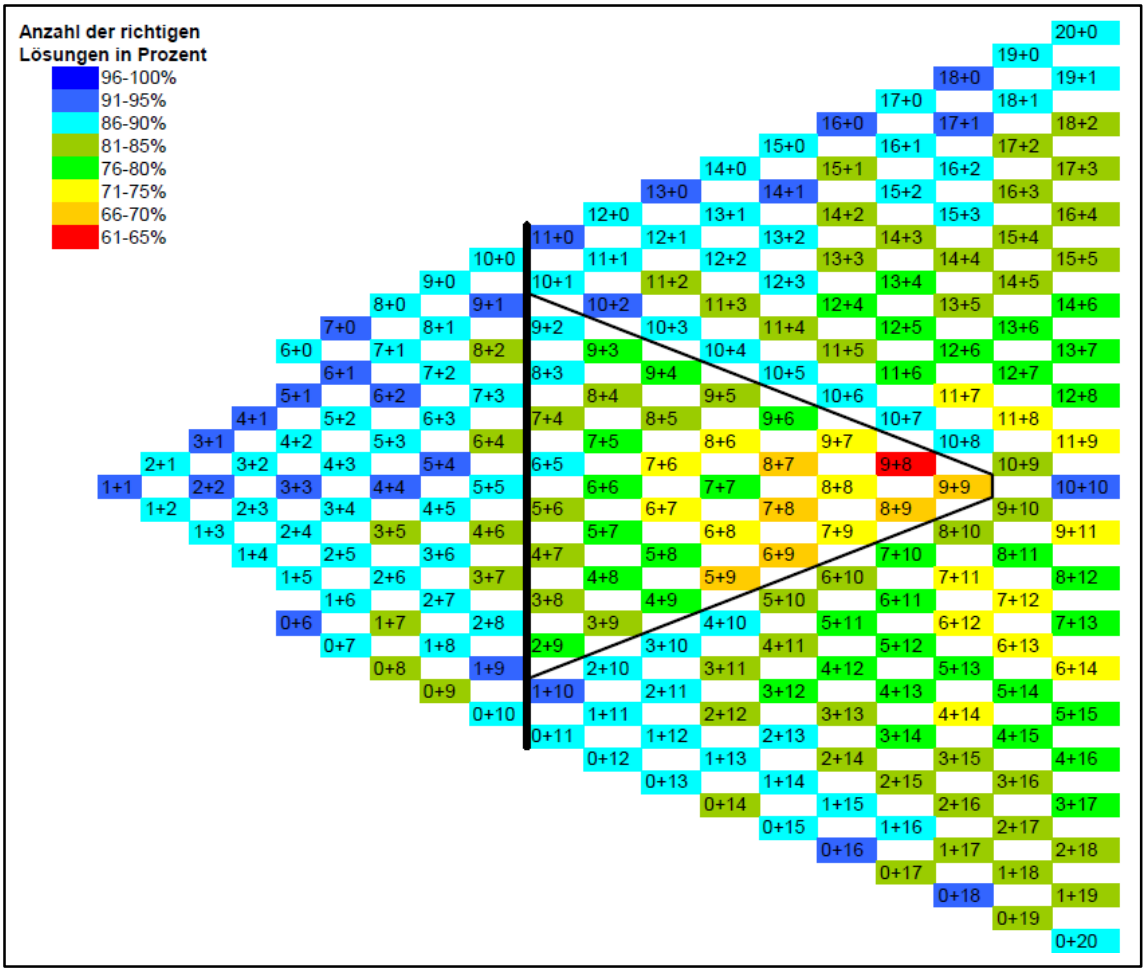


Abbildung 29: Prozentuale Anteile richtiger Lösungen von Additionsaufgaben im Zahlenraum bis 20, bei einer Anzahl an Probanden zwischen 146 und 284 Erstklässlern (entnommen aus: Rinkens & Eilerts, 2001, S. 7)

Die Ergebnisse weisen darauf hin, dass mit zunehmender Größe der Summanden die Lösungshäufigkeiten der Aufgaben abnehmen. Es zeigt sich, dass die 10 dabei eine besondere Rolle spielt. Einerseits stellt sie einen gewissen Schwellenwert dar, denn „Aufgaben im Zahlenraum bis Zehn werden häufiger richtig gelöst als die Aufgaben im Zahlenraum von 10 bis 20“, so Rinkens und Eilerts (2001, S. 10). Andererseits scheinen viele Kinder die besondere Stellung der 10 im Dezimalsystem nutzen zu können, denn „Aufgaben mit Zehn als Summand oder Ergebnis werden zu einem hohen Prozentsatz richtig gelöst (Rinkens & Eilerts, 2001, S. 10). Ebenso werden Aufgaben, bei denen einer der Summanden 0 oder 1 ist, von vielen Kindern richtig gelöst. Im Zahlenraum bis 10 fallen zudem Verdopplungsaufgaben

durch hohe Lösungshäufigkeiten auf. Die Verdopplungsaufgaben im Zahlenraum bis 20 werden hingegen nur von 67 % bis 89 % der Kinder richtig gelöst. Besonders schwer fallen den Kindern dabei die Aufgaben 7 + 7 (ca. 72 %), 8 + 8 (ca. 69 %) und 9 + 9 (ca. 67 %).

Rinkens und Eilerts (2001) bezogen in ihren Untersuchungen ausschließlich Additionsaufgaben der Form a + b = __ ein, bei denen zur Lösung jeweils die Summe der beiden Summanden ermittelt werden muss. Dazu kann ein Schüler, abhängig von der Ausprägung seiner mathematischen Kompetenzen, verschiedene Lösungswege nutzen (Siegler, 1987). In Abschnitt 2.7 wurden die Ergebnisse der Studie von Siegler (1987) erläutert, in welcher die Strategien von Kindern beim Lösen von Rechenaufgaben geschildert werden. In ihrer Untersuchung zeigte sich, dass viele Kinder zunächst ressourcenaufwändige Zählstrategien (counting-all-strategy, counting-on-strategy, min-strategy) nutzen, jedoch im Laufe der Zeit eine große Wissensbasis anlegen können, welche einen schnellen Faktenabruf zur Lösung einer gestellten Aufgabe zulässt. Speziell rechenschwache Kinder verfügen jedoch oft nicht über ein breit ausgeprägtes basales Faktenwissen (Bull & Johnston, 1997; Geary, 1990; Geary et al., 1991; Geary et al., 1992; Grube 2006; Ostad, 1998; Temple & Sherwood, 2002), sodass sie meist auf zählbasierte Strategien zurückgreifen, welche mehr Zeit in Anspruch nehmen und darüber hinaus auch recht fehleranfällig sind (Kaufmann & Wessolowski, 2005, 2006; Ostad, 1997; Schipper, 2003; Wehrmann, 2007). Es ist demnach davon auszugehen, dass Kinder mit Schwierigkeiten im Rechnen durch eine nur geringe Anzahl an richtig gelösten Aufgaben in dem CBM „Addition im Zwanzigerraum“ auffällig werden.

Neben Additionsaufgaben der Form a + b = __ fragen verschiedene Items des hier vorliegenden CBM „Addition im Zwanzigerraum“ darüber hinaus auch die Fähigkeit zum Lösen sog.

Platzhalteraufgaben ab. Bei diesen Aufgaben ist jeweils die Summe sowie einer der Summanden, in der Form a + __ = c bzw. __ + b = c, vorgegeben. Solche Platzhalteraufgaben prüfen ein grundlegendes Verständnis der Kinder für die Zerlegbarkeit und die Relationen von Zahlen, da die Lösung der Aufgabe nicht ohne weiteres direkt abzählbar ist. Ebenso ist ein umfassendes Verständnis der Rechenoperationen Addition und Subtraktion nötig, da zur Lösung der Aufgaben die jeweilige Umkehroperation gebildet werden muss. Jiban und Deno (2007) weisen zudem darauf hin, dass beim Lösen solcher Platzhalteraufgaben das Gleichheitszeichen von den Schülern nicht lediglich als Aufforderung zum Rechnen aufgefasst werden kann, sondern, dass es dabei eine Balance - im Sinne von „ist gleich“ oder „ist genauso viel wie“ - der beiden Seiten angibt. Neben dem einfachen Abruf von basalem arithmetischen Faktenwissen, werden demnach komplexe Einsichten wie das Teile-Ganzes-Verständnis, der relationale Zahlbegriff sowie das Verständnis für die Operationen Addition und Subtraktion getestet, welche grundlegend für den Erwerb weiterführenden mathematischen Wissens sind (Fritz, et al., 2007; Krajewski & Schneider, 2006). Wie in Abschnitt 2.10 beschrieben, verfügen rechenschwache Kinder lediglich über basale mathematische Kompetenzen und Einsichten, sodass diese Kinder bei der Lösung der Platzhalteraufgaben Schwierigkeiten haben werden. Somit erweisen sich Platzhalteraufgaben als ein sinnvolles Aufgabenformat, um Kinder mit Schwierigkeiten zu identifizieren.

Auf Grundlage oben genannter Vorüberlegungen wurden Items für das CBM „Addition im Zwanzigerraum“ konzipiert. In einem ersten Schritt sind dabei alle Items auf Grundlage ihrer Struktur verschiedenen Kategorien zugeordnet worden, wobei davon auszugehen ist, dass Items innerhalb einer Kategorie einen ähnlichen Schwierigkeitsgrad aufweisen. In Tabelle 9 ist differenziert aufgeschlüsselt, welche Aufgabenstrukturen in dem CBM „Addition

im Zwanzigerraum“ enthalten sind. In einem zweiten Schritt wurden zwei CBM-Versionen gebildet, bei denen darauf geachtet wurde, dass jeweils an der gleichen Position der Aufgabenblätter Items derselben Kategorie verwendet wurden. Es ist davon auszugehen, dass die beiden auf diese Weise entstandenen Tests von gleicher bzw. ähnlicher Schwierigkeit sind und somit als parallel angenommen werden können. Von diesen beiden Versionen wurden in einem dritten Schritt, durch spaltenweises Umordnen, jeweils vier unterschiedliche Pseudoparallelformen erstellt, von denen wiederum angenommen werden kann, dass die Schwierigkeitsverteilung der Aufgaben zwischen den Verfahren homogen ausfällt, sodass es sich bei ihnen vermutlich um parallele Tests handelt.

Tabelle 9: Im CBM „Addition im Zwanzigerraum“ enthaltene Aufgabenstrukturen

	a + b = x	a + x = b	x + a = b
a, b, x ≤ 10	✓	✓	✓
a, b ≤ 10; 10 < x ≤ 20	✓		
a, b ≤ 10; x ≤ 20	✓		
b ≤ 10; 10 < a, x ≤ 20	✓		
a ≤ 10; 10 < b, x ≤20	✓	✓	✓
a, x ≤ 10; 10 < b ≤20		✓	✓
x ≤ 10; 10 < a, b ≤20			✓
a, b, x ≤ 10; a = b	✓		
a, b ≤ 10; 10 < x ≤ 20, a = b	✓		
a, x ≤ 10; b = 0	✓		
10 < a, x ≤ 20; b = 0	✓		

Erläuterung: ✓ - bedeutet, dass der Aufgabentyp im CBM „Addition im Zwanzigerraum“ vorkommt

Insgesamt stehen für das CBM „Addition im Zwanzigerraum“ acht verschiedene Versionen zur Verfügung. In Abbildung 30 sind Beispielitems des CBM „Addition im Zwanzigerraum“ dargestellt.

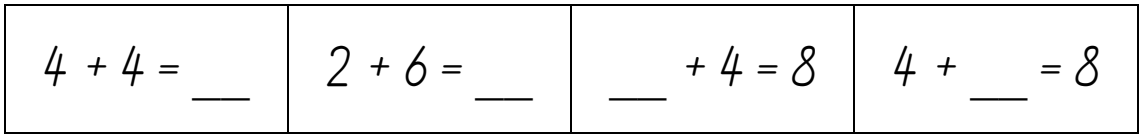


Abbildung 30: Beispielitems des CBM „Addition im Zwanzigerraum“

4.8 CBM „Subtraktion im Zwanzigerraum“ (Niveaustufe 2)

Die Konzeption des CBM „Subtraktion im Zwanzigerraum“ erfolgte analog zum CBM „Addition im Zwanzigerraum“. Das Verfahren wurde ebenso in Anlehnung an die in Abschnitt 3.2.4 beschriebenen CBM des Typs computation entwickelt. Innerhalb einer Bearbeitungszeit von einer Minute sollen dabei möglichst viele der präsentierten Aufgaben zur Subtraktion im Zahlenraum bis 20 gelöst werden. Jede Parallelform des CBM beinhaltet 40 Items. Auch in diesem Fall sollen die Entwicklungsverläufe aller Kinder, speziell jedoch auch die von Kindern mit Schwierigkeiten beim Rechnen erfasst werden. Daher müssen die CBM jeweils aus Items eines breiten Schwierigkeitsspektrums bestehen. Auch zur Konzeption der CBM zur Subtraktion liefern die Ergebnisse der Studie von Rinkens und Eilerts (2001) wichtige Informationen hinsichtlich der Schwierigkeitsgrade von Subtraktionsaufgaben. Innerhalb ihrer Studie im Jahr 2001 überprüften die Autoren neben den Lösungshäufigkeiten von Additionsaufgaben im Zahlenraum bis 20 auch die Lösungshäufigkeiten von Subtraktionsaufgaben im selben Zahlenraum. Dabei erhielten 287 Erstklässler zum Ende des ersten Schulhalbjahres aus einem Aufgabenpool von 231 Subtraktionssaufgaben im Zahlenraum bis 20 randomisiert zusammengestellte Aufgabenblätter mit je 21 Items, welche durch eine zuvor festgelegte Clusterung der Aufgaben als strukturgleich bzw. -ähnlich gestalten. Abbildung 31 gibt einen Überblick über alle gestellten Aufgaben sowie die entsprechenden prozentualen Anteile der Kinder, die die jeweilige Aufgabe richtig gelöst haben.

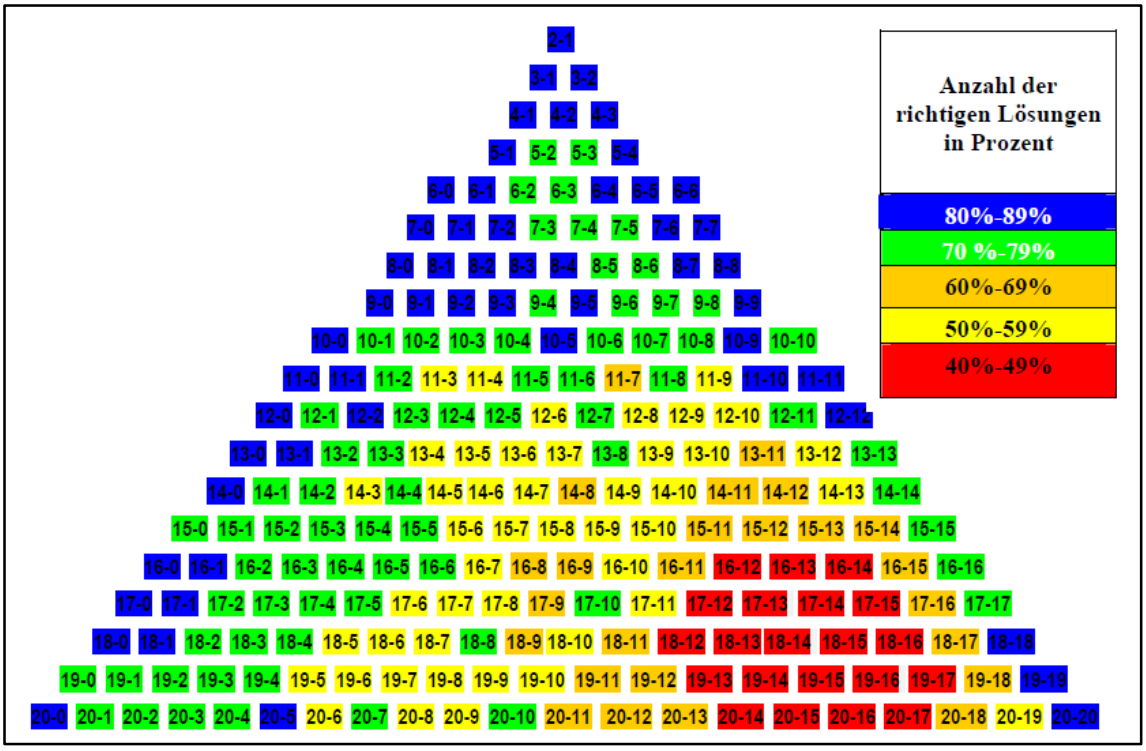


Abbildung 31: Prozentuale Anteile richtiger Lösungen von Subtraktionsaufgaben im Zahlenraum bis 20, bei einer Anzahl an Probanden zwischen 141 und 287 Erstklässlern (entnommen aus: Rinkens & Eilerts, 2001, S. 13)

Ähnlich wie bei den Additionstests zeigt sich, dass den Kindern Aufgaben mit Zahlen von zunehmender Größe immer schwerer fallen. Während Aufgaben im Zahlenraum bis zehn noch von zumindest 70 % der Kinder richtig gelöst werden konnten, sind die Lösungshäufigkeiten im Zahlenraum bis zwanzig geringer. Die Ergebnisse von Rinkens und Eilerts (2001) deuten darüber hinaus darauf hin, dass Subtraktionsaufgaben mit Überschreiten der Zehn eine besondere Schwierigkeit für Kinder darstellen (Rinkens & Eilerts, 2001).

> *„So wurden zum Beispiel Aufgaben mit der Zehn als Minuend durchgehend schlechter gelöst als Aufgaben mit der Neun oder der Acht als Minuend. Aufgaben mit einem großen Minuend und einem großen Subtrahend (beide über Zehn) bereiteten den Kindern die größten Schwierigkeiten“* (Rinkens & Eilerts, 2001, S. 13).

Zwischen 50 % und 80 % der Kinder konnten Subtraktionsaufgaben mit Nachbaraufgaben im Zahlenraum zwischen zehn und zwanzig (11 – 10, 12 – 11, 13 – 12 etc.) korrekt lösen. Deutlich geringer ist der Prozentsatz, wenn der Subtrahend mehr als eins vom Minuend verschieden ist. Die Aufgabe 17 – 14 = __ konnte nur von knapp 40 % der Kinder richtig gelöst werden (S. 16). Dies kann als Indiz dafür angesehen werden, dass die Kinder nicht erkennen, dass sie die Aufgabe leicht durch Ergänzen lösen können, sondern alle Subtraktionsaufgaben „stur“ durch Abziehen des Subtrahenden lösen. Es ist demnach davon auszugehen, dass auch Platzhalteraufgaben hinsichtlich der Subtraktion in der Form 15 – __ = 12 seltener von den Kindern richtig gelöst werden, also schwerer sind, als Aufgaben mit einer geringeren Differenz wie 16 - __ = 15.

In der Studie von Rinkens und Eilerts (2001) zur Lösungshäufigkeit von Subtraktionsaufgaben wurden ausschließlich Rechnungen der Form a – b = __ abgefragt. Um die gesuchte Differenz zu erhalten, mussten die Kinder dabei den Subtrahend vom Minuend anziehen. Analog zum Addieren können auch beim Subtrahieren abhängig von verfügbaren mathematischen Kompetenzen und Einsichten, verschiedene Lösungswege genutzt werden. In Frage kommen zählbasierte Strategien, die Ergänzungsstrategie und der Faktenabruf aus der Wissensbasis (Siegler, 1987). Auch hinsichtlich der Subtraktion fallen rechenschwache Kinder oft durch ein nur in geringem Maße ausgeprägtes Faktenwissen auf (Bull & Johnston, 1997; Geary, 1990; Geary et al., 1991; Geary et al., 1992; Grube 2006; Ostad, 1998; Temple & Sherwood, 2002). Sie müssen vermutlich meist auf ressourcenaufwändige, fehleranfällige zählbasierte Strategien zurückgreifen (Kaufmann & Wessolowski, 2005, 2006; Ostad, 1997; Schipper, 2003; Wehrmann, 2007). Kinder mit Schwierigkeiten im Rechnen müssten demnach auch im CBM „Subtraktion im Zwanzigerraum“ durch geringe Anzahlen an richtig

gelösten Aufgaben innerhalb der Bearbeitungsdauer von einer Minute auffällig werden.

Analog zum CBM „Addition im Zwanzigerraum“ fragen auch in dem hier vorliegenden CBM zur Subtraktion verschiedene Items die Fähigkeit zum Lösen von Platzhalteraufgaben ab. Dabei ist jeweils die Differenz sowie der Minuend, in der Form a – __ = c bzw. die Differenz und der Subtrahend, in der Form __ – b = c, vorgegeben. Die gestellten Aufgaben sind dadurch nicht ohne weiteres direkt abzählbar. Wie beim CBM „Addition im Zwanzigerraum“ beschrieben, wird damit das Vorhandensein eines Verständnisses der Kinder für die Zerlegbarkeit und die Relationen von Zahlen sowie der Rechenoperationen Addition und Subtraktion und deren Beziehung als Umkehroperationen abgeprüft, welches grundlegend für den Erwerb von weiterführendem mathematischen Wissen ist (Fritz, et al., 2007; Krajewski & Schneider, 2006). Den Ausführungen in Abschnitt 2.10 folgend werden rechenschwache Kinder speziell bei der Lösung dieser Platzhalteraufgaben Schwierigkeiten haben, da sie lediglich über basale mathematische Kompetenzen und Einsichten verfügen. Auch hinsichtlich der Subtraktion erscheinen Platzhalteraufgaben als ein sinnvolles Aufgabenformat, um Kinder mit Schwierigkeiten zu identifizieren.

Ausgehend von den hier dargestellten Überlegungen wurden die Items für das CBM „Subtraktion im Zwanzigerraum“ zusammengestellt. Dabei wurde erneut in mehreren Schritten vorgegangen. Zuerst wurden alle Items zur Subtraktion auf Grundlage ihrer Struktur verschiedenen Kategorien zugeordnet. Diese Kategorien sind in Tabelle 10 differenziert aufgeschlüsselt. Es wird dabei angenommen, dass Items innerhalb einer Kategorie einen ähnlichen Schwierigkeitsgrad aufweisen. Dann wurden zwei CBM-Versionen mit Aufgaben aus allen genannten Kategorien gebildet. Dabei wurde darauf geachtet, dass jeweils an der gleichen Position

der Aufgabenblätter Items derselben Kategorie verwendet wurden, um Tests von gleicher bzw. ähnlicher Struktur zu erhalten. Aufgrund der Vorgehensweise der Itemzusammenstellung wird ein gleicher Schwierigkeitsgrad der entstandenen Tests angenommen. Zuletzt wurden von diesen beiden Versionen durch spaltenweises Umordnen jeweils vier unterschiedliche Pseudoparallelformen erstellt, von denen wiederum angenommen werden kann, dass die Schwierigkeitsverteilung der Aufgaben zwischen den Verfahren homogen ausfällt, sodass es sich auch bei diesen Verfahren um Paralleltests handelt.

Tabelle 10: Im CBM „Subtraktion im Zwanzigerraum“ enthaltene Aufgabenstrukturen

	a – b = x	a – x = b	x – a = b
a, b, x ≤ 10	✓		✓
a, b ≤ 10; 10 < x ≤ 20			✓
b, x ≤ 10; 10 < a ≤ 20	✓	✓	
x ≤ 10; 10 < a, b ≤ 20	✓	✓	
b ≤ 10; 10 < a, x ≤ 20	✓	✓	✓
a, x ≤ 10; b = 0		✓	
10 < a, x ≤ 20; b = 0		✓	

Erläuterung: ✓ – bedeutet, dass der Aufgabentyp im CBM „Addition im Zwanzigerraum“ vorkommt

Insgesamt stehen somit für das CBM „Subtraktion im Zwanzigerraum“ acht verschiedene Versionen zur Verfügung. In Abbildung 32 sind Beispielitems des CBM „Subtraktion im Zwanzigerraum“ dargestellt.

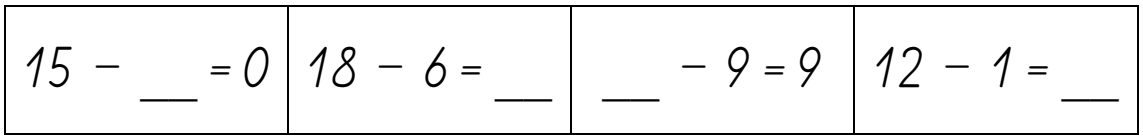


Abbildung 32: Beispielitems des CBM „Subtraktion im Zwanzigerraum“

5 Forschungsfragen und Hypothesen der Untersuchung

Mit der vorliegenden Untersuchung wird geprüft, inwieweit sich an US-amerikanischen Verfahren orientierte curriculumbasierte Messverfahren im Bereich des mathematischen Erstunterrichts auch im deutschen Schulsystem einsetzen lassen. D. h, es wird geprüft, ob die Verfahren eine ausreichende diagnostische Güte aufweisen und ob sie für den Zweck der Früherkennung von Rechenschwäche bzw. der Prävention genutzt werden können. Nachfolgend werden zentrale Forschungsfragen sowie Forschungshypothesen formuliert, die auf den Darstellungen der vorangehenden Kapitel beruhen. Das von L.S. Fuchs (2004) vorgeschlagene dreistufige Vorgehen bei der Evaluation von CBM (siehe dazu Abschnitt 3.3) bildet dabei die zentrale Grundlage. Entsprechend L.S. Fuchs‘ Postulat sind Untersuchungen auf jeder der drei Stufen notwendig, um einen validen und effektiven Einsatz von CBM zu gewährleisten. Alle dargestellten Forschungsfragen und -hypothesen sind auf jedes der in Kapitel 4 dargestellten CBM-Verfahren zu beziehen.

5.1 Überlegungen zu Forschungsfragen der Stufe 1 (L.S. Fuchs, 2004)

Forschungsarbeiten der von L.S. Fuchs (2004) vorgeschlagenen Stufe 1 beziehen sich dabei auf die Eignung der Verfahren als statische Indikatoren für Risiken bei der Entwicklung mathematischer Kompetenzen, d. h. ohne jegliche Beachtung der Entwicklung über die Zeit, sondern als punktuelle Messung im Sinne einer Statusdiagnostik. Damit sind technische Aspekte der Tests angesprochen. Folglich beziehen sich Forschungsfragen der Stufe 1 auf die in der Diagnostik üblichen Testgütekriterien, wie Objektivität, Reliabilität sowie Validität (Moosbrugger & Kevala, 2012). Die dazu in

Abschnitt 3.3.2 geschilderten theoretischen und empirischen Erkenntnisse führen zu den folgenden zentralen Forschungsfragen:

F_1 Handelt es sich bei den hier vorgestellten CBM um objektive Verfahren?

F_2 Handelt es sich bei den hier vorgestellten CBM um reliable Verfahren?

F_3 Handelt es sich bei den hier vorgestellten CBM um valide Verfahren?

Ganz wesentlich erscheint zudem die Voraussetzung, dass regelmäßige Messungen mit Tests gleicher Schwierigkeit erfolgen. Man spricht in diesem Fall von der Parallelität der Verfahren (Deno, 2003b; L.S. Fuchs, 2004; Klauer, 2010). Wäre diese Voraussetzung nicht gegeben, könnten sich ergebende Verlaufskurven nicht auf die tatsächliche Leistungsentwicklung eines Kindes, sondern auf unterschiedliche Schwierigkeitsgrade der eingesetzten Verfahren zurückführen lassen. Ein weiteres wichtiges Kriterium bei der Konzeption eines Verfahrens im Sinne der Lernverlaufsdiagnostik stellt die Kontent- bzw. Inhaltsvalidität dar (Klauer, 2010). D. h., die Items müssen in einem engen Zusammenhang mit einem bestimmten Kriterium stehen, in diesem Fall den mathematischen Kompetenzen eines Kindes der ersten Grundschulklasse. Die Items müssen diese Kompetenzen also operationalisieren. Da CBM-Ergebnisse im Rahmen schulischer Prävention einen Screeningcharakter annehmen, ist ebenso die Bestimmung der prognostischen Validität ein wesentlicher Aspekt der Analysen. Die Forschungsfragen beziehen sich dabei auf jedes der oben beschriebenen CBM, ebenso müssen die nachfolgenden Hypothesen bezüglich jedes dieser Verfahren geprüft werden.

Hypothesen hinsichtlich der Objektivität der Verfahren

$H_{1|1}$ Die Durchführungsobjektivität der CBM ist gewährleistet.

$H_{1|2}$ Die Auswertungsobjektivität der CBM ist gewährleistet.

$H_{1|3}$ Die Interpretationsobjektivität der CBM ist gewährleistet.

Hypothesen hinsichtlich der Reliabilität der Verfahren

$H_{2|1}$ Die Koeffizienten hinsichtlich der Splithalf-Reliabilität liegen im hohen Bereich ($r > .7$).

$H_{2|2}$ Die in Abständen von mehreren Schulwochen erhobenen Koeffizienten hinsichtlich der Retest-Reliabilität liegen zumindest im mittleren Bereich ($r > .5$).

$H_{2|3}$ Korrelationen der Ergebnisse zweier als parallel angenommener CBM, welche zeitgleich erhoben werden, liegen im hohen Bereich ($r > .7$).

Hypothesen hinsichtlich der Validität der Verfahren

$H_{3|1}$ Es bestehen zumindest mittlere Zusammenhänge ($r > .7$) zwischen den einzelnen CBM der Niveaustufen 1 und 2 (Konstruktvalidität).

$H_{3|2}$ Es bestehen mittlere Zusammenhänge ($.5 < r \leq .7$) zwischen den einzelnen CBM der Niveaustufen 1 und den mathematischen Schülerleistungen gemessen mit dem Diagnoseprogramm Kalkulie Teil 1 (Fritz et al., 2007) zu Beginn der Klasse eins (konvergente bzw. Übereinstimmungsvalidität).

$H_{3|3}$ Es bestehen mittlere Zusammenhänge ($.5 < r \leq .7$) zwischen den einzelnen CBM der Niveaustufe 2 und den mathematischen Schülerleistungen gemessen mit dem DEMAT 1+ (Krajewski et al., 2002) zum Ende der Klasse eins (konvergente bzw. Übereinstimmungsvalidität).

$H_{3|4}$ Es bestehen mittlere Zusammenhänge ($.5 < r \leq .7$) zwischen den einzelnen CBM der Niveaustufen 1 und 2 und den mathematischen Schülerleistungen gemessen mit dem Inventar Rechenfische (Wagner & Hartke, 2006) zu

verschiedenen Messzeitpunkten innerhalb der Klasse eins (konvergente bzw. Übereinstimmungsvalidität).

$H_{3|5}$ Es bestehen lediglich geringe Zusammenhänge ($r < .5$) zwischen den einzelnen CBM der Niveaustufe 1 und den Voraussetzungen zum erfolgreichen Leseerwerb der Schüler gemessen mit dem MÜSC (Mannhaupt, 2005) zu Beginn der Klasse eins (diskriminante Validität).

$H_{3|6}$ Es bestehen lediglich geringe Zusammenhänge ($r < .5$) zwischen den einzelnen CBM der Niveaustufen 1 und 2 und den Leseleistungen der Schüler gemessen mit der WLLP (Küspert & Schneider, 1998) zum Ende der Klasse eins (diskriminante Validität).

$H_{3|7}$ Die über die Messzeitpunkte 1 bis 4 erhobenen Schülerergebnisse innerhalb der CBM der Niveaustufe 1 erlauben eine valide Prognose der Mathematikleistungen zum Ende der Klasse eins erhoben mit dem DEMAT 1+ (Krajewski et al., 2002) (prognostische Validität).

$H_{3|8}$ Die über die Messzeitpunkte 5 bis 9 erhobenen Schülerergebnisse innerhalb der CBM der Niveaustufe 2 erlauben eine valide Prognose der Mathematikleistungen zum Ende der Klasse eins erhoben mit dem DEMAT 1+ (Krajewski et al., 2002) (prognostische Validität).

$H_{3|9}$ Die erhobenen Schülerergebnisse aller CBM der Niveaustufe 1 zum Messzeitpunkt 1 geben kombiniert betrachtet eine valide Prognose der Mathematikleistungen zum Ende der Klasse eins bzw. zwei erhoben mit dem DEMAT 1+ (Krajewski et al., 2002) bzw. mit dem DEMAT 2+ (Krajewski et al., 2004) (inkrementelle Validität).

$H_{3|10}$ Die erhobenen Schülerergebnisse aller CBM der Niveaustufe 2 zum Messzeitpunkt 5 geben kombiniert betrachtet eine valide Prognose der Mathematikleistungen zum Ende der Klasse eins bzw. zwei erhoben mit dem DEMAT 1+ (Krajewski

et al., 2002) bzw. mit dem DEMAT 2+ (Krajewski et al., 2004) (inkrementelle Validität).

5.2 Überlegungen zu Forschungsfragen der Stufe 2 (L.S. Fuchs, 2004)

Da durch curriculumbasierte Messverfahren der Lernverlauf von Kindern erfasst werden soll, um daraus Schlüsse hinsichtlich des Unterrichts bzw. der Förderung zu ziehen, reichen Analysen hinsichtlich der beschrieben Stufe 1 nicht aus. Wesentlich sind Erkenntnisse im Zusammenhang mit der folgenden Frage:

F_4 Können die Lernverläufe der Schüler durch die hier konzipierten CBM reliabel und sensibel erfasst werden?

Für die Beantwortung dieser zentralen Forschungsfrage ist es erforderlich, weitere Analysen hinsichtlich der von L.S. Fuchs (2004) beschriebenen Stufe 2 vorzunehmen. Dazu muss gezeigt werden, dass mittels CBM gemessene Verläufe auch die tatsächliche Leistungsentwicklung der Kinder abbilden. Geht es darüber hinaus darum, Leistungsverläufe speziell von schwächeren Schülern zu verfolgen, so müssen die eingesetzten CBM sehr sensibel gegenüber Leistungsänderungen sein, also auch minimale Veränderungen erfassen. Es erscheint wenig sinnvoll, ein CBM einzusetzen, in welchem Kinder einen durchschnittlichen Leistungszuwachs von einem halben Punkt je Monat aufweisen. Gerade bei leistungsschwachen Kindern würden sich Fortschritte dann erst nach mehreren Monaten zeigen. Nachfolgende Hypothesen beziehen sich im Anschluss an Klauer (2011) und Shin et al. (2004) auf Aspekte hinsichtlich der Änderungssensibilität der CBM (vgl. Abschnitt 3.3.3).

Hypothesen hinsichtlich der Änderungssensibilität der Verfahren

$H_{4|1}$ Die Korrelationen zwischen den CBM-Ergebnissen werden kleiner, je größer der Zeitraum zwischen den Erhebungszeitpunkten ist.

$H_{4|2}$ Die Splithalf-Reliabilität der CBM ist höher als die Retest- bzw. Paralleltest-Reliabilität.

$H_{4|3}$ Die Varianzen der Schülerergebnisse in den CBM erhöhen sich von Messzeitpunkt zu Messzeitpunkt.

$H_{4|4}$ Zwischen den Mittelwerten der Schülerergebnisse in den CBM zu aufeinander folgenden Messzeitpunkten gibt es signifikante Unterschiede.

$H_{4|5}$ In den einzelnen CBM zeigen die untersuchten Schüler einen durchschnittlichen monatlichen Leistungsanstieg von zumindest einem Punkt.

$H_{4|6}$ Die durch die CBM erhobenen Daten weisen auf verschiedene Leistungsentwicklungen zwischen Kindergruppen mit unterschiedlich stark ausgeprägtem mathematischen Vorwissen hin: Kinder mit einem Risikostatus im mathematischen Vorwissen zu Beginn der ersten Klasse erzielen geringere Lernzuwächse als Kinder ohne einen solchen Risikostatus.

5.3 Überlegungen zu Forschungsfragen der Stufe 3 (L.S. Fuchs, 2004)

Auf der Stufe 3 postuliert L.S. Fuchs (2004) die Analyse der Einsetzbarkeit und Effektivität von CBM im schulischen Alltag. Wenn auch erstrebenswert, so ist eine klassische Effektivitätsprüfung nicht möglich, da im vorliegenden Forschungsdesign keine unbehandelte Kontrollgruppe gegeben ist. All den von L.S. Fuchs (2004) gestellten Forderungen vollends nachzukommen, ist somit innerhalb dieser Dissertation nicht realisierbar. Dennoch können Aussagen zu

folgenden Forschungsfragen hinsichtlich der Einsetzbarkeit der CBM getroffen werden:

F_5 Sind die entwickelten CBM in den Unterricht integrierbar und liefern sie den Lehrkräften Informationen zur Identifikation sowie zur Förderung von rechenschwachen Kindern?

Daraus lassen sich folgende Hypothesen ableiten, welchen durch eine Lehrerbefragung nachgegangen werden soll.

$H_{5|1}$ Die Mehrheit der befragten Lehrkräfte schätzt die innerhalb dieser Forschungsarbeit konzipierten CBM als sinnvoll ein.

$H_{5|2}$ Die Mehrheit der befragten Lehrkräfte schätzt den monatlichen Einsatz innerhalb dieser Forschungsarbeit konzipierter CBM als realistisch ein.

$H_{5|3}$ Die Mehrheit der befragten Lehrkräfte empfindet den regelmäßigen Einsatz innerhalb dieser Forschungsarbeit konzipierter CBM als erheblichen Mehraufwand.

$H_{5|4}$ Die Mehrheit der befragten Lehrkräfte, die den regelmäßigen Einsatz innerhalb dieser Forschungsarbeit konzipierter CBM als großen Mehraufwand empfindet, schätzt diesen Mehraufwand als angemessen ein und nimmt ihn in Kauf.

$H_{5|5}$ Die Mehrheit der befragten Lehrkräfte nimmt an, dass es anhand der Schülerergebnisse, der innerhalb dieser Forschungsarbeit konzipierten CBM gelingt, Schüler mit schwachen Rechenleistungen zu identifizieren.

$H_{5|6}$ Die Mehrheit der befragten Lehrkräfte sieht in der Auswertung der Schülerergebnisse in den innerhalb dieser Forschungsarbeit konzipierten CBM eine Hilfestellung für Unterrichtsgestaltung und für die Planung von Fördermaßnahmen.

$H_{5|7}$ Der Einsatz der CBM im Mathematikunterricht der Klassenstufe 1 ist den Kindern zumutbar.

6 Methode der Untersuchung

Die hier beschriebene Untersuchung ist eine Teilstudie des am Institut für Sonderpädagogische Entwicklungsförderung und Rehabilitation der Universität Rostock angesiedelten Forschungsprojektes „Prävention und Integration im RTI-Paradigma“. Ziel dieses vom Ministerium für Bildung, Wissenschaft und Kultur des Landes Mecklenburg-Vorpommern geförderten Forschungsvorhabens liegt in der Entwicklung eines Präventions- und Inklusionskonzepts für die Eingangsstufe der Grundschule. Seit dem Schuljahr 2010/11 wird das „Rügener Inklusionsmodell“ (RIM, Mahlau et al., 2011) als eine Adaption des in Abschnitt 3.1.4 beschrieben Response-to-Instruction-Ansatzes in den ersten Klassen der öffentlichen Schulen der Insel Rügen implementiert. Zentraler Baustein ist neben dem systematischen Einsatz schulischer Präventionsmaßnahmen auf gestuften Förderebenen sowie dem Einsatz evidenzbasierter Unterrichtsmaterialien und -methoden die regelmäßige Kontrolle des Lernverlaufs aller Kinder anhand der Ergebnisse curriculumbasierter Messungen. Eine detaillierte inhaltliche Beschreibung des Forschungsprojektes findet sich in dem Beitrag von Mahlau et al. (2011) sowie Diehl et al. (2012) verwiesen. Neben der Implementation steht die Wirksamkeitsprüfung des RIM anhand eines 2-Gruppen-Versuchsdesigns mit Prä- und Posttest im Fokus des Forschungsprojektes (einen differenzierten Überblick über die Evaluationsmethodik und erste Ergebnisse finden sich bei Voß et al., 2012). Die innerhalb dieser Hauptstudie gewonnenen CBM-Daten sowie weiterer Messungen der untersuchten Schülergruppe auf Rügen bilden in der vorliegenden Teilstudie die Grundlage der durchgeführten Analysen (Substudie 1, siehe Abschnitt 6.2.1). Es handelt sich daher um eine Reanalyse der in dem Rahmen des Forschungsprojektes „Prävention und Integration im RTI-Paradigma“ erhobenen Daten. Mit der Verwendung dieser Daten ergeben sich einige Restriktionen für Analysen, da das Untersuchungsdesign nicht explizit auf die Forschungsfragen der vorliegenden Dissertation

ausgerichtet ist. Die sich durch das Design des Forschungsprojektes „Prävention und Integration im RTI-Paradigma“ ergebende Spezifika bei den Analysen werden innerhalb der Diskussion in Kapitel 9 kommentiert. Da die im Rahmen der Hauptstudie erfassten Daten nicht genügen, um die hier entwickelten CBM hinreichend zu prüfen, wurden weitere Substudien durchgeführt. Im Folgenden werden zunächst die Stichproben der Rügener Studie und weiteren Substudien vorgestellt (Abschnitt 6.1), im Anschluss daran der Untersuchungsverlauf (Abschnitt 6.2), die dabei verwendeten Untersuchungsinstrumente (Abschnitt 6.3) sowie die für die Analysen gewählten Auswertungsmethoden (Abschnitt 6.4).

6.1 Beschreibung der Stichproben

Die vorliegende Studie umfasst vier einzelne Teiluntersuchungen mit jeweils separaten Stichproben. Tabelle 11 gibt eine Übersicht über diese Kohorten, deren Spezifika sowie ihren Zweck. Da keine zufällige Stichprobenauswahl vorgenommen wurde, handelt es sich hier jeweils um quasi-experimentelle Versuchsanordnungen (Bortz & Döring, 2002).

Tabelle 11: Übersicht über die Untersuchungsgruppen sowie deren Besonderheiten und Zweck

Substudie	Untersuchungsgruppe	Region	Personengruppe	N	Besonderheiten	Zweck der Substudie
1	UG 1	Rügen	Erstklässler	472	Beschulung nach RIM-Konzept inkl. monatlicher Durchführung der CBM	Hinweise zur Bestimmung der Validität der CBM
2	UG 2	NVP	Erstklässler	700	monatliche Durchführung der CBM jeweils in zwei Versionen	Hinweise zur Bestimmung der Parallelität der CBM
3	UG 3	Rügen, NVP	Grundschul-lehrkräfte	37	Befragung zum Einsatz von CBM	Hinweise zur Bestimmung von Nutzen und Effektivität der CBM
4	UG 4	Rostock	Erstklässler	174	Durchführung der CBM N1 ohne Begrenzung der Bearbeitungszeit	Hinweise zur Bestimmung der Homogenität der Verfahren

Erläuterungen: UG –Untersuchungsgruppe; N – Stichprobenumfang; NVP – Nordvorpommern; CBM N1 – curriculumbasierte Messverfahren der Niveaustufe 1

Untersuchungsgruppe 1 – Erstklässler Rügens

Um Hinweise zur Bestimmung der Validität der konzipierten CBM zu erhalten, wurden in einer Substudie CBM-Daten sowie Daten weiterer Verfahren (ausführlicher dazu in Abschnitt 6.2 sowie 6.3) der UG 1 erhoben. Wie in Abschnitt 6.1 erwähnt, handelt es sich bei der UG 1 um die Probanden der Wirksamkeitsstudie des RIM und setzt sich aus nahezu allen Schulanfängern der Insel Rügen des Schuljahres 2010/11 zusammen. Diese 472 Erstklässler verteilten sich auf 23 Klassen in 12 staatlichen Grundschulen der Insel Rügen. Schüler, die an Privatschulen unterrichtet werden, wurden in der Untersuchung

nicht erfasst. Aus organisatorischen Gründen wurde die Grundschule Hiddensee nicht in die Stichprobe einbezogen. Die Stichprobe besteht aus 241 Jungen (51.1 %) und 227 Mädchen (48.1 %). Für 4 Kinder (0.8 %) liegen keine Angaben zum Geschlecht vor. Das durchschnittliche Alter der Kinder betrug bei Einschulung 6;7 Jahre (SD = 0;4 Jahre). Angaben zu ethnischen Charakteristika der untersuchten Kinder liegen nicht vor. Die regelmäßigen Messungen mittels CBM in einer etwa vierwöchentlichen (bezogen auf Schulwochen) Frequenz gehören für die Schüler der UG 1 zum Treatment des Präventions- und Integrationskonzepts des RIM.

Im Rahmen des RIM wurde zur Einschätzung des mathematischen Vorwissens in allen Rügener Projektschulen zu Beginn der ersten Klasse der erste Teil des Diagnoseprogramms Kalkulie (Fritz et al., 2007) durchgeführt. Die Rügener Erstklässler erzielten darin einen Mittelwert von 24.8 Punkten bei einer Standardabweichung von 7.9 Punkten, was im Vergleich zur angegebenen Normstichprobe einem Prozentrang von etwa 53, also einer normalen Leistung, entspricht. Zur gleichen Zeit wurde das schriftsprachliche Vorwissen mit dem Münsteraner Screening (MÜSC; Mannhaupt, 2005) erfasst. Bei der Auswertung des MÜSC werden für jede Subskala Risikopunkte vergeben. Die Summe der Risikopunkte entscheidet schließlich, ob die Leistung als auffällig oder unauffällig eingeschätzt werden kann. Von den untersuchten Kindern wurden 83 % als unauffällig, die restlichen 17 % als auffällig in ihrem schriftsprachlichen Vorwissen klassifiziert. Dies deckt sich ungefähr mit den im Manual des MÜSC angegebenen Angaben der Normstichprobe (84.3 % unauffällige und 15.7 % auffällige Kinder). Zur Einschätzung der kognitiven Fähigkeiten der Rügener Kinder wurde ebenfalls zu Beginn des ersten Schuljahres der Grundintelligenztest Skala 1 (CFT 1; Cattell, Weiß & Osterland, 1997) eingesetzt. Im Durchschnitt erzielten die Schüler einen IQ von etwa 97.6 (Standardabweichung 14.5 IQ-Punkte), was einer normalen kognitiven Leistung entspricht.

Untersuchungsgruppe 2 – Erstklässler Nordvorpommerns

In einer weiteren Substudie zur Einschätzung der Parallelität der entwickelten CBM wurden in Absprache mit dem zuständigen Schulamt alle ersten Grundschulklassen der Region Nordvorpommern für eine weitere Substudie rekrutiert. Diese UG 2 setzt sich aus nahezu allen Schulanfängern der Region Nordvorpommern des Schuljahres 2010/11 zusammen. Es handelt sich dabei um 700 Erstklässler verteilt auf 32 Klassen in 21 staatlichen Grundschulen sowie eine Klassen an einer privaten Schule in kirchlicher (evangelischer) Trägerschaft. Die Stichprobe besteht aus 279 Jungen (39.9 %) und 273 Mädchen (39.0 %). Von 8 Klassen (N = 148, 21.1 %) fehlen Angaben zum Geschlecht. Das Durchschnittsalter der teilnehmenden Kinder wurde nicht erfasst und kann somit nicht angegeben werden. Angaben zu ethnischen Charakteristika der untersuchten Kinder liegen ebenfalls nicht vor. Ebenso wie in der UG 1 wurden auch innerhalb dieser Stichprobe regelmäßige CBM-Testungen in einem Abstand von jeweils etwa vier Schulwochen vorgenommen. Darüber hinaus wurde zu jedem der ca. monatlichen Messzeitpunkte eine zweite Version eines jeden CBM durchgeführt. Die Ergebnisse der Schüler aus Nordvorpommern dienen zum einen als Referenzwerte, um Leistungseinschätzungen der Rügener Kinder über die Messzeitpunkte zu realisieren. Zum anderen erlaubt die zeitgleiche Durchführung zweier CBM-Versionen Rückschlüsse auf die Parallelität der CBM. Bedingt durch z. B. vereinzelte krankheitsbedingte Ausfälle, Wegzüge, Verweigerungen der Eltern variierte die Anzahl der teilnehmenden Schüler an den Testungen über den Untersuchungszeitraum. Teilweise sehr hohe Dropout-Raten zu einzelnen Messzeitpunkten (bis zu 57 %) resultierten insbesondere aus schulorganisatorischen Gründen wie z. B. anstehenden Projektwochen bzw. Klassenfahrten oder durch einen als zu hoch erlebten Mehraufwand, speziell zum Schuljahresende, aufgrund der CBM-Testungen.

Untersuchungsgruppe 3 – Grundschullehrkräfte Rügens und Nordvorpommerns

Hinweise zum Nutzen und zur Effektivität der entwickelten CBM für den Unterricht wurden mithilfe einer Befragung einer Lehrerstichprobe innerhalb der Substudie 3 ermittelt. 37 Lehrkräfte der Regionen Rügen und Nordvorpommern, welche mit den innerhalb dieser Forschungsarbeit entwickelten CBM arbeiteten, wurden gebeten, innerhalb einer Onlineumfrage 14 Fragen zur Handhabbarkeit und zum Nutzen der CBM zu beantworten. Von den insgesamt 37 Lehrkräften (♀: N = 36; ♂: N = 1) kamen 20 Grundschullehrkräfte (54.1 %) aus der Region Rügen und 17 Grundschullehrkräfte (45.1 %) aus der Region Nordvorpommern.

Untersuchungsgruppe 4 – Erstklässler Rostocks

In einer weiteren Substudie sollen die entwickelten CBM für den Mathematikunterricht in Klassenstufe 1 differenzierter in Hinblick auf ihre Splithalf-Reliabilität hin untersucht werden. Da die Items für die Verfahren der Niveaustufe eins randomisiert zusammengestellt wurden, stellt sich hier die Frage, inwieweit diese Tests als homogen angesehen werden können (vgl. dazu auch die Ausführungen in Abschnitt 3.3.2.2). Diesem Aspekt soll innerhalb der Substudie 4 nachgegangen werden. Dazu wurde eine zusätzliche Untersuchungsgruppe in die Studie einbezogen: die UG 4. Die UG 4 besteht aus 174 Erstklässlern der Hansestadt Rostock des Schuljahres 2012/13. Die Kinder verteilen sich auf 8 Klassen in 3 staatlichen Grundschulen. Die Stichprobe besteht aus 86 Jungen (49.4 %) und 88 Mädchen (50,6 %). Angaben zum Alter der untersuchten Kinder wurden dabei nicht erhoben. Angaben zu ethnischen Charakteristika der untersuchten Kinder liegen ebenfalls nicht vor.

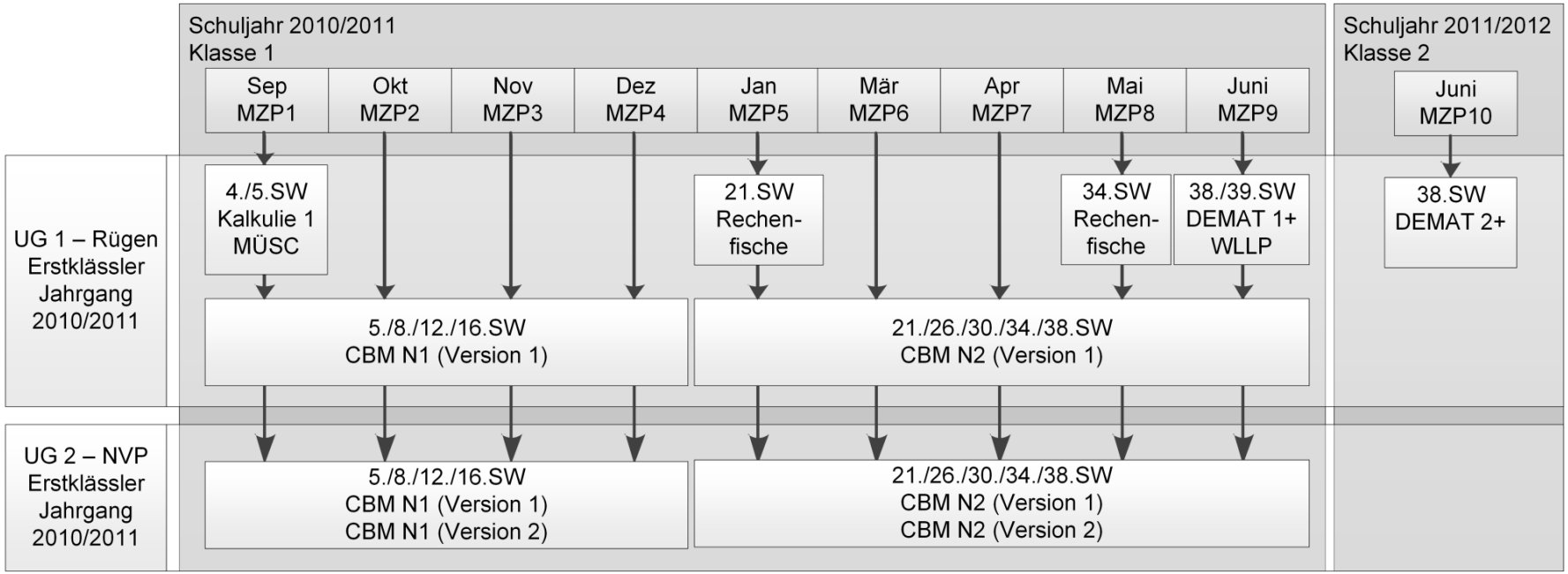
6.2 Untersuchungsdurchführung und -verlauf

In den nachfolgenden Abschnitten sollen die Untersuchungsabläufe der einzelnen Substudien vorgestellt werden. Dabei werden Substudie 1 und 2 zusammengefasst betrachtet, da diese ähnlich aufgebaut waren. Das Vorgehen für die Substudie 3 und 4 wird separat betrachtet.

6.2.1 Teilstudie 1 und 2

Im Vorfeld der Studie, im August bzw. September 2010, wurden die Schulleiter der teilnehmenden Grundschulen sowie die betroffenen Lehrkräfte der ersten Klassen auf der Insel Rügen und in der Region Nordvorpommern über theoretische Aspekte, Ablauf und Ziele einer Lernfortschrittsdokumentation mittels curriculumbasierter Messverfahren informiert. Innerhalb der Substudie 1 wurden in der 5. Schulwoche des Schuljahres 2010/11 (MZP 1) das mathematische und schriftsprachliche Vorwissen der UG 1 mit dem Diagnoseprogramm Kalkulie (Fritz et al., 2007) sowie dem Münsteraner Screening (MÜSC; Mannhaupt, 2005) erhoben. Abbildung 33 gibt einen Überblick über die innerhalb der Substudien 1 und 2 eingesetzten Verfahren. Eine genaue Beschreibung der Erhebungsinstrumente ist in Abschnitt 6.3 nachzulesen. Beginnend ab der 5. Schulwoche wurden im Verlauf des Schuljahres die Schülerleistungen der UG 1 anhand der in dieser Arbeit beschriebenen CBM in regelmäßigen, möglichst vierwöchigen Abständen (bezogen auf Schulwochen) erhoben. Da einige Schulwochen aufgrund verschiedener Feier- bzw. Ferientage kürzer ausfielen, umfassen einzelne Zeitabstände zwischen den Testungen mehr als die ursprünglich geplanten vier Schulwochen. Zu Beginn des Schuljahres ist der zeitliche Abstand zwischen dem Messzeitpunkt 1 und 2 aus organisatorischen Gründen verkürzt worden. Welche CBM-Version zum jeweiligen Messzeitpunkt eingesetzt wurde ist Tabelle 12, Tabelle 13 und Tabelle 14 zu entnehmen. Zur 21. Schulwoche

(MZP 5) und zur 34. Schulwoche (MZP 8) erfolgte innerhalb der UG 1 eine wiederholte Messung der mathematischen Kompetenzen anhand des Inventars Rechenfische (Wagner & Hartke, 2006). Diese Daten dienen zur Bestimmung der Konstrukt- bzw. Kriteriumsvalidität. Darüber hinaus wurden zum Schuljahresende der 1. Klasse (9. MZP) in der UG 1 die Leistungen in den Lernbereichen Mathematik, erhoben mit dem Deutschen Mathematiktest für 1. Klassen (DEMAT 1+; Krajewski et al., 2002) und Lesen, erhoben mit der Würzburger Leise Leseprobe (WLLP; Küspert & Schneider, 1998), durchgeführt. Auch diese Ergebnisse können zur Bestimmung der Konstrukt- bzw. Kriteriumsvalidität herangezogen werden. Zum Ende der Klasse zwei (MZP 10), also nach einer Schulbesuchszeit von zwei Jahren, wurden mit dem Deutschen Mathematiktest für zweite Klassen (DEMAT 2+; Krajewski, Liehm & Schneider, 2004) zusätzlich weitere Schulleistungsdaten im Fachbereich Mathematik für weiterführende Analysen hinsichtlich der prognostischen Validität der CBM erhoben.



Erläuterungen: UG –Untersuchungsgruppe; NVP – Nordvorpommern, MZP – Messzeitpunkt; SW – Schulwoche; CBM N1 – Curriculumbasierte Messungen Niveaustufe 1; CBM N2 – Curriculumbasierte Messungen Niveaustufe 2

Abbildung 33: Überblick über die zu den verschiedenen Messzeitpunkten eingesetzten Testverfahren

Im Rahmen der Substudie 2 nahmen die Erstklässler des Schuljahres 2010/2011 der Region Nordvorpommern (UG 2) zeitgleich an den regelmäßigen, in Substudie 1 beschriebenen CBM-Testungen teil. Zu allen Messzeitpunkten bearbeitete die UG 2 jeweilig zwei Versionen eines jeden CBM. Die teilnehmenden Lehrkräfte wurden darauf hingewiesen, diese Testungen idealerweise an zwei aufeinanderfolgenden Schultagen zu realisieren. Aus organisatorischen Gründen war es jedoch nicht möglich, die Einheitlichkeit der Durchführung in teilnehmenden Klassen abzusichern, wodurch eine damit möglicherweise einhergehende Verzerrung der Schülerergebnisse in den einzelnen Klassen nicht vollständig ausgeschlossen werden kann.

Tabelle 12 gibt Aufschluss darüber, welche Versionen des CBM „Arithmetische Basiskompetenzen“ zu den verschiedenen Messzeitpunkten innerhalb des ersten Schuljahres durchgeführt wurden. Anhand der in der Region Nordvorpommern (UG 2) zu den jeweiligen Messzeitpunkten zeitgleich eingesetzten CBM können Informationen zur Paralleltest-Reliabilität, d. h. zur Parallelität der Verfahren, gewonnen werden. Dazu werden die Schülerergebnisse der zeitgleich eingesetzten Verfahren mit einander korreliert, wie bspw. die Ergebnisse für CBM1 und CBM2 usw. (CBM1↔CBM2, CBM3↔CBM4, CBM5↔CBM6, CBM1↔CBM7, CBM1↔CBM3, CBM4↔CBM5, CBM6↔CBM2, CBM7↔CBM8, CBM9↔CBM10). Da einige CBM-Versionen zu verschiedenen Messzeitpunkten eingesetzt wurden, können Angaben zur Retest-Reliabilität getroffen werden. Dazu werden die Korrelationen der Schülerergebnisse wiederholt eingesetzter CBM-Verfahren analysiert (CBM1: MZP1↔MZP4↔MZP5; CBM2: MZP1↔MZP7; CBM3: MZP2↔MZP5; CBM4: MZP2↔MZP6; CBM5: MZP3↔MZP6; CBM6: MZP3↔MZP7; CBM7: MZP4↔MZP8).

Tabelle 12: Eingesetzte Versionen der CBM „Arithmetische Basiskompetenzen“ in den Regionen Rügen (UG 1) und Nordvorpommern (UG 2) innerhalb des ersten Schuljahres (Niveaustufe 1 und 2)

MZP	eingesetzte CBM-Version in den Regionen Rügen und NVP	zusätzlich eingesetzte CBM-Version in der Region NVP
1	CBM1	CBM2
2	CBM3	CBM4
3	CBM5	CBM6
4	CBM1	CBM7
5	CBM1	CBM3
6	CBM4	CBM5
7	CBM6	CBM2
8	CBM7	CBM8
9	CBM9	CBM10

Erläuterungen: MZP – Messzeitpunkt; NVP - Nordvorpommern

In Tabelle 13 ist dargestellt, welche Versionen der CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ und „Mengenvergleich im Zwanzigerraum“ zu den verschiedenen Messzeitpunkten innerhalb der ersten Schuljahreshälfte der ersten Klassenstufe (Niveaustufe 1) durchgeführt wurden. Anhand der in der Region Nordvorpommern (UG 2) zu den jeweiligen Messzeitpunkten zeitgleich eingesetzten CBM können Informationen über die Paralleltest-Reliabilität gewonnen werden (CBM1↔CBM2, CBM3↔CBM4, CBM5↔CBM6, CBM1↔CBM7). Da die CBM der Version CBM1 zum ersten als auch zum vierten Messzeitpunkt eingesetzt wurden, können Koeffizienten hinsichtlich der Retest-Reliabilität geschätzt werden (CBM1: MZP1↔MZP4).

Tabelle 13: Eingesetzte Versionen der CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ und „Mengenvergleich im Zwanzigerraum“ in den Regionen Rügen und Nordvorpommern in der ersten Hälfte des ersten Schuljahres (Niveaustufe 1)

MZP	eingesetzte CBM-Version in den Regionen Rügen und NVP	zusätzlich eingesetzte CBM-Version in der Region NVP
1	CBM1	CBM2
2	CBM3	CBM4
3	CBM5	CBM6
4	CBM1	CBM7

Erläuterungen: MZP – Messzeitpunkt; NVP - Nordvorpommern

Welche Versionen der CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ zu den verschiedenen Messzeitpunkten innerhalb der zweiten Schuljahreshälfte der ersten Klassenstufe (Niveaustufe 2) eingesetzt wurden, ist in Tabelle 14 dargestellt. Anhand der in Region Nordvorpommern zu den unterschiedlichen Messzeitpunkten zeitgleich eingesetzten CBM können auch hier Analysen der Paralleltest-Reliabilität vorgenommen werden (CBM1↔CBM2, CBM3↔CBM4, CBM5↔CBM6, CBM7↔CBM8, CBM1↔CBM4). Da die CBM der Version CBM1 bzw. CBM4 zum fünften als auch zum neunten bzw. zum sechsten und zum neunten Messzeitpunkt eingesetzt wurden, können verschiedene Koeffizienten hinsichtlich der Retest-Reliabilität ermittelt und eingeschätzt werden (CBM1: MZP5↔MZP9; CBM4: MZP6↔MZP9).

Tabelle 14: Eingesetzte Versionen der CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ in den Regionen Rügen und Nordvorpommern in der zweiten Hälfte des ersten Schuljahres (Niveaustufe 2)

MZP	eingesetzte CBM-Version in den Regionen Rügen und NVP	zusätzlich eingesetzte CBM-Version in der Region NVP
5	CBM1	CBM2
6	CBM3	CBM4
7	CBM5	CBM6
8	CBM7	CBM8
9	CBM1	CBM4

Erläuterungen: MZP – Messzeitpunkt; NVP - Nordvorpommern

Alle in der Substudie 1 und 2 aufgeführten Testungen wurden im Rahmen des regulären Unterrichts durchgeführt. Die Durchführung wurde von den Grundschullehrkräften selbst angeleitet. Ebenso wurde die Auswertung von ihnen übernommen. Lediglich die Erhebung der Mathematikleistungen mit dem DEMAT 2+ (Krajewski et al., 2004) zum Ende Klasse zwei wurden im Rahmen der Substudie 1 durch geschulte studentische Hilfskräfte des Instituts für Sonderpädagogische Entwicklungsförderung und Rehabilitation der Universität Rostock durchgeführt.

Um die Objektivität bei der Durchführung der CBM-Testungen zu gewährleisten, wurden standardisierte Durchführungsbedingungen vorgegeben, sodass die Testbedingungen jeweils konstant gehalten wurden. Es wurden den an der Studie teilnehmenden Lehrkräften allgemeine Bedingungen mitgeteilt, welche folgende Aspekte umfassen:

- Die CBM sollten in einer ruhigen, störungsfreien Atmosphäre bearbeitet werden.
- Aufgaben-Instruktionen werden wörtlich vorgelesen.
- Die Bearbeitungszeit von jeweils 1 Minute muss eingehalten werden.
- Es muss verhindert werden, dass Kinder voneinander abschreiben.
- Der Testleiter sollte nach Möglichkeit die unterrichtende Lehrkraft sein.

Darüber hinaus lagen neben diesen allgemeinen Hinweisen für jedes CBM separate Durchführungshinweise vor. Zusätzlich zu den benötigten Materialien (bspw. Stoppuhr) wurde hier auch die vorzulesende wörtliche Instruktion aufgeführt. Damit das Testergebnis nicht davon abhängt, wer den Test auswertet, wurden allen teilnehmenden Lehrkräften klare Auswertungsregeln vorgegeben. Alle Aufgaben der innerhalb dieser Studie vorgestellten CBM haben eine eindeutige Lösung. Richtig gelöste Aufgaben wurden dabei mit einem Punkt bewertet, falsch gelöste Aufgaben mit null Punkten. Der Gesamtwert wurde durch Addition der einzelnen Punkte jeweils für jedes CBM ermittelt. Dieser Wert wurde zum Beobachten des Lernverlaufs dokumentiert. Zur weiteren Gewährleistung der Objektivität lagen für die meisten Verfahren Lösungsfolien vor, welche zur Kontrolle der Rechnungen auf die Aufgabenvorlage gelegt werden können. Als Hilfsmittel zur Dokumentation der Lernverläufe der Schüler wurden allen

teilnehmenden Lehrkräften in Microsoft Excel programmierte Dateien ausgehändigt. Diese Dateien ermöglichten eine automatische grafische Auswertung der eingetragenen Werte für jeden Schüler und für jedes CBM (vgl. Abbildung 34). Zusätzlich wurde automatisch der Klassendurchschnitt ermittelt und ebenfalls in der grafischen Auswertung eines jeden Schülers als Referenzwert ausgegeben.

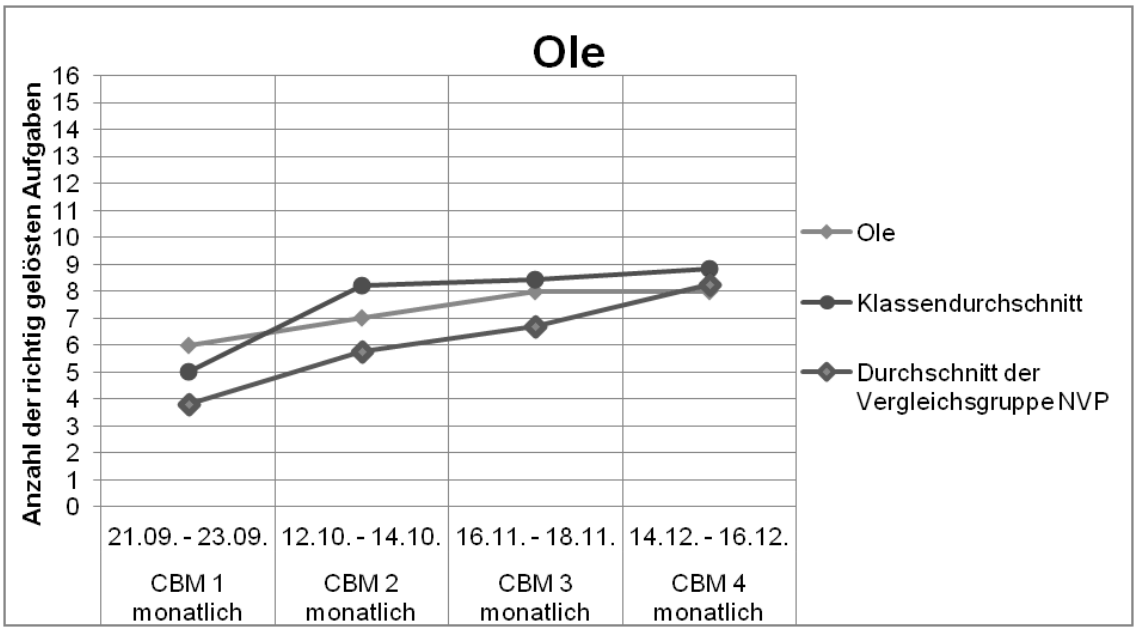


Abbildung 34: Beispiel einer mithilfe der vorprogrammierten Microsoft-Excel-Datei automatisch generierten grafischen Ausgabe

Um die erzielten Leistungen der Kinder detaillierter einschätzen zu können, wurden weiterhin ausgewählte Prozentrangwerte (PR < 10, PR ≤ 25, PR ≥ 75, PR > 90) für die monatlichen Messungen an die Lehrkräfte zurückgemeldet. Diese Vergleichswerte ergaben sich aus den Schülerergebnissen der Stichprobe der Erstklässler der Region Nordvorpommern. Diese können jedoch, aufgrund ihrer eingeschränkten Repräsentativität, nicht als Normierungswerte, sondern lediglich als ein Vergleichswert zur Abschätzung der durchschnittlichen Schülerleistungen in den Verfahren angesehen werden. Der in dieser Stichprobe ermittelte Durchschnittswert konnte ebenfalls in den Auswertungsdateien notiert werden und ging in die automatisch generierte grafische Auswertung ein. Tabelle 15 zeigt die

Zuordnung von erzielten Leistungseinschätzungen zu Prozentrangbereichen, welche den Lehrkräften mitgeteilt wurde.

Tabelle 15: Leistungseinschätzung auf Grundlage erreichter Prozentrangwerte

Schülerleistungen gelten als	weit überdurchschnittlich	bei einem Prozentrang	> 90.
	überdurchschnittlich		≥ 75 und ≤ 90.
	durchschnittlich		< 75 und > 25.
	unterdurchschnittlich		≤ 25 und ≥ 10.
	weit unterdurchschnittlich		< 10.

Um die Erstellung der Entwicklungsgrafen zusätzlich zu objektivieren und zu erleichtern, wurden Auswertungsdateien per Excel programmiert, welche die systematische Sammlung der Ergebnisse einer Klasse sowie die automatische grafische Auswertung der erhobenen Daten ermöglichen sollten und zur Übermittlung der Daten an die Universität Rostock dienten. Den Lehrkräften wurden nach jedem Messzeitpunkt die ermittelten Referenzwerte sowie zusätzliche Hinweise zur Leistungseinschätzung der Schüler mitgeteilt (siehe dazu auch Abbildung 34).

6.2.2 Teilstudie 3

Innerhalb der Substudie 3 wurden die an den CBM-Testungen beteiligten Lehrkräfte der Regionen Rügen und Nordvorpommern nach dem 1. Schuljahr gebeten, innerhalb einer Onlineumfrage, 13 Aussagen zu den im Rahmen der vorliegenden Arbeit konzipierten curriculumbasierten Messungen einzuschätzen. Dabei wurden neben allgemeinen Aspekten, eine Einschätzung der Verfahren selbst und ihr Nutzen in der pädagogischen Arbeit sowie die Umsetzbarkeit im Unterricht im Sinne von Praktikabilität abgefragt. Die Umfrage war technisch derart konzipiert, dass jede Lehrkraft nur einmalig teilnehmen konnte, um sicherzustellen, dass die Ergebnisse nicht durch mehrmaliges Bearbeiten einzelner Personen verzerrt werden

konnten. Darüber hinaus konnte die Onlineumfrage nur durch Einschätzung aller aufgeführten Items beendet werden, eine Ausweichkategorie der Art „keine Angabe“ oder ähnliches wurde nicht gegeben.

6.2.3 Teilstudie 4

Auch die Testungen in der Hansestadt Rostock innerhalb der Substudie 4 wurden im Vorfeld mit dem zuständigen Schulamt abgesprochen, ebenso erhielten die Eltern der teilnehmenden Kinder ein Informationsschreiben zu Sinn und Zweck der Untersuchung sowie den Hinweis, dass sie ihre Kinder aus der Untersuchung ausschließen lassen könnten, wenn dies gewünscht werde. Um den Aufwand für die teilnehmenden Schüler möglichst gering zu halten, wurde die Schülerkohorte der Region Rostock in zwei Gruppen aufgeteilt. Die eine Gruppe bearbeitete ausschließlich die ersten vier bzw. fünf Versionen eines jeden CBM, die zweite Gruppe die jeweils weiteren vier bzw. fünf Versionen. Jeweils zwei Versionen der CBM „Zahlenreihe im Zwanzigerraum“, „Mengenvergleich im Zwanzigerraum“ und zwei bzw. drei Versionen des CBM „Arithmetische Basiskompetenzen“ wurden dabei zu einem Testheft zusammengefasst (Tabelle 16). Jede Gruppe bearbeitete jeweils ein solches Testheft in zwei Schulstunden. Je Schulstunde werden dabei ungefähr 30 Minuten der Unterrichtszeit beansprucht. Für jede Version eines jeden CBM stand den Kindern jeweils zweimal eine 30-sekündige Bearbeitungszeit zur Verfügung. Immer nach Ablauf einer jeden Bearbeitungszeit wurden die Kinder durch ein Signal des Testleiters darauf hingewiesen, die Bearbeitung zu unterbrechen und einen Strich als Markierung der zuletzt bearbeiteten Aufgabe zu setzen. Sobald zwei 30-sekündige Bearbeitungszeiträume für eine Version durchlaufen waren, wurde auf Hinweis des Testleiters zum nächsten CBM weitergeblättert.

Tabelle 16: Angaben zum Einsatz der verschiedenen CBM-Versionen innerhalb der Substudie 3 in der Region Rostock

CBM	Gruppe 1		Gruppe 2	
	Inhalte des Testhefts der Schulstunde 1	Inhalte des Testhefts der Schulstunde 2	Inhalte des Testhefts der Schulstunde 1	Inhalte des Testhefts der Schulstunde 2
Zahlenreihe im Zwanzigerraum	Version 1	Version 3	Version 5	Version 7
	Version 2	Version 4	Version 6	Version 8
Mengenvergleich im Zwanzigerraum	Version 1	Version 3	Version 5	Version 7
	Version 2	Version 4	Version 6	Version 8
Arithmetische Basiskompetenzen	Version 1	Version 4	Version 6	Version 9
	Version 2	Version 5	Version 7	Version 10
	Version 3		Version 8	

Da das CBM „Zahlenlesen im Zwanzigerraum“ in Einzeltestungen durchgeführt werden musste, war der Testaufwand deutlich höher. Die Testungen wurden jeweils parallel zum Unterricht durchgeführt. Nacheinander wurden die Schüler, jeweils einzeln, zum Testleiter gerufen, um die entsprechenden vier Versionen des CBM nacheinander zu absolvieren. Für jede Version wurde jeweils nach 30 Sekunden durch den Testleiter vermerkt, welche Zahl das Kind zuletzt (richtig oder falsch) benannt hatte. Nach zweimaligem Ablauf der 30-sekündigen Bearbeitungsdauer für ein Testverfahren wurde zum nächsten übergegangen.

Der Testzeitraum erstreckte sich über drei Wochen im September bzw. Oktober 2012. Anschließend wurden die Daten von eingewiesenen studentischen Hilfskräften entsprechend der Vorgaben ausgewertet.

6.3 Erhebungsinstrumente

Nachfolgend sollen alle Erhebungsinstrumente, welche zur Beantwortung der dieser Studie zugrunde liegenden Forschungsfragen verwendet wurden, näher beschrieben werden. Neben allgemeinen Informationen zu den einzelnen Verfahren sollen

vor allem auch Angaben in Hinblick auf ihre Güte beschrieben werden.

Um die Validität der hier beschriebenen CBM abzuschätzen, wurden folgende Verfahren eingesetzt:

- das Diagnoseprogramm Kalkulie (Fritz et al., 2007)
- das Münsteraner Screening (MÜSC, Mannhaupt, 2005)
- die Rechenfische (Wagner & Hartke, 2006)
- die Würzburger Leise Leseprobe (WLLP; Küspert & Schneider, 1998)
- der Deutsche Mathematiktest für erste Klassen (DEMAT 1+; Krajewski et al., 2002)
- der Deutsche Mathematiktest für zweite Klassen (DEMAT 2+; Krajewski et al., 2004)

Zur Auswahl der eingesetzten Testverfahren wurden neben inhaltlichen und testtheoretischen auch Überlegungen hinsichtlich der Ökonomie einbezogen. So waren angesichts des Umfangs der Stichprobe neben der Güte der Verfahren die Möglichkeit der Testung in Gruppen sowie eine möglichst kurze Bearbeitungsdauer bezüglich Durchführung und Auswertung zentrale Aspekte.

Zur Abschätzung der Ökonomie, der Nützlichkeit und der Zumutbarkeit der CBM wurde zusätzlich eine internetgestützte Befragung der teilnehmenden Lehrkräfte durchgeführt. Dabei handelt es sich um eine speziell für diesen Zweck entworfene Sammlung von Aussagen bezüglich der in der vorliegenden Arbeit konzipierten CBM sowie deren Einsatz, welche durch die Teilnehmer eingeschätzt werden sollen. In den nachfolgenden Abschnitten werden detaillierte Auskünfte über alle eingesetzten Erhebungsverfahren gegeben. Aus Gründen der Übersichtlichkeit erfolgt die Darstellung der Erhebungsverfahren jeweils nach einem einheitlichen Schema ergänzt um Angaben zur Anwendung in der Studie.

6.3.1 Kalkulie (Fritz et al., 2007)

Testart	Entwicklungsstanddiagnose, Test zur Entwicklung mathematischer Kompetenzen
Zweck	Die Testaufgaben sind in Anlehnung an das Modell der mathematischen Kompetenzentwicklung nach Fritz et al. (2007) konstruiert. Es werden grundlegende mathematische Einsichten und Konzepte überprüft wie bspw. Kenntnis der Zahlenreihe, Erkennen der Kardinalität von Zahlen oder Einsicht in das Teile-Ganzes-Konzept. Das Verfahren besteht aus 3 Teilen, welche jeweils aus verschiedenen Subtests zusammengesetzt sind: • Teil 1: Aufgaben zum Mengenvergleich, zur Zahlwortsequenz, zum Zahlvergleich, zum Teile-Ganzes-Konzept mit und ohne Situationsbezug, zum Abzählen, zu Zählzahl und Anzahl sowie zu Zahlrelationen (insgesamt 9 Untertests) • Teil 2: Aufgaben zum Mengenordnen, zu Anzahlen, zu Zahlpositionen, zu 5er- und 10er-Strukturen, Zahlbeziehungen sowie zum Kopfrechnen (insgesamt 6 Untertests) • Teil 3: Aufgaben zur Kraft der 5, Kraft der 10, zum Zerlegen in Teilmengen, zu Umkehraufgaben, zur Kompensation bei der Addition, zum gleichsinnigen Verändern, zur Kovarianz sowie zu Beziehungen zwischen Rechenaufgaben (insgesamt 9 Untertests) Das Verfahren ist in zwei Pseudo-Parallelformen A und B verfügbar.
Einsatzbereich	Das Verfahren kann von Klasse eins bis drei zu verschiedenen Zeitpunkten eingesetzt werden: • Teil 1: Beginn, Mitte, Ende Klasse eins • Teil 2: Mitte, Ende Klasse eins sowie Mitte, Ende Klasse zwei • Teil 3: Mitte, Ende Klasse zwei sowie Mitte, Ende Klasse drei
Testform	• Individual- oder Gruppentest
Durchführungszeit	• Teil 1: ca. 45min • Teil 2: ca. 45min • Teil 3: ca. 60min
Objektivität	• durch standardisierte Durchführungs- und Auswertungshinweise gegeben
Reliabilität	Zur Reliabilitätsanalyse wurde die interne Konsistenz des Verfahrens errechnet. Es ergeben sich Werte für Cronbachs α zwischen: • .77 und .87 für Teil 1 • .81 und .82 für Teil 2 • .79 und .89 für Teil 3
Validität	• Kriteriumsvalidität ermittelt durch Korrelation Kalkulie Teil 1 mit dem OTZ: r = .65 (N = 32) • prognostische Validität ermittelt durch Korrelation Kalkulie Teil 1 mit dem DEMAT-1+: r = .69 (N = 74)

Normen	Es liegen Prozentrangwerte auf Basis der Ergebnisse von 2513 Kindern vor. Die Normen sind nach Messzeitpunkt getrennt aufgeführt: • Teil 1: Beginn, Mitte, Ende Klasse eins • Teil 2: Mitte, Ende Klasse eins sowie Mitte, Ende Klasse zwei • Teil 3: Mitte, Ende Klasse zwei sowie Mitte, Ende Klasse drei
Erläuterungen:	OTZ – Osnabrücker Test zur Zahlbegriffsentwicklung (van Luit, van de Rijt & Hasemann, 1994), DEMAT 1+ – Deutscher Mathematiktest für erste Klassen (Krajewski et al., 2002) ; N – Stichprobenumfang, r – Korrelationskoeffizient

In der vorliegenden Studie wurden Informationen zur Entwicklung der mathematischen Kompetenzen der Kinder der Region Rügen mit dem Teil 1 von Kalkulie (Fritz et al., 2007) zu Beginn des ersten Schuljahres gemessen. Es wurde zu beiden Messzeitpunkten ausschließlich die Version A eingesetzt.

6.3.2 Münsteraner Screening (MÜSC, Mannhaupt, 2005)

Testart	Entwicklungsstanddiagnose, Test zur Entwicklung schriftsprachlicher Kompetenzen
Zweck	Das MÜSC dient der Bestimmung der Lesevoraussetzungen bei Kindern und soll die Identifikation von Kindern mit besonderem Förderbedarf im Schriftspracherwerb ermöglichen. Es umfasst Aufgaben zur Überprüfung: • der phonologischen Bewusstheit, • der Kurzzeitgedächtniskapazität, • der Abrufgeschwindigkeit und • der visuellen Aufmerksamkeit. Das Verfahren ist in zwei Parallelformen A und B verfügbar.
Einsatzbereich	• Beginn der Grundschule Klasse eins
Testform	• Individual- oder Gruppentest
Durchführungszeit	• 2 mal ca. 25min
Objektivität	• durch standardisierte Durchführungs- und Auswertungshinweise gegeben
Reliabilität	• Reliabilität der UT: r_{tt} zwischen .65 und .88
Validität	Kriteriumsvalidität ermittelt durch Ermittlung der Höhe der Übereinstimmung mit dem BISC: RATZ = 60.1 Hinweise zur Konstruktvalidität durch Faktorenanalyse gegeben
Normen	Es liegen Normen von N = 2896 Kindern aus Thüringen und Nordrhein-Westfalen vor.
Erläuterungen:	UT – Untertests, BISC - Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (Jansen, Mannhaupt, Marx &

Skowronek, 1999); r_{tt} – Korrelationskoeffizient; RATZ – Relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote; N – Stichprobenumfang

In der vorliegenden Studie wurden die Kinder nur mit der Version A des MÜSC (Mannhaupt, 2005) getestet.

6.3.3 Rechenfische (Wagner & Hartke, 2006)

Testart	Lernfortschrittsdiagnostikum, Inventar zur Erfassung mathematischer Kompetenzen
Zweck	Die Rechenfische dienen der Identifikation von Schwierigkeiten beim Erlernen arithmetischer Kenntnisse bei Grundschulkindern in Kassenstufe 1. Weiterhin soll es Ansatzpunkte für notwendige Fördermaßnahmen liefern. Das Verfahren orientiert sich am Curriculum sowie an aktuellen fachdidaktischen und entwicklungspsychologischen Modellen. Neben Subtests, welche das konkrete Lösen von Rechenaufgaben erfordern, gibt es auch Aufgabenbereiche, welche erste arithmetische Einsichten abfragen: • Zahl-Mengen-Zuordnung • Addition: a+b=x, x≤10 • Addition: a+b=x, 10<x≤20 • Subtraktion: a-b=x, a≤10 • Subtraktion: a-b=x, 10<a≤20 • Addition: a+x=c, c≤20 • Addition: x+b=c, c≤20 • Subtraktion: x-b=c, x≤20 • Subtraktion: a-x=c, a≤20 • Addition: a+b=x, x≤10 in Päckchen • Addition: a+b=x, 10<x≤20 in Päckchen • Subtraktion: a-b=x, a≤10 in Päckchen • Subtraktion: a-b=x, 10<a≤20 in Päckchen • Zahlen zerlegen • Bild-Rechenoperation-Zuordnung • Zahlreihen ergänzen • Textaufgaben • Größer-Kleiner-Vergleiche • Zahlenstrahl • Kettenaufgaben
Einsatzbereich	• 3 Messzeitpunkte in der Grundschule Klasse eins 1. MZP: 21. - 23. Schulwoche 2. MZP: 33. - 34. Schulwoche 3. MZP: 43. - 44. Schulwoche
Testform	Individual- oder Gruppentest
Durchführungszeit	• ca. 2 mal 45 Min

Objektivität	• durch standardisierte Durchführungs- und Auswertungshinweise gegeben
Reliabilität	• Retest-Reliabilität: 1. zu 2. MZP: r = .73, 2. zu 3. MZP: r = .76, 1. zu 3. MZP: r = .68 • Cronbachs α: 1. MZP: r = .98, 2. MZP: r = .99, 3. MZP: r = .99 • Splithalf-Reliabilität: 1. MZP, 2. MZP, 3. MZP: r = .91
Validität	Durch die Orientierung der Testaufgaben an allen deutschen Mathematiklehrplänen ist die Lehrplanvalidität gegeben. Die Kriteriumsvalidität wurde durch Korrelation mit dem DEMAT 1+ (Krajewski et al., 2002) zum Ende Klasse eins ermittelt: 1. MZP: r = .73, 2. MZP: r = .78, 3. MZP: r = .79. Mittels Regressionen (zwischen den MZP und dem DEMAT 1+; Krajewski et al., 2002) wurde die prognostische Validität überprüft: 1. MZP: R^2 = .61. 2. MZP: R^2 = .64, 1. und 2. MZP zusammen: R^2 = .70
Normen	• Es sind Prozentrangnormen getrennt nach MZP für den Gesamtwert sowie für die einzelnen Subskalen angegeben (N = 1352 bis N = 1455).

Erläuterungen: MZP – Messzeitpunkt; N – Stichprobenumfang, r – Korrelationskoeffizient

6.3.4 Würzburger Leise Leseprobe – WLLP (Küspert & Schneider, 1998)

Testart	Schulleistungstest zur Erfassung der Lesegeschwindigkeit
Zweck	Das Verfahren erfasst die Dekodiergeschwindigkeit von Grundschulkindern mittels 140 Items. Dabei soll immer einem geschriebenen Wort jeweils eine der vier vorgegebenen Bildalternativen zugeordnet werden. Die Dekodiergeschwindigkeit (Lesegeschwindigkeit) wird als bedeutsames Maß der Leseleistung angesehen (Küspert & Schneider, 1998). Das Verfahren ist in zwei Pseudo-Parallelformen A und B verfügbar.
Einsatzbereich	• Grundschule Klasse eins bis vier (jeweils in den letzten beiden Monaten des Schuljahres) • Sonderschule für die Bereiche Sprache und Lernen
Testform	• Individual- oder Gruppentest
Durchführungszeit	• ca. 20 Min
Objektivität	• durch standardisierte Durchführungs- und Auswertungshinweise gegeben
Reliabilität	Es wurde die Paralleltest-Reliabilität zu allen Messzeitpunkten ermittelt (Intervall 14 Wochen): Klasse eins: r = .87, Klasse zwei: r = .92, Klasse drei: r = .93, Klasse vier: r = .82. Ebenso zu den vier Messzeitpunkten wurde die Retest-Reliabilität berechnet: Klasse eins: r = .75, Klasse zwei: r = .81, Klasse drei: r = .88, Klasse vier: r = .78.

Validität	Kriteriumsvalidität ermittelt durch Korrelation mit verschiedenen Außenkriterien: • DLF 1-2: r = .68 (Klasse eins, N = 74), r = .51 (Klasse zwei, N = 67) • BLT 3-4: r = .79 (Klasse drei, N = 67), r = .72 (Klasse vier, N = 73) • Lehrerurteil: r = .75 (Klasse eins, N = 74), r = .58 (Klasse zwei, N = 67), r = .39 (Klasse drei, N = 67), r = .56 (Klasse vier, N = 73) • Deutschnote: r = .45 (Klasse drei, N = 67), r = .43 (Klasse vier, N = 73)
Normen	• Es sind Prozentrangnormen für die Klassenstufen 1 bis 4 getrennt nach Geschlecht angegeben (N = 2820).

Erläuterungen: DLF – Diagnostischer Lesetest zur Frühdiagnose von Lesestörungen (Müller, 1984), BLT 3-4 – Bremer Lesetest für 3. und 4. Klassen (Niemeyer, 1976); r – Korrelationskoeffizient; N – Stichprobenumfang

In der vorliegenden Studie wurden die Kinder nur mit der Version A der WLLP (Küspert & Schneider, 1998) getestet.

6.3.5 Deutscher Mathematiktest für erste Klassen – DEMAT 1+ (Krajewski et al., 2002)

Testart	Schuleistungstest, Test zur Überprüfung mathematischer Kompetenzen
Zweck	Der DEMAT 1+ ist ein an den Lehrplänen aller deutschen Bundesländer orientiertes Verfahren zur Ermittlung mathematischer Kompetenzen von Grundschülern zum Ende der ersten bzw. Anfang der zweiten Klasse. Das Verfahren besteht aus folgenden 8 Untertests: • Mengen-Zahlen (MZ) • Zahlenraum (ZR) • Addition (AD) • Subtraktion (SU) • Zahlenzerlegung-Zahlenergänzung (ZZ) • Teile-Ganzes (TG) • Kettenaufgaben (KA) • Ungleichungen (UG) • Sachaufgaben (SA) Das Verfahren ist in zwei Pseudo-Parallelformen A und B verfügbar.
Einsatzbereich	Ende erstes Schuljahr, Anfang zweites Schuljahr
Testform	Individual- oder Gruppentest
Durchführungszeit	• ca. 45 Min
Objektivität	• durch standardisierte Durchführungs- und Auswertungshinweise gegeben

Reliabilität	Zur Analyse der Reliabilität wurde die Retest-Reliabilität (r = .65 bei einem zeitlichen Abstand von 3 Monaten, N = 52) errechnet. Zusätzlich wurde Cronbachs α für Klasse eins (r = .89, N = 1354) sowie Klasse zwei (r = .88, N = 1582) ermittelt.
Validität	Durch die Orientierung der Testaufgaben an allen deutschen Mathematiklehrplänen ist die Lehrplanvalidität gegeben. Die Kriteriumsvalidität wurde durch Korrelation mit dem DBZ 1, dem Lehrerurteil sowie einem informellen Tests zur Ermittlung der Schnelligkeit beim Kopfrechnen ermittelt: • DBZ 1: r = .77 (N = 97) • Lehrerurteil: r = .66 (N = 195) • Schnelligkeit beim Kopfrechnen: r = .57 (N = 195)
Normen	Es liegen Prozentrang-, T-Werte und T-Wert-Bänder auf Basis der Ergebnisse von insgesamt 2936 Kindern vor. Die Normen sind für den Gesamttest, die Untertests und getrennt nach Geschlecht zu zwei Messzeitpunkten aufgeführt: • Letzter Schuljahresmonat Klasse eins (N = 1354) • Erste drei Schuljahresmonate Klasse zwei (N = 1582)
Erläuterungen:	DBZ 1 – Diagnostikum: Basisfähigkeiten im Zahlenraum 0 bis 20 (Wagner & Born, 1994); r – Korrelationskoeffizient; N – Stichprobenumfang

In der vorliegenden Studie wurden die Kinder nur mit der Version A des DEMAT 1+ (Krajewski et al., 2002) getestet.

6.3.6 Deutscher Mathematiktest für zweite Klassen – DEMAT 2+ (Krajewski et al., 2004)

Testart	Schuleistungstest, Test zur Überprüfung mathematischer Kompetenzen zum Ende der zweiten Klasse
Zweck	Der DEMAT 2+ ist ein an den Lehrplänen aller deutschen Bundesländer orientiertes Verfahren zur Ermittlung mathematischer Kompetenzen von Grundschülern zum Ende der zweiten bzw. Anfang der dritten Klasse. Das Verfahren besteht aus folgenden 10 Untertests: • Zahleneigenschaften, • Längenvergleich, • Addition, • Subtraktion, • Verdoppeln, • Division, • Halbieren, • Rechnen mit Geld, • Sachaufgaben und • Geometrie Das Verfahren ist in zwei Pseudo-Parallelformen A und B verfügbar.
Einsatzbereich	Ende zweites Schuljahr, Anfang drittes Schuljahr
Testform	Individual- oder Gruppentest
Durchführungszeit	• ca. 45 Min
Objektivität	• durch standardisierte Durchführungs- und Auswertungshinweise gegeben
Reliabilität	Zur Analyse der Reliabilität wurde Cronbachs α für Klasse zwei (α = .93, N = 2298) sowie Klasse drei (α = .91, N = 1716) errechnet. Zusätzlich wurde die Splithalf-Reliabilität (korrigiert nach Spearman-Brown) ermittelt: 2. Klasse: r = .95, 3. Klasse: r = .94.

Validität	Durch die Orientierung der Testaufgaben an allen deutschen Mathematiklehrplänen ist die Lehrplanvalidität gegeben. Die Kriteriumsvalidität wurde durch Korrelation mit der Mathematiknote, den Rechenheften von Grube, Barth und Hasselhorn (2002), einem informellen Test zur Ermittlung der Schnelligkeit beim Kopfrechnen (Krajewski, 2003) sowie dem DEMAT 1+ (Krajewski et al., 2002) ermittelt: • Mathematiknote: r = .66 (N = 681) • Rechenhefte (Grube et al., 2002): r = .66 (N = 103) • Schnelligkeit beim Kopfrechnen (Krajewski, 2003): r = .53 (N = 134) • DEMAT 1+ (Krajewski et al., 2002): r = .67 (N = 1185) Darüber hinaus wurde die prognostische Validität anhand der Ergebnisse im DEMAT 3+ (Roick, Gölitz & Hasselhorn, 2004) und DEMAT 4+ (Gölitz, Roick & Hasselhorn, 2006) sowie den Mathematiknoten in Klassenstufe 3 und 4 bestimmt. • DEMAT 3+ (Roick et al., 2004): r = .65 (N = 846) • DEMAT 4+ (Gölitz et al., 2006): r = .67 (N = 741) • Mathematiknote Klasse drei: r = .67 (N = 846) • Mathematiknote Klasse vier: r = .64 (N = 784)
Normen	Es liegen Prozentrang-, T-Werte und T-Wert-Bänder auf Basis der Ergebnisse von insgesamt 4014 Kindern vor. Die Normen sind für den Gesamttest, die Untertests (hier nur Prozentränge) und getrennt nach Geschlecht zu zwei Messzeitpunkten aufgeführt: • letzter Schuljahresmonat Klasse zwei (N = 2298) • erste drei Schuljahresmonate Klasse drei (N = 1716)

Erläuterungen: DEMAT 1+ – Deutscher Mathematiktest für erste Klassen (Krajewski et al., 2002); DEMAT 3+ – Deutscher Mathematiktest für dritte Klassen (Roick et al., 2004); DEMAT 4+ – Deutscher Mathematiktest für vierte Klassen (Gölitz et al., 2006); r – Korrelationskoeffizient; α – Cronbachs α; N – Stichprobenumfang

In der vorliegenden Studie wurden die Kinder nur mit der Version A des DEMAT 2+ (Krajewski et al., 2004) getestet.

6.3.7 Aussagenkatalog zur Einschätzung der CBM sowie deren Einsatz

Wie in Abschnitt 3.2.5 erläutert, ist der wiederholte systematische Einsatz von CBM zur Analyse des Lernverlaufs der Schüler in deutschen Schulen relativ neuartig. Es ist davon auszugehen, dass ein solch standardisiertes Vorgehen für die Lehrkräfte, die innerhalb der vorliegenden Studie teilnahmen, bisher unüblich war. Der ein

Schuljahr überdauernde, etwa monatlich frequentierte Einsatz verschiedener CBM-Tests im Bereich Mathematik für alle Schüler der Klasse stellte für die meisten der Lehrkräfte somit wahrscheinlich einerseits ein Novum sowie andererseits einen Mehraufwand dar. Durch die Ähnlichkeit zu in Schulen üblichen täglichen bzw. zumindest regelmäßigen Übungsphasen sowie die leichte Durchführung und Auswertung der CBM wird jedoch von einer überwiegenden Akzeptanz der Tests bei teilnehmenden Lehrkräften ausgegangen. Diese Annahme wird auch durch Ergebnisse verschiedener Studien (Shapiro et al., 2005; Souvignier et al., 2010) gestützt.

Innerhalb ihrer Studie befragten Shapiro et al. (2005) 18 an der Untersuchung beteiligte Sonderpädagogen und Schulkoordinatoren zum wöchentlichen Einsatz von CBM-Tests in Mathematik mit Schülern, welche als Kinder mit Schwierigkeiten im mathematischen Lernen identifiziert wurden. Dabei konnten die Autoren festhalten: „Teachers reported that conducting the monitoring was feasible within their routine schedule and that using the probes helped to shape their instructional decisions for all students" (Shapiro et al., 2005, S. 28). Ähnliche Ergebnisse liefern eine Lehrerbefragung zur zweiwöchentlichen Testung der Schüler anhand von CBM für die Bereiche Lesen und Mathematik von Souvignier et al. (2010). Diese zeigen, dass 79 % der insgesamt 78 Befragten eine zweiwöchentliche Frequenz der innerhalb der Studie eingesetzten CBM als realistisch einschätzen, 87 % geben dabei an, dass der dafür zu investierende Mehraufwand angemessen sei.

Zur Abschätzung der Ökonomie, der Nützlichkeit und der Zumutbarkeit der in der vorliegenden Arbeit entwickelten CBM wurde ein Aussagenkatalog zur Einschätzung der CBM sowie deren Einsatz im Unterricht konzipiert. Darin sollten die Umfrageteilnehmer 14 Aussagen bewerten, welche sich zum einen eher auf allgemeine

Aspekte von CBM, zum anderen jedoch auch auf den Nutzen und Praktikabilität der Verfahren bezogen. Dazu musste bei allen bis auf einem Item jeweils eine der Antwortkategorien „trifft voll und ganz zu“, „trifft überwiegend zu“, „trifft eher nicht zu“ oder „trifft überhaupt nicht zu“ angekreuzt werden. Dabei wurde bewusst auf eine neutrale Mittelkategorie verzichtet, da diese laut Moosbrugger und Kevala (2012) oft als Ausweichkategorie verwendet wird, „wenn die Testperson den angegebenen Wortlaut als unpassend beurteilt, die Frage nicht verstanden hat, die Antwort verweigert oder aber diese nicht kennt“ (S. 53). Die Bearbeitung der Aussage 11) stellt eine Besonderheit innerhalb des Aussagenkatalogs dar. Die Teilnehmer, welche bei der Einschätzung der Aussage 10) die Kategorie „trifft eher nicht zu“ bzw. „trifft nicht zu“ gewählt haben, sollten zur Beantwortung angeben, in welchem zeitlichen Abstand sie CBM-Testungen als realistisch erachten. Die Bearbeitung der Aussage 13) war nur erforderlich, sofern die Aussage 12) mit einer der Kategorien „trifft eher zu“ bzw. „trifft zu“ bewertet wurde. Nachfolgend sind die Aussagen aufgeführt:

1) Den Einsatz der von der Universität Rostock erarbeiteten CBM zur Dokumentation des Lernverlaufs meiner Schüler finde ich sinnvoll.
2) Das Bearbeiten der CBM hat den Kindern Spaß gemacht.
3) Eine Dokumentation des Lernverlaufs meiner Schüler finde ich sinnvoll.
4) Die Auswertung der Ergebnisse meiner Schüler stellt eine Hilfestellung für meine Unterrichtsgestaltung und für die Planung von Fördermaßnahmen dar.
5) Anhand der Ergebnisse der Schüler gelingt mir die Identifikation schwacher Schüler.
6) Ich nutze die Ergebnisse der CBM (tabellarische oder grafische Auswertung) in Gesprächen mit Eltern.

7) Ich nutze die Ergebnisse der CBM (tabellarische oder grafische Auswertung) in Gesprächen mit Kollegen.
8) Eine zusätzliche Beratung/Weiterbildung zur Bedeutung sowie Durchführung und Auswertung von CBM würde mir bei der Arbeit helfen.
9) Ich brauche zusätzliche Beratung/Information, wie ich identifizierte Risikokinder spezifisch fördern kann.
10) Der monatliche Einsatz von CBM in meiner Klasse ist realistisch.
11) Falls nicht bzw. eher nicht: Den regelmäßigen Einsatz von CBM in einem Abstand von ___ Wochen finde ich realistisch.
12) Der regelmäßige Einsatz von CBM bedeutet einen großen Mehraufwand für mich.
13) Falls dies zutrifft bzw. eher zutrifft: Der Mehraufwand ist angemessen. Ich nehme ihn gern in Kauf.
14) Ich finde die computergestützte Auswertung der CBM nützlich.

Die Aussagen 10) bis 13) beziehen sich dabei auf Gesichtspunkte hinsichtlich der Ökonomie der innerhalb der vorliegender Arbeit konzipierten CBM. Angaben zur Nutzbarkeit sollen aus Bewertungen der Aussagen 1), 3) bis 9) sowie 14) gefolgert werden. Hinweise zur Zumutbarkeit der CBM liefert die Aussage 2).

6.4 Statistische Methoden zur Datenauswertung

6.4.1 Allgemeine Angaben

Ziel der hier vorliegenden Arbeit ist, neben der Konzeption von CBM für den mathematischen Erstunterricht an der Grundschule, deren Analyse hinsichtlich der Erfüllung wissenschaftlicher

Hauptgütekriterien wie der Objektivität, Reliabilität und Validität (Bortz & Döring, 2006; Moosbrugger & Kevala, 2012). Ergänzend dazu soll mittels Analysen zur Änderungssensibilität der entwickelten CBM den in Abschnitt 3.3 benannten Forderungen der von L.S. Fuchs (2004) beschriebenen Forschungsebene 2 entsprochen werden. Neben den Hauptgütekriterien der Objektivität, Reliabilität und Validität (Moosbrugger & Kevala, 2012) soll innerhalb der vorliegenden Untersuchung auch die Erfüllung weiterer Aspekte wie der Ökonomie, der Nützlichkeit und der Zumutbarkeit der CBM betrachtet werden, welche in der Literatur zwar als Nebengütekriterien (Moosbrugger & Kevala, 2012) beschrieben werden, jedoch bei der Anwendung im schulischen Alltag eine wesentliche Rolle spielen. Angaben zur Ökonomie und Nützlichkeit der CBM geben Aufschluss über die Einsetzbarkeit der Verfahren, welche nach L.S. Fuchs (2004) einen wesentlichen Untersuchungsaspekt der Forschungsebene 3 darstellt. Eine Übersicht über die in der vorliegenden Arbeit durchgeführten statistischen Analysen zur Einschätzung der Güte der CBM liefert Tabelle 17.

Tabelle 17: Übersicht über in der vorliegenden Arbeit durchgeführte statistische Analysen zur Einschätzung der Güte der erarbeiteten CBM

Analysen hinsichtlich der		**vorgenommene Analysen**	**Ergebnisse ab Seite**
Reliabilität	Splithalf-Reliabilität	korrelative Analysen	352
	Retest-Reliabilität	korrelative Analysen	356
	Paralleltest-Reliabilität	korrelative Analysen Mittelwertvergleiche	360
Validität	Augenscheinvalidität	deskriptive Ergebnisse der Onlineumfrage	370
	Konstruktvalidität (konvergente und diskriminante Validität)	Multitrait-Multimethod-Matrix korrelative Analysen	371
	Kriteriumsvalidität (Übereinstimmungs- und prognostische Validität)	korrelative Analysen Regressionsanalysen *klassifikatorische Analysen*	387
Änderungssensibilität		korrelative Analysen *Hierarchisch-lineare Modellierung*	412
Ökonomie		deskriptive Ergebnisse der Onlineumfrage	450
Nützlichkeit		deskriptive Ergebnisse der Onlineumfrage	450
Zumutbarkeit		deskriptive Ergebnisse der Onlineumfrage	450

Erläuterungen: Kursiviert – Analysen, welche nachfolgend detaillierter vorgestellt werden

Die Schätzung der Reliabilität der entwickelten CBM, verstanden als „Grad der Genauigkeit, mit dem das geprüfte Merkmal gemessen wird“ (Bortz & Döring, 2006, S. 196), wird im Rahmen der vorliegenden Studie durch verschiedene Schritte vorgenommen. In Substudie 2 wird die Retest-Reliabilität aller entwickelten CBM als Korrelation der Ergebnisse zweier wiederholt eingesetzter CBM-Versionen (Bortz & Döring, 2006) ermittelt (vgl. Abschnitt 7.2.2). Zur Bestimmung der Retest-Reliabilität wird ein- und dasselbe Verfahren wiederholt bei einer Probandengruppe eingesetzt. Die Korrelation der Daten beider Messzeitpunkte ergibt dann den Retest-Reliabilitätskoeffizienten. Voraussetzung dabei ist, dass sich das zu messende Merkmal über die Zeit nicht unsystematisch verändert. Eine hohe Korrelation spricht in diesem Fall dafür, dass die gesamte Stichprobe eine gleiche Entwicklung der Merkmalsausprägung

aufweist (Schermelleh-Engel & Werner, 2012). Hingegen ist bei der Entwicklung mathematischer Kompetenzen in der Grundschule nicht von einer gleichschrittartigen Entwicklung der Schüler auszugehen. Wie in Kapitel 2 bereits beschrieben, kommen Schulanfänger schon mit unterschiedlichsten Lernvoraussetzungen in die Schule. Die innerhalb dieser Studie ermittelten Korrelationen zur Beschreibung der Retest-Reliabilität sind durch die Varianz, welche sich aufgrund des individuellen Lernzuwachses über die Zeit zwischen den Erhebungszeiträumen ergibt, beeinflusst. Dieser Einfluss steigt mit zunehmendem zeitlichem Abstand zwischen den Messzeitpunkten. Insofern ist bei der Interpretation von Retest-Reliabilitätskoeffizienten der Zeitraum zwischen den Messungen zu berücksichtigen.

Weiterhin wird im Rahmen der Substudie 2 die Paralleltest-Reliabilität der Verfahren errechnet (Abschnitt 7.2.3). Die Bestimmung der Paralleltest-Reliabilität ist möglich, wenn für ein Testverfahren sog. parallele Versionen existieren, also solche, die dasselbe Merkmal mit gleicher Genauigkeit messen (Schermelleh-Engel & Werner, 2012). Um Hinweise zur Paralleltest-Reliabilität zu erhalten, werden die als parallel angenommenen Versionen mit derselben Stichprobe durchgeführt. Diese ergibt sich als Korrelation zweier paralleler Verfahren, sofern Mittelwerte und Standardabweichungen der beiden Versionen vergleichbar sind (Bühner, 2006). Daher werden zur differenzierteren Analyse der Parallelität Mittelwertvergleiche der jeweils zeitgleich durchgeführten CBM in der Untersuchungsgruppe Nordvorpommern vorgenommen. Je nach Verteilung der Daten werden dabei t-Tests (bei normalverteilten Daten) bzw. u-Tests (bei nicht-normalverteilten Daten) durchgeführt. Ggf. auftretende statistisch signifikante Unterschiede geben dabei Hinweise über die Existenz eines Effektes, nicht jedoch hinsichtlich dessen Bedeutsamkeit (Bortz & Döring, 2006). Um zu prüfen, welche praktische Relevanz signifikante Mittelwertunterschiede zwischen CBM haben, werden die Effektstärken nach Cohen (Cohens d)

berechnet. Die Effektstärke für abhängige Stichproben d_z wird über die Differenzen zwischen den beiden Messwertereihen nach folgender Formel ermittelt:

$$d_z = \frac{M_d}{\sigma_d},$$

wobei M_d die Differenz der Mittelwerte und σ_d die Differenz der Standardabweichungen darstellt (Cohen, 1988). Folgt man Cohen (1992), so spricht man bei Effektstärken von

- $d < .20$ von keinem bedeutsamen Effekt,
- $.20 \leq d < .50$ von einem kleinen Effekt,
- $.50 \leq d < .80$ von einem mittleren Effekt und
- $d \geq .80$ von einem großen Effekt (S. 157).

Treten bei als parallel angenommenen CBM statistisch signifikante Unterschiede der Mittelwerte auf, wird die Bedeutsamkeit des Effektes mittels Cohens d analysiert. Dabei werden die Verteilungen der Versionen bei einer Effektstärke $d < .50$ als vergleichbar angesehen, sodass die Korrelation zwischen den Ergebnissen als Maß für die Paralleltest-Reliabilität interpretiert wird.

Weiterhin werden in der Substudie 4 Analysen hinsichtlich der Verteilung der Itemparameter aller CBM der Niveaustufe 1 vorgenommen. Dazu werden jeweils die Koeffizienten zur Splithalf-Reliabilität berechnet, um Auskunft über die Homogenität der Verfahren zu erhalten (Abschnitt 7.2.1). Zur Berechnung der Splithalf-Reliabilität (auch Testhalbierungsreliabilität) wird das zu analysierende Verfahren zunächst in zwei Testhälften geteilt. Die Korrelation der Daten der beiden Hälften gibt dann Aufschluss über die Splithalf-Reliabilität (Schermelleh-Engel & Werner, 2012). Zur Halbierung eines Tests werden in der Literatur verschiedene

Vorgehensweisen (Teilung nach der Hälfte der Itemanzahl, Zufallsauswahl, Odd-Even-Methode, Methode der Itemzwillinge etc.) beschrieben (Bortz & Döring, 2006; Schermelleh-Engel & Werner, 2012), bei Speedtests ist jedoch die sog. Zeitpartitionierungsmethode (Schermelleh-Engel & Werner, 2012) üblich. Da bei einer Testhalbierung jedoch nur die Hälfte der Items Berücksichtigung bei der Analyse finden, wird die Reliabilität, die in Relation zur Gesamtitemanzahl zunimmt, unterschätzt (Bortz & Döring, 2006; Schermelleh-Engel & Werner, 2012). Um dies zu kompensieren, werden bei den Analysen in nachfolgenden Abschnitten Korrekturrechnungen der Reliabilität auf Basis der Spearman-Brown-Formel vorgenommen:

$$r_{korr} = \frac{2 \cdot r_{unkorr}}{1+r_{unkorr}},$$

wobei r_{korr} den korrigierten und r_{unkorr} den unkorrigierten Reliabilitätskoeffizienten angibt (vgl. u. a. Schermelleh-Engel & Werner, 2012).

Auf eine differenzierte Analyse der CBM der Niveaustufe 2 auf Itemebene wird in der vorliegenden Arbeit verzichtet, da davon ausgegangen wird, dass die Homogenität der CBM, aufgrund der in Abschnitt 3.2.8.1 beschriebenen Zusammenstellungsweise ihrer Items durch eine stratifizierte Anordnung, hinreichend gewährleistet ist. Die Reliabilitätsanalysen für die Niveaustufe 2 beschränken sich demnach auf Untersuchungen hinsichtlich der Paralleltest- und Retest-Reliabilität.

Zur Interpretation der errechneten Korrelationskoeffizienten wird in vorliegender Studie auf die von Bühl (2006) beschriebene Einteilung zurückgegriffen, diese ist in Tabelle 18 dargestellt.

Tabelle 18: Interpretation des Korrelationskoeffizienten r nach Bühl (2006, S. 342)

Korrelationskoeffizient	Interpretation nach Bühl (2006)
r > .9	sehr hohe Korrelation
.9 ≥ r >.7	hohe Korrelation
.7 ≥ r >.5	mittlere Korrelation
.5 ≥ r >.2	geringe Korrelation
r ≤ .2	sehr geringe Korrelation

Zur Überprüfung der Validität der konzipierten CBM sollen verschiedene Aspekte untersucht werden. Den Zusammenhang der verschiedenen Validitätsaspekte liefert Abbildung 35. Neben Gesichtspunkten zur Inhaltsvalidität sollen die CBM demnach auch hinsichtlich ihrer Augenschein-, Konstrukt- und Kriteriumsvalidität geprüft werden. Dabei müssen wiederum verschiedene Teilaspekte analysiert werden. So ergibt sich die Konstruktvalidität als Zusammenspiel von der konvergenten und der diskriminanten Validität, die Kriteriumsvalidität wird durch Angaben zur Übereinstimmungs- und prognostischen Validität bestimmt (Moosbrugger & Kelava, 2012).

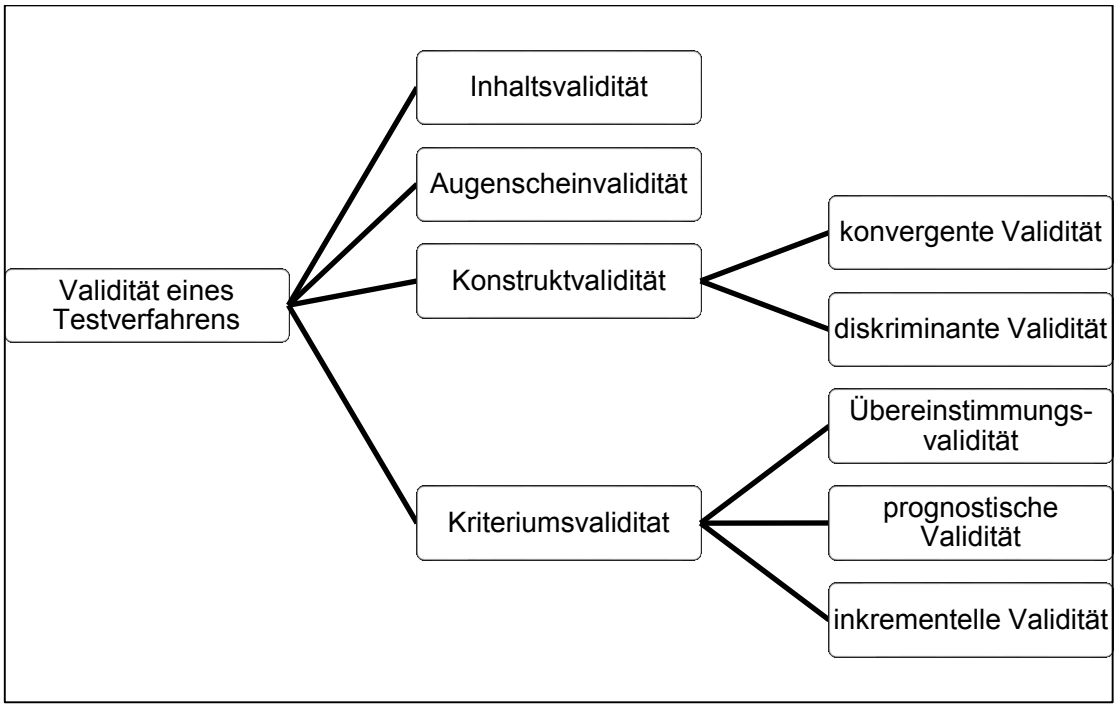


Abbildung 35: Teilaspekte der Validität eines Testverfahrens (angelehnt an: Bühner, 2006; Moosbrugger & Kelava, 2012)

Zur Bestimmung der Inhaltsvalidität wird überprüft, inwieweit die vorgenommene Operationalisierung des zu messenden Konstrukts dessen relevante Inhalte tatsächlich widerspiegelt. Die Überprüfung auf Inhaltsvalidität wird in der Regel nicht rechnerisch vollzogen, sie „erfordert vielmehr eine Betrachtung der Iteminhalte, wobei der Bezug zwischen diesen Inhalten und dem zu erfassenden Merkmal durch theoretische Argumente zu stützen ist“ (Hartig, Frey & Jude, 2012, S. 149). Für eine solche kritische Beurteilung der Testinhalte ist es notwendig, das zu erfassende Merkmal hinreichend zu definieren. Anschließend muss geprüft werden, ob und inwieweit die Aussagen des Tests es ermöglichen, das Testergebnis über die verwendeten Items zu generalisieren. Dies soll durch theoretische Abwägungen geschehen. Angaben zur Augenscheinvalidität können der zuvor erwähnten Onlinebefragung des Lehrpersonals teilnehmender Klassen entnommen werden. Dabei beziehen sich einzelne Fragen auf die Bewertung der innerhalb der vorliegenden Arbeit konzipierten CBM. Eine positive Einschätzung der CBM seitens der Lehrkräfte spricht für die Augenscheinvalidität der Verfahren.

Zur Bestimmung der Konstruktvalidität soll in einem ersten Schritt die Interkorrelation der einzelnen CBM zu den verschiedenen Messzeitpunkten analysiert werden. Darüber hinaus sollen die konvergente sowie die diskriminante Validität der CBM als Teilaspekte der Konstruktvalidität ermittelt werden (Schermelleh-Engel & Schweizer, 2012). Konvergente Validität liegt vor, wenn ein hoher Zusammenhang zwischen den Ergebnissen zweier Testverfahren, welche gleiche Konstrukte erfassen, besteht (Schermelleh-Engel & Schweizer, 2012). Zur Prüfung der konvergenten Validität wird die Korrelation zu den Ergebnissen eines etablierten Tests ermittelt, welcher ein gleiches oder ähnliches Merkmal misst. Konvergente Validität ist dann gegeben, wenn die Ergebnisse beider Testverfahren hoch miteinander korrelieren. Diskriminante oder auch divergente Validität liegt vor, wenn die

Ergebnisse von zwei Testverfahren, welche voneinander unabhängige Konstrukte erfassen niedrig miteinander korrelieren (Schermelleh-Engel & Schweizer, 2012). Bspw. soll ein Rechentest ein anderes Konstrukt erfassen als z. B. ein Lesetest. Niedrige Korrelationen weisen in diesem Fall darauf hin, dass es keinen bzw. nur einen geringen Zusammenhang der Ergebnisse in beiden Verfahren gibt, sie also unterschiedliche Konstrukte messen. Eine Methode die konvergente und die diskriminante Validität zu untersuchen stellt die auf Campbell und Fiske (1959) zurückgehende Multitrait-Multimethod-Methode dar.

> *„Diese Validierungsmethode erfordert, dass mehrere Konstrukte (Multitrait) durch mehrere Erhebungsmethoden (Multimethod) erfasst werden. Eine systematische, regelgeleitete Analyse der wechselseitigen Beziehungen zwischen Konstrukten und Methoden erlaubt es, die Höhe der Konstruktvalidität abzuschätzen“* (Bortz & Döring, 2006, S. 203)

Dabei werden das mathematische und das schriftsprachliche Vorwissen zum Beginn der Klasse eins bzw. die mathematischen und die Lesekompetenzen zum Ende der Klasse eins als verschiedene Konstrukte (Multitrait) einerseits anhand von CBM und andererseits anhand weiterer standardisierter Testverfahren ohne zeitliche Begrenzung (Multimethod) erhoben. Zur Bestimmung der Konstruktvalidität werden zunächst die Interkorrelationen aller dieser Verfahren ermittelt und in einer Matrix, ähnlich wie in Abbildung 36, zusammengetragen. Innerhalb der Matrix lassen sich verschiedene Teilmatrizen finden, welche jeweils Aufschluss über die konvergente und die diskriminante Validität der Verfahren geben. Dabei sind r_1 und r_6 die Korrelationskoeffizienten der sog. Monotrait-Heteromethod-Blöcke (mittleres grau; Bortz & Döring, 2009; Schermelleh-Engel & Schweizer, 2012). Hier sind gleiche bzw. ähnliche Konstrukte mit verschiedenen Methoden erhoben worden. Eine hohe Korrelation weist hier auf eine hohe konvergente Validität hin. Die Koeffizienten

r_2 und r_5 weisen die sog. Heterotrait-Monomethod-Blöcke (dunkles grau; Bortz & Döring, 2009; Schermelleh-Engel & Schweizer, 2012) aus. In diesen Feldern sind die Korrelationen verschiedener Konstrukte, die mit gleichen Methoden erhoben worden notiert, entsprechend geben geringe Koeffizienten Hinweise auf diskriminate Validität. r_3 und r_4 sind in den sog. Heterotrait-Heteromethod-Blöcken (helles grau; Bortz & Döring, 2009; Schermelleh-Engel & Schweizer, 2012) notiert. Sie bilden den Zusammenhang verschiedener Konstrukte, erhoben mit verschiedenen Methoden, ab. „Hier werden die geringsten Korrelationen erwartet, da weder methodische noch inhaltliche Übereinstimmungen vorliegen“ (Bortz & Döring, 2009, S. 205).

		CBM		standardisiertes Verfahren	
		MK	LK	MK	LK
CBM	MK	1			
	LK	r_1	1		
standardisiertes Verfahren	MK	r_2	r_4	1	
	LK	r_3	r_5	r_6	1

Erläuterungen: MK – mathematische Kompetenzen, LK - Lesekompetenzen

Abbildung 36: schematische Darstellung der Multitrait-Multimethod-Matrix zweier Traits (mathematische und Lesekompetenz) erhoben mittels zweier Methods (CBM und weiteres standardisiertes Verfahren)

Von wesentlichem Interesse sind dabei die Korrelationskoeffizienten r_1, r_2 und r_3, da sich diese direkt auf das zu untersuchende CBM beziehen. Anhand von vier von Campbell und Fiske (1959) festgelegten Kriterien kann die Matrix hinsichtlich der Konstruktvalidität untersucht werden:

1. die Monotrait-Heteromethod-Korrelationen sind größer als Null: r_2, $r_5 > 0$ (konvergente Vaildität)

2. die Heterotrait-Monomethod-Korrelationen sind kleiner als die Monotrait-Heteromethod-Korrelationen: r_1, r_6 < r_2, r_5 (diskriminante Validität)
3. die Heterotrait-Heteromethod-Korrelationen sind kleiner als die Monotrait-Heteromethod-Korrelationen und Heterotrait-Monomethod-Korrelationen: r_3, r_4 < r_2, r_5 r_1, r_6 (diskriminante Validität)
4. Konstruktvalidität ist gegeben, wenn konvergente und diskriminante Validität vorliegen (siehe 1. – 3.)

Gegebene Kriteriumsvalidität eines Testverfahrens deutet darauf hin, dass der Schluss von einem Ergebnis des zu untersuchenden Tests auf ein praktisch relevantes Außenkriterium zulässig ist. Empirisch wird das Vorliegen der Konstruktvalidität anhand von Zusammenhangsanalysen zwischen dem Testwert und den Ergebnissen des Außenkriteriums bestimmt (Hartig et al., 2012). Je größer die somit ermittelten Zusammenhänge sind, umso sicherer ist der Nachweis der Kriteriumsvalidität. „Dabei können externe Kriterien gewählt werden, die zeitlich parallel existieren (Übereinstimmungsvalidität), oder solche, die eine Prognose über die zukünftige Ausprägung des Merkmals beinhalten (prognostische Validität)“ (Hartig et al., 2012, S. 157). Die Bestimmung der Kriteriumsvalidität, als Zusammenspiel von Übereinstimmungs-, prognostischer- und inkrementeller Validität (Moosbrugger & Kevala, 2012; Hartig et al., 2012), geschieht in der vorliegenden Arbeit in mehreren Schritten. Zum einen geben die Analysen zur konvergenten Validität ebenso Aufschluss über die Übereinstimmungsvalidität, zum anderen soll die prognostische Validität der einzelnen CBM ermittelt werden. Angaben zur prognostischen Validität, als ein besonders wichtiges Kriterium für Screening-Verfahren (Marx, 1992; Marx & Lenhard, 2010), sollen anhand klassifikatorischer Analysen ermöglicht werden. Da es sich hierbei um Analysen handelt, welche nicht unbedingt bei jedem veröffentlichtem Testverfahren aufgeführt

werden (Marx & Lenhard, 2010), wird das dazu nötige Vorgehen in Abschnitt 6.4.2 detaillierter beschrieben.

Die inkrementelle Validität beschreibt das „Ausmaß, in dem die Vorhersage des praktisch relevanten externen Kriteriums verbessert werden kann, wenn zusätzliche Testaufgaben oder Testskalen zu den bereits existierenden hinzugenommen werden“ (Hartig et al., 2012, S. 158). Die inkrementelle Validität eines jeden hier vorgestellten CBM soll mithilfe der binär-logistischen Regression unter Einbeziehung der Schülerergebnisse weiterer innerhalb der vorliegenden Arbeit konzipierten CBM zum gleichen Messzeitpunkt geprüft werden. D. h., es wird untersucht, inwieweit sich die Prognose schulischer Minderleistungen im Bereich Mathematik zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) bzw. zum Ende der Klasse zwei (PR ≤ 15 im DEMAT 2+; Krajewski et al., 2004) durch die Berücksichtigung der Daten aller eingesetzter CBM der Niveaustufe 1 bzw. 2 in Kombination verbessert. Der Zuwachs an aufgeklärter Varianz bei der Vorhersage des externen Kriteriums gibt dann Aufschluss über den prognostischen Zugewinn (Hartig et al., 2012). Als Kriterium bzw. abhängige Variable wird die Zugehörigkeit zur Gruppe von Risikokindern mit Minderleistungen im Rechnen in die Regressionsanalyse einbezogen. Dabei gehören Kinder, welche zum Ende der Klassenstufe 1 bzw. 2 nur geringe Leistungen im Rechnen (operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 1+ bzw. DEMAT 2+) haben, der Risikogruppe an und werden als 1 codiert. Schüler mit Leistungsdaten, welche einem Prozentrang über 15 entsprechen, zählen zur Gruppe von Kindern ohne schulische Minderleistungen (PR > 15 im DEMAT 1+ bzw. DEMAT 2+) und werden als 0 codiert. Da der Einsatz von CBM eine möglichst frühzeitige Prognose schulischer Minderleistungen ermöglichen soll, werden in den Analysen diejenigen CBM-Daten als unabhängige Variablen berücksichtigt, welche zum ersten Messzeitpunkt in der fünften Schulwoche (Niveaustufe 1) bzw. zum

fünften Messzeitpunkt in der 21. Schulwoche (Niveaustufe 2) des ersten Schuljahres erhoben wurden.

Inwieweit die konzipierten Verfahren die Veränderungen in den Ausprägungen der arithmetischen Kenntnisse der Schüler im Laufe des Schuljahres der ersten Klasse erfassen, soll in einem mehrstufigen Analyseprozess ermittelt werden. In einem ersten Schritt werden die zuvor ermittelten Koeffizienten hinsichtlich der Retest- und der Paralleltest-Reliabilität verglichen. Dabei werden die von Klauer (2011) konstatierten Kriterien geprüft.

1. Ein änderungssensibler Test weist nur mittlere Retest- bzw. Paralleltest-Reliabilitäten auf.
2. Die Korrelationen zwischen änderungssensiblen Tests werden kleiner, je größer der Zeitraum zwischen den Erhebungszeitpunkten ist.
3. Die Splithalf-Reliabilität eines änderungssensiblen Tests ist höher als die Retest- bzw. Paralleltest-Reliabilitäten.

In einem zweiten Schritt sollen Mittelwertvergleiche in den einzelnen innerhalb der vorliegenden Arbeit konzipierten CBM zwischen den Untersuchungsgruppen 1 (Rügen) und 2 (Nordvorpommern) über die verschiedenen Messzeitpunkte vorgenommen werden. Dies geschieht mittels Hierarchisch-linearer Modellierung (Bryk & Raudenbush, 1992). Dabei wird geprüft, inwieweit es statistisch signifikante Unterschiede in den zu verschiedenen Messzeitpunkten erreichten mittleren CBM-Ergebnissen gibt. Da es sich hierbei ebenso um eine sehr spezifische Methode handelt, soll diese später, in Abschnitt 6.4.3, genauer beschrieben werden.

6.4.2 Klassifikatorische Analysen

Analysen hinsichtlich der prognostischen bzw. Vorhersagevalidität gehen der Frage nach, inwieweit Ergebnisse eines durchgeführten

Tests mit einem sich später einstellenden Merkmal übereinstimmen. Ein Schultest, welcher valide Prognosen über zukünftige Leistungsentwicklungen eines Schülers ermöglicht und somit auch frühzeitig auf Entwicklungsprobleme hindeuten kann, ist im Rahmen schulischer Präventionsbemühungen besonders wertvoll (Marx & Lenhard, 2010). Damit kommt auch der Ermittlung der prognostischen Validität eines Screeningverfahrens zur Identifikation von Kindern mit Problemen in bestimmten Lernbereichen eine große Bedeutung zu. Bezogen auf Testverfahren zur Identifikation von Kindern mit erhöhtem Risiko eine Rechenschwäche zu entwickeln, führen Hasselhorn, Roick und Gölitz (2005) verschiedene Möglichkeiten an, die prognostische Validität zu ermittelten. Zum einen:

> *„„[…] indem das Testverfahren an einer größeren Stichprobe durchgeführt und zu einem späteren Zeitpunkt das Ausmaß der unabhängig bewerteten Störungssymptomatik quantifiziert wird: Eine hohe Korrelation zwischen Testwert und später auftretender Störung gilt dann als Beleg für die gute prognostische Validität des Testverfahrens“* (S. 190f.).

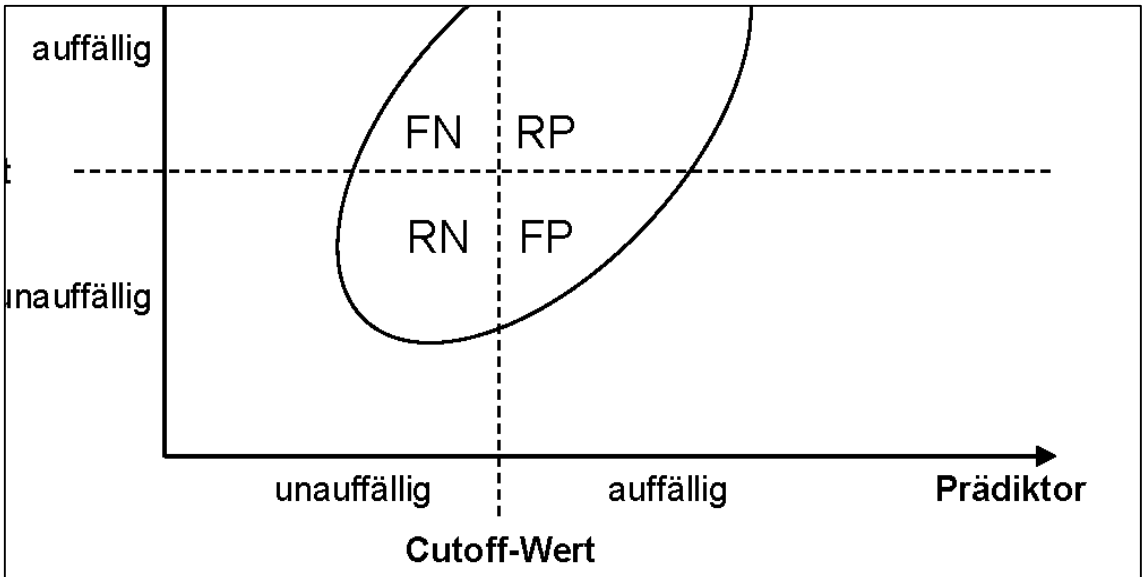
Als einen weiteren Weg mit weitaus differenzierterem Informationsgehalt beschreiben die Autoren die Ermittlung des Zuordnungsfehlers bei der Klassifikation anhand der Ergebnisse eines Tests. Im einfachsten Fall weisen Prädiktor (also die erzielten Ergebnisse innerhalb des zu untersuchenden Tests) und Kriterium eine zweidimensionale Ausprägung auf, bspw. „auffällig“ bzw. „unauffällig“. Abhängig von den gewählten Cutoff-Werten des Testverfahrens (Prädiktor) und des Kriteriums ergeben sich damit vier mögliche Klassifikationsvarianten, welche in Abbildung 37 dargestellt sind und von denen zwei Klassifikationen als Fehlzuordnung anzusehen sind. Es wird demnach zum einen der Anteil an Kindern, welche durch den Test zwar als auffällig klassifiziert wurden, in einer späteren Messung des interessierenden Kriteriums jedoch eine

unauffällige Merkmalsausprägung aufweisen, bzw. zum anderen der Anteil an Kindern, welche durch den Test zwar als unauffällig klassifiziert wurden, in einer späteren Messung des interessierenden Kriteriums jedoch eine auffällige Merkmalsausprägung aufweisen, analysiert. Erstere Fehlklassifikationen werden als Fehler erster Art bzw. α-Fehler, zweitere als Fehler zweiter Art bzw. β-Fehler bezeichnet (Hasselhorn et al., 2005). Je nachdem, zu welchem Zweck das Testverfahren eingesetzt wird, ziehen diese Fehlklassifikation mehr oder weniger gravierende Folgen für die falsch klassifizierten Kinder nach sich (Hartke & Koch, 2008; Krapp, 1993; Suchodoletz, 2005). Als falsch positiv identifizierte Kinder werden zu Unrecht Interventionsmaßnahmen unterzogen, wodurch verfügbare Ressourcen unnötig verbraucht werden. Gleichzeitig sind diese Kinder einem erhöhten Risiko von Etikettier- oder Stigmatisierungsprozessen, unter Umständen selektiven Maßnahmen ausgesetzt (Hasselhorn et al., 2005). Für falsch negativ klassifizierte Kinder hingegen, wird ein notwendiger Förderbedarf nicht angezeigt, wodurch keine zeitnahe Förderung einsetzt. In Kapitel 1 wurde die gravierende Bedeutung von Wissenslücken auf die gesamte schulische Laufbahn aufgezeigt.

Auch wenn falsch positiv klassifizierte Kinder unnötige Ressourcen für Unterricht bzw. Förderung beanspruchen, erscheinen im Rahmen eines RtI-Ansatzes (vgl. Abschnitt 3.1.4) jedoch eher die Fehlklassifikationen zweiter Art als die gravierenderen (Irvin, Park, Anderson, Alonzo & Tindal, 2011). Während eine hohe Chance besteht, dass Kinder mit einem möglicherweise falsch zugewiesenen Risikostatus durch die intensivierten wiederholten Leistungserfassungen zeitnah korrekt reklassifiziert werden, erhalten Kinder, welche fälschlicherweise als nicht auffällig eingeordnet werden, nicht die entsprechende Förderung, die sie eigentlich benötigen.

Zur praktischen Bestimmung der Vorhersagegüte eines Tests werden die Zellbesetzungen der vier möglichen Klassifikationen sowie die Beziehungen von den Zellhäufigkeiten zu den Randsummen genauer betrachtet. Es ergeben sich folgende Kennwerte, welche die Güte der Klassifikationsfähigkeit eines Testverfahrens beschreiben (Hasselhorn et al., 2005; Marx, 1992; Marx & Lenhard, 2010):

- Die Sensitivität eines Verfahrens gibt an, wie viele der Probanden mit Auffälligkeiten in der Kriteriumsleistung auf Basis eines Screenings als auffällig eingestuft wurden.
- Die Spezifität eines Verfahrens gibt an, wie viele der Probanden ohne Auffälligkeiten in der Kriteriumsleistung auf Basis eines Screenings als unauffällig eingestuft wurden.
- Der positive prädiktive Wert (Prädiktortrefferquote) gibt an, mit welcher Wahrscheinlichkeit ein anhand des Prädiktors als auffällig klassifiziertes Kind später auch tatsächlich innerhalb der Kriteriumsmessung auffällig wird.
- Der negative prädiktive Wert gibt an, mit welcher Wahrscheinlichkeit ein anhand des Prädiktors als unauffällig klassifiziertes Kind später auch tatsächlich innerhalb der Kriteriumsmessung unauffällig bleibt.
- Die Trefferquote (TQ) gibt den Anteil der korrekten Klassifikationen an.
- Die maximale Trefferquote (max.TQ) gibt die Obergrenze der Trefferquote (TQ) an.
- Die Zufallstrefferquote (ZTQ) gibt den Anteil richtiger Vorhersagen an, welche bei einer zufälligen Klassifikation zu erwarten sind.
- Zusätzlich wird der **r**elative **A**nstieg der **T**refferquote gegenüber einer **Z**ufallstrefferquote (RATZ-Index) unter Berücksichtigung der maximalen Trefferquote der Verfahren ermittelt.



en: FN – Anteil der falsch-negativ klassifizierten Untersuchungspe
(β-Fehler); RP – Anteil der richtig-positiv klassifi
Untersuchungspersonen; RN – Anteil der richtig-negativ klassifi
Untersuchungspersonen; FP – Anteil der falsch-positiv klassifi
Untersuchungspersonen (α-Fehler)

		Kriterium		gesamt
		auffällig	unauffällig	
pr	auffällig	Richtig-Positive (RP)	Falsch-Positive (FP)	RP + FP
	unauffällig	Falsch-Negative (FN)	Richtig-Negative (RN)	FN + RN
gesamt		RP + FN	FP + RN	RP + FP + FN + RN
		$\frac{RP}{RP + FN}$		
		$\frac{RN}{FP + RN}$		
rädiktiver Wert efferquote)		$\frac{RP}{RP + FP}$		
prädiktiver Wert		$\frac{FN}{FN + RN}$		
te (TQ)		$\frac{RP + RN}{}$		

Abbildung 37: Hypothetische Klassifikation anhand eines Screenings (Prädiktor) sowie eines Kriteriums untersuchter Personen in Anlehnung an Marx und Lenhard (2010)

Wesentliche Kennwerte, welche Hinweise zur Vorhersagegüte eines Tests liefern sind die Sensitivität, die Spezifität, der negative sowie der positive prädiktive Wert und der RATZ-Index (Hasselhorn et al., 2005; Marx & Lenhard, 2010). Während alle anderen Kennwerte einzeln nur bedingt aussagekräftig sind, ist der RATZ-Index als kombinierter Wert auch für sich gehaltvoll und gilt somit als der bedeutendste Kennwert, wenn es darum geht, die prognostische Validität eines testdiagnostischen Verfahrens zu beurteilen (Hasselhorn et al., 2005). Er beschreibt den Gewinn eines Tests im Vergleich zu einer zufälligen Klassifikation. In der Regel liegen die Werte zwischen 0 und 1, können jedoch auch negativ ausfallen. Üblich ist die Angabe des RATZ-Indexes in Prozent. Laut Jansen et al. (1999) bedeuten Werte

- kleiner als 0, dass eine zufällig vorgenommene Klassifikation bessere Ergebnisse als der Test erbringt.
- von 0, dass das Testverfahren keine bessere Vorhersage als eine zufällige Klassifikation liefern kann.
- ab 34 %, dass ein Testverfahren eine leichte Verbesserung gegenüber einer zufälligen Zuteilung und somit eine gute Klassifikationsleistung aufweist.
- ab 66 %, dass ein Testverfahren eine deutliche Verbesserung gegenüber einer zufälligen Zuteilung und somit eine sehr gute Klassifikationsleistung aufweist.

In der vorliegenden Arbeit sollen Angaben zur Sensitivität, zur Spezifität, zu positiven und negativen prädiktiven Werten, zur Trefferquote, zur maximalen Trefferquote, zur Zufallstrefferquote sowie zum RATZ-Index für jedes CBM über die Messzeitpunkte ermittelt werden. Dabei wird für jedes CBM ein Cutoff-Wert entsprechend einem Prozentrang kleiner bzw. gleich 25 angelegt. Kinder mit einer Leistung entsprechend dieses Cutoff-Kriteriums erzielen, werden damit einer Schülergruppe zugeordnet, welche dem

Risiko unterliegt, zukünftig schwache Mathematikleistungen aufzuweisen. Zur Erfassung einer Minderleistung im mathematischen Bereich wird als Kriterium ein Prozentrang kleiner bzw. gleich 15 im Deutschen Mathematiktest für erste Klassen (DEMAT 1+; Krajewski et al., 2002) zum Ende der Klasse eins bzw. im Deutschen Mathematiktest für zweite Klassen (DEMAT 2+; Krajewski et al., 2004) zum Ende der Klasse zwei angenommen. Schüler mit Leistungen entsprechend unterhalb oder gleich einem Prozentrang 15 weisen demnach Minderleistungen im Fach Mathematik auf. Es wird dann untersucht, inwieweit es anhand der CBM jeweils über die verschiedenen Messzeitpunkte hinweg gelingt, die Zugehörigkeit zu den oben benannten Gruppen (mit oder ohne mathematische Minderleistung) zu prognostizieren. Zur Ermittlung der Klassifikationsqualität der CBM wird jeweils für den Anteil von Kindern, welche zum entsprechenden Messzeitpunkt Ergebnisse im unteren Quartil erzielten, geprüft, inwieweit diese zur Gruppe der Kinder mit Minderleistungen im Bereich Rechnen (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) zum Ende der Klasse eins gehörten. Umgekehrt wird ebenso geprüft, inwieweit der Anteil der Kinder, welche durch die CBM als nicht gefährdet eingestuft wurden, zum Ende der Klasse eins auch tatsächlich keine Minderleistungen im Rechnen (PR > 15 im DEMAT 1+; Krajewski et al., 2002) aufwiesen.

6.4.3 Hierarchisch-lineare Modellierung (HLM)

Zur Schätzung von Verlaufsparametern der CBM-Daten einer Schülergruppe schlagen Shapiro et al. (2005) das statistische Verfahren der Regression auf der Grundlage der zu verschiedenen Testzeitpunkten erhobenen Schülerdaten vor. Auch Walter (2011) nutzt die Modellierung linearer Trends der CBM-Daten von Schülern mittels Regressionsanalyse zur Bestimmung ihres Lernverlaufs. Dieses Vorgehen erlaubt eine Annährung einer Geraden durch die erhobenen Datenpunkte, sodass CBM-Werte vor, zwischen und nach

den Testungen prognostiziert werden können. Weiterhin können Angaben über das mittlere Ausgangsniveau der Schüler zum ersten Testzeitpunkt sowie mittlere Leistungsanstiege über die Testzeitpunkte getroffen werden. Zur Untersuchung der Änderungssensibilität schlagen Shin et al. (2004) jedoch das statistische Vorgehen der Hierarchisch-linearen Modellierung (HLM) vor, da dies hinsichtlich der Analyse von längsschnittlich erhobenen Daten gegenüber weiteren möglichen statistischen Vorgehensweisen wie der ANOVA mit Messwiederholung oder Strukturgleichungsmodellen eine Reihe von Vorteilen bietet. So werden für die HLM nicht unbedingt vollständige Fälle zu allen Messzeitpunkten vorausgesetzt, auch muss die Datenerhebung nicht mit allen Probanden zum gleichen Zeitpunkt durchgeführt werden, was bei größeren Stichprobengrößen oftmals aus organisatorischen Gründen vorteilhaft ist. Darüber hinaus erlaubt die HLM, im Gegensatz zu Strukturgleichungsmodellen, auch Parameterschätzungen anhand relativ kleiner Stichproben (Bryk & Raudenbush, 1992).

Bei der HLM handelt sich um ein regressionsbasiertes statistisches Verfahren, welches die Effekte geschachtelter, hierarchisch geordneter Systeme analysiert. So können bspw. Schülerdaten unter gleichzeitiger Berücksichtigung der Spezifika einer Klasse, in der die Schüler geschachtelt sind, untersucht werden. Dieses Vorgehen ermöglicht Vergleiche von Personen- bzw. Gruppen (between subjects) unter Berücksichtigung der personenspezifischen Variabilität (within subjects) und somit eine genauere Modellierung realer Zusammenhänge. Die Hierarchisch-lineare Modellierung ist eine statistische Methode, welche sich darüber hinaus ebenfalls zur Untersuchung längsschnittlich erhobener Daten, also zur Analyse von Veränderungen, eignet (Bryk & Raudenbush, 1992). Dabei werden die Messzeitpunkte als innerhalb der Person geschachtelte Systeme aufgefasst. Diese Strukturierung der Daten am Beispiel der

Lernverlaufsdokumentation mittels CBM ist in Abbildung 38 grafisch dargestellt. Dabei wird zum einen die interindividuelle Variabilität (zwischen verschiedenen Schülern, Ebene 2) der Testergebnisse als auch die intraindividuelle Variabilität (innerhalb des Schülers über die Schulwochen, Ebene 1) der Testergebnisse beachtet. Anhand Hierarchisch-linearer Modellierung werden somit individuelle Spezifika eines Kindes, wie bspw. die Zugehörigkeit zu einer Gruppe von Kindern mit Risiken im mathematischen Vorwissen, als Prädiktoren auf der Ebene 2 in das Analysemodell einbezogen.

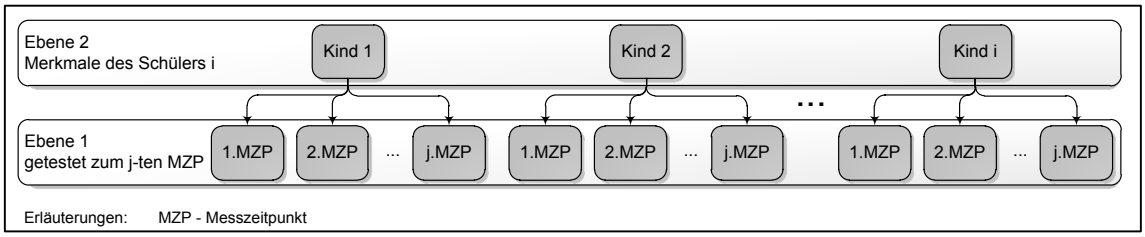


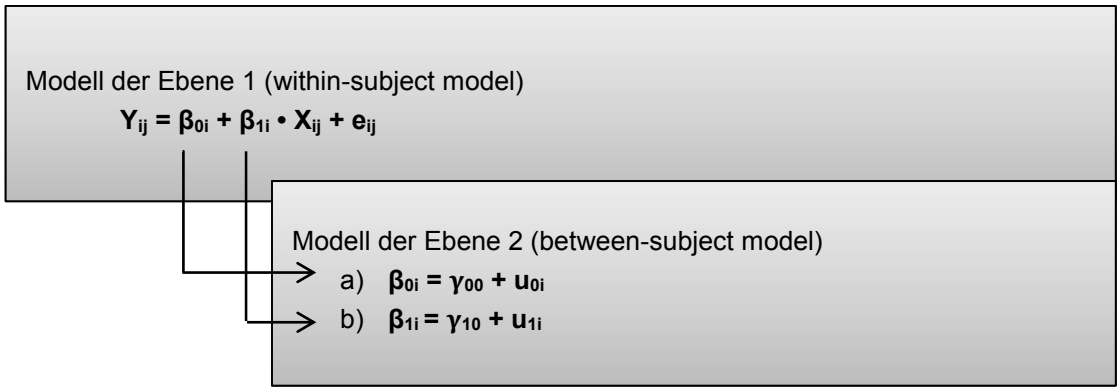
Abbildung 38: Datenstruktur einer wiederholten Messung (z. B. mittels CBM) im Rahmen des hierarchisch-linearen Mehrebenenmodells (entnommen und modifiziert aus Langer, 2009, S. 223)

Ähnlich wie bei der einfachen Regression werden bei der HLM die Einflüsse von Prädiktoren X_{ij} (hier die Schulwoche, in welcher CBM-Daten erhoben werden) sowie eines gewissen Fehlers e_{ij} auf das Kriterium Y_{ij} (hier das Ergebnis im CBM) anhand einer Regressionsgleichung definiert. Bezogen auf den Sachverhalt der vorliegenden Arbeit bedeutet dies:

Von jedem Kind i liegen zu j Zeitpunkten CBM-Daten vor, welche mit Y_{ij} bezeichnet werden. Die Variable X_{ij} gibt die Schulwoche j an, in welcher das Kind i anhand des CBM getestet wurde und geht somit als Prädiktor auf Ebene 1 in das Modell. Es ergibt sich eine Gleichung der Regressionsgeraden auf Ebene 1 für den Lernverlauf der in Abbildung 39 dargestellten Form. Der Geradenverlauf wird durch den Steigungskoeffizient (β_{1i}, slope) und die Regressionskonstante (β_{0i}, intercept) beschrieben. Wobei der Steigungskoeffizient der Regressionsgeraden als ein Maß für den monatlichen

Leistungszuwachs (bzw. Leistungsabfall oder Stagnation) eines Kindes i und die Regressionskonstante als der Ausgangswert eines Kindes i zum ersten Testzeitpunkt aufgefasst wird. Die Indexierung der Regressionsparameter mit i (für den Schüler i) weist darauf hin, dass die Parameter des dargestellten Modells der Ebene 1 über die Level-2-Einheiten, im vorliegenden Fall also die Schüler, hinweg variieren und somit spezifisch für den Schüler i sein können. Der Steigungskoeffizient β_{1i} und die Regressionskonstante β_{0i} werden dann wiederum als Kriterium bzw. outcome durch jeweils eigene Regressionsgleichungen auf Ebene 2 beschrieben.

In der vorliegenden Arbeit werden zur Analyse der Änderungssensibilität der Verfahren jeweils zwei Modelle berechnet. Zum einen wird innerhalb eines sog. *unkonditionierten Modells* (Ditton, 1998) geprüft, ob es signifikante Unterschiede in den Ausgangsniveaus und den linearen Leistungsanstiegen der CBM zwischen den Schülern gibt. „Hierbei werden zwar die Regressionskonstanten und [Hervorh. des Verfassers] -steigungen als Zufallskoeffizienten behandelt, es werden aber keine Prädiktoren aufgenommen, um die Variabilität dieser Koeffizienten vorherzusagen“ (Ditton, 1998, S. 68). Die entsprechenden Regressionsgleichungen der Ebenen 1 und 2 sind in Abbildung 39 dargestellt. Das unkonditionierte Modell an sich liefert allein allerdings nur wenig aussagekräftige Informationen, da unklar bleibt, wodurch eine mögliche Variabilität hervorgerufen wird. Auf der Grundlage der Ergebnisse des unkonditionierten Modells kann jedoch entschieden werden, welche Prädiktoren aufgrund ihrer Varianzanteile als feste, also nicht über die Schüler hinweg variierende Effekte angenommen werden können.



Y_{ij}	CBM-Ergebnis von Schüler i zur Schulwoche j (Kriterium)
$_i$	Index für den Schüler
$_j$	Index für die Schulwoche
β_{0i}	personenspezifischer CBM-Wert von Schüler i zum ersten Testzeitpunkt (intercept)
β_{1i}	personenspezifischer Leistungsanstieg (slope) von Schüler i von einem Testzeitpunkt zum nächsten
X_{ij}	Angabe der Schulwoche (Prädiktor auf Ebene 1)
e_{ij}	personenspezifischer Fehlerterm von Schüler i zur Schulwoche j
γ_{00}	mittleres CBM-Ergebnis aller Schüler zum ersten Testzeitpunkt
u_{0i}	individuelle Abweichung des Schülers i von der über die Schüler gemittelten Regressionskonstanten γ_{00}
γ_{10}	mittlerer Anstieg der CBM-Werte aller Schüler von einem Testzeitpunkt zum nächsten
u_{1i}	individuelle Abweichung des Schülers i von der über die Schüler gemittelten Regressionskonstanten γ_{10}

Abbildung 39: Datenstruktur einer wiederholten Messung mittels CBM im Rahmen des unkonditionierten hierarchisch-linearen Mehrebenenmodells (entnommen und modifiziert aus Ditton, 1998, S. 67)

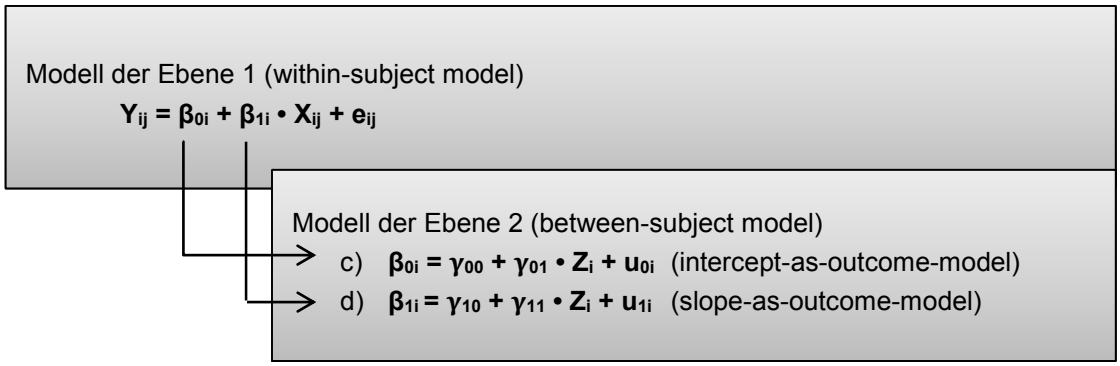
In einem nächsten Schritt wird dann das sog. *allgemeine Modell* (Ditton, 1998) definiert, anhand dessen es ermöglicht werden soll, im unkonditionierten Modell festgestellte Unterschiede der Regressionsparameter zwischen den Schülern durch geeignete Prädiktoren auf der Schülerebene zu erklären (vgl. Abbildung 40). Hierbei wird Z_i als weiterer Prädiktor auf der Ebene 2 in das Modell gegeben, welcher in der vorliegenden Arbeit Angaben über die Zugehörigkeit zur Gruppe von Kindern mit Risiken im mathematischen Vorwissen beinhaltet. Operationalisiert ist dieser

Risikostatus durch einen Prozentrang kleiner bzw. gleich 15 im Diagnoseprogramm Kalkulie (Fritz et al., 2007). Es handelt sich bei Z_i um eine dichotome Variable mit den Ausprägungen „0“, was auf keine Zugehörigkeit zur Risikogruppe hinweist, und „1“, was die Zugehörigkeit zur Risikogruppe anzeigt.

In der vorliegenden Arbeit wird in der Regel davon ausgegangen, dass alle Kinder jeweils ein individuelles Ausgangsniveau sowie einen individuellen Anstieg in den CBM-Daten zeigen, daher wurden u_{0i} und u_{1i} in die Modelle auf der Ebene 2 eingefügt, welche jeweils als Maß für personenspezifische Abweichungen von den gemittelten Regressionsparametern betrachtet werden. D. h., es wird der Einfluss des Prädiktors Z_i auf den Steigungskoeffizienten β_{1i} als auch auf die Regressionskonstante β_{0i} angenommen, sodass zwei Regressionsgleichungen auf der Ebene 2 entstehen, das intercept-as-outcome- und das slope-as-outcome-model. Zeigen sich bei einem oder mehreren CBM keine signifikanten Unterschiede in den berechneten Regressionskonstanten oder Steigungskoeffizienten der untersuchten Schüler, können diese Effekte demnach als fest aufgefasst werden, entsprechend wird dann nur der Einfluss des Prädiktors Z_i auf das slope-as-outcome- bzw. das intercept-as-outcome-model berechnet.

Beim intercept-as-outcome-model hängt der Regressionsparameter β_{0i} von γ_{00}, γ_{01} sowie u_{0i} ab. Während γ_{00} das mittlere CBM-Ergebnis aller Schüler zum ersten Messzeitpunkt angibt, drückt γ_{01} die Differenz der mittleren CBM-Ergebnisse von Kindern ohne und von Kindern mit Risiken im mathematischen Vorwissen (Schülermerkmal Z_i) zum ersten Messzeitpunkt aus. Durch u_{0i} wird die individuelle Abweichung des Schülers i von dem über alle Schüler gemittelten CBM-Ausgangsniveau (γ_{00}) zum ersten Messzeitpunkt berücksichtigt. Der Regressionsparameter β_{1i} hingegen hängt von γ_{10}, γ_{11} sowie u_{1i} ab. Dabei gibt γ_{10} den mittleren Anstieg der CBM-Ergebnisse von

einem Messzeitpunkt zum nächsten an. γ_{11} drückt die Differenz der mittleren CBM-Anstiege von Kindern ohne und von Kindern mit Risiken im mathematischen Vorwissen (Schülermerkmal Z_i) aus und durch u_{1i} wird die individuelle Abweichung des Schülers i von den über alle Schüler gemittelten CBM-Anstiege (γ_{10}) berücksichtigt. Bei der HLM werden die gemittelten Regressionsparameter (mittlere Ausgangswerte und Anstiege) als sog. feste Effekte (fixed effects) bezeichnet, hingegen werden Varianzparameter (individuelle Abweichungen von den mittleren Ausgangswerten und Anstiegen) als zufällige Effekte (random effects) benannt (Bryk & Raudenbush, 1992).



Y_{ij}	CBM-Ergebnis von Schüler i zur Schulwoche j (Kriterium)
i	Index für den Schüler
j	Index für die Schulwoche
β_{0i}	personenspezifischer CBM-Wert von Schüler i zum ersten Testzeitpunkt (intercept)
β_{1i}	personenspezifischer Leistungsanstieg (slope) von Schüler i von einem Testzeitpunkt zum nächsten
X_{ij}	Angabe der Schulwoche (Prädiktor auf Ebene 1)
e_{ij}	personenspezifischer Fehlerterm von Schüler i zur Schulwoche j
γ_{00}	mittleres CBM-Ergebnis aller Schüler zum ersten Testzeitpunkt
γ_{01}	Differenz des mittleren CBM-Ausgangsniveaus von Kindern ohne und mit Risiken im mathematischen Vorwissen (Schülermerkmal Z_i)
Z_i	Angabe zur Zugehörigkeit zur Gruppe der Kinder mit Risiken im mathematischen Vorwissen (Prädiktor auf Ebene 2)
u_{0i}	individuelle Abweichung des Schülers i von der über die Schüler gemittelten Regressionskonstanten γ_{00}
γ_{10}	mittlerer Anstieg der CBM-Werte aller Schüler von einem Testzeitpunkt zum nächsten
γ_{11}	Differenz der mittleren Anstiege der CBM-Werte der Kinder ohne und mit Risiken im mathematischen Vorwissen (Schülermerkmal Z_i) von einem Testzeitpunkt zum nächsten
u_{1i}	individuelle Abweichung des Schülers i von der über die Schüler gemittelten Regressionskonstanten γ_{10}

Abbildung 40: Datenstruktur einer wiederholten Messung (z. B. mittels CBM) im Rahmen des allgemeinen hierarchisch-linearen Mehrebenenmodells (entnommen und modifiziert aus Langer, 2009, S. 225)

Nach der Berechnung der beiden oben aufgeführten Modelle wird der Anteil der durch das allgemeine Modell im Vergleich zum

unkonditionierten Modell erklärten Varianz anhand der nachfolgenden Formel errechnet:

$$Anteil\ d.erklärten\ Var. = \frac{Var(unkond.Modell) - Var(allgem.Modell)}{Var(unkond.Modell)}$$

Zur besseren Interpretierbarkeit der Ergebnisse können die Prädiktoren auf Ebene 1 und auf Ebene 2 zentriert werden. In der vorliegenden Arbeit werden die Angaben der Variablen X_{ij}, also die Schulwoche, in der eine CBM-Testung stattfindet, als Prädiktor auf Ebene 1 umkodiert, sodass der erste Wert 0 ergibt, alle weiteren Ausprägungen der Variable werden entsprechend angepasst. Für die CBM der Niveaustufe 1, welche ausschließlich in der ersten Hälfte des ersten Schuljahres durchgeführt wurden, ergibt sich somit folgende Codierung: 5.SW → X_{i5} = 0; 8.SW → X_{i8} = 3; 12.SW → X_{i12} = 7; 16.SW → X_{i16} = 11. CBM der Niveaustufe 2, welche ausschließlich in der zweiten Hälfte des ersten Schuljahres durchgeführt wurden, werden folgendermaßen codiert: 21.SW → X_{i21} = 0; 26.SW → X_{i26} = 5; 30.SW → X_{i30} = 9; 34.SW → X_{i34} = 13; 38.SW → X_{i38} = 17. Für das CBM „Arithmetische Basiskompetenzen“, welche über das gesamte Schuljahr eingesetzt werden wird folgende Codierung verwendet: 5.SW → X_{i5} = 0; 8.SW → X_{i8} = 3; 12.SW → X_{i12} = 7; 16.SW → X_{i16} = 11; 21.SW → X_{i21} = 16; 26.SW → X_{i26} = 21; 30.SW → X_{i30} = 25; 34.SW → X_{i34} = 29; 38.SW → X_{i38} = 33. Eine Zentrierung des Prädiktors auf Ebene 2 erfolgt nicht. Eine Ausprägung von 1 deutet hier darauf hin, dass der Schüler der Gruppe der Kinder zugehörig ist, welche Risiken im Bereich des mathematischen Vorwissens aufweisen, bei einer Ausprägung von 0 ist er es nicht.

6.4.4 Angaben zur Datenauswertung

Für die statistischen Analysen wird im Wesentlichen das Programm Statistical Package for the Social Sciences (SPSS) in der Version 19 benutzt. Für Analysen zur prognostischen Validität wird auf die von Lenhard (2010) bereitgestellte Auswertungshilfe zur Berechnung von Testkennwerten von Screeningverfahren zurückgegriffen. Es handelt sich dabei um eine Exceldatei mit vordefinierten Rechenalgorithmen, welche ausgehend von den einzugebenden Zellbesetzungen der Klassifikationen wesentliche Kennwerte (Sensitivität, Spezifität, Trefferquote und RATZ-Index) automatisch berechnet. Zur Hierarchisch-linearen Modellierung wird das Programm HLM 7 von Raudenbush, Bryk und Congdon (2011) benutzt.

7 Ergebnisse zur Güte und Anwendbarkeit der entwickelten CBM

Zur Ergebnisdarstellung werden in einem ersten Schritt deskriptive Statistiken der CBM-Ergebnisse der Regionen Rügen und Nordvorpommern analysiert (Abschnitt 7.1). Auf Grundlage dieser und zusätzlich in der Hansestadt Rostock erhobenen Daten soll in einem zweiten Schritt Auskunft über die Reliabilität der CBM gegeben werden (Abschnitt 7.2). Anschließende Untersuchungen der Validität der CBM erfolgen auf Basis der Rügener Schülerergebnisse in den CBM und weiteren Kriteriumsmaßen (Abschnitt 7.3). Analysen zur Änderungssensibilität der CBM werden anhand der erhobenen Daten der Schüler der Region Rügen in Abschnitt 7.4 vorgenommen. Abschließend werden mittels der Ergebnisse einer durchgeführten Lehrerbefragung Hinweise zur Ökonomie, Nutzbarkeit und Zumutbarkeit abgeleitet (Abschnitt 7.5).

7.1 Deskriptive statistische Angaben

In einem ersten Schritt werden im nachfolgenden Abschnitt deskriptive Statistiken der erhobenen CBM-Daten getrennt nach den CBM-Typen sowie getrennt nach Region Rügen und Nordvorpommern dargestellt. Da in der Region Nordvorpommern im Rahmen der Abschätzung der Parallelität der CBM zu jedem Messzeitpunkt jeweils zwei Versionen der Verfahren durchgeführt wurden, sind nachfolgende Analysen für diese Untersuchungsgruppe jeweils getrennt für beide Versionen ausgewiesen.

7.1.1 CBM „Zahlenlesen im Zwanzigerraum“

7.1.1.1 Untersuchungsgruppe Nordvorpommern

Für das CBM „Zahlenlesen im Zwanzigerraum“ liegen für die beiden zum jeweiligen Messzeitpunkt durchgeführten Versionen 1 und 2 zwischen 318 bis 535 gültige Werte von Schülern aus der Region NVP vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle, ungültige Einträge und schulorganisatorische Umstände. Zum Ende des Schuljahres nahm die Bereitschaft zur Teilnahme an den curriculumbasierten Messungen in der Region Nordvorpommern deutlich ab, da die monatlichen Messungen doch einen deutlichen Aufwand für die Lehrkräfte bedeuteten. Dies scheint ebenso der Grund zu sein, weshalb für die Version 2 deutlich weniger Ergebnisse vorliegen. In Tabelle 19 sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Nordvorpommerns für das erste Schuljahr getrennt nach den Versionen tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM der Version 1 liegt zum ersten Messzeitpunkt (5. Schulwoche) bei 29.76 Punkten mit einer Standardabweichung von 14.62 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich an (MZP 2: 36.26 Punkte; MZP 3: 39.26 Punkte; MZP 4: 43.5 Punkte). Die Standardabweichung bleibt über die Messzeitpunkte hinweg mit Werten zwischen 14.27 und 14.83 Punkten recht konstant. Den Messzeitpunkt 3 (Minimum: 5 Punkte) in der zwölften Schulwoche ausgenommen, hat mindestens ein Kind bei jedem der Messzeitpunkte keinen Punkt erzielt (Minimum: 0 Punkte). Zu jeder Messung gibt es jedoch jeweils Schüler, die die Maximalpunktzahl von 64 Punkten erreicht haben. Die Verteilungen der Version 2 zeigen ein ähnliches Bild. Im Laufe der Schulwochen erhöht sich auch hier der Mittelwert der Verteilungen kontinuierlich (MZP 1: 32.87 Punkte; MZP 2: 37.91

Punkte; MZP 3: 40.24 Punkte; MZP 4: 45.23 Punkte). Die Standardabweichung für die Version 2 bleibt mit Werten im Intervall von 14.09 und 14.86 über die Messzeitpunkte konstant. Die Mittelwerte und Standardabweichungen der Versionen 1 und 2 sind zu jedem Zeitpunkt vergleichbar. Zum ersten Messzeitpunkt beträgt das Minimum 5 Punkte, zu den restlichen Testungen fällt das Minimum jeweils mit 0 Punkten aus. Es gibt jedoch auch zu jeder Messung jeweils Schüler, die die Maximalpunktzahl von 64 Punkten erreicht haben.

Tabelle 19: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	CBM „Zahlenlesen im Zwanzigerraum“ Version 1					CBM „Zahlenlesen im Zwanzigerraum“ Version 2				
	N_{V1}	M_{V1}	SD_{V1}	Min_{V1}	Max_{V1}	N_{V2}	M_{V2}	SD_{V2}	Min_{V2}	Max_{V2}
1 (05.SW)	465	29.76	14.62	0	64	318	32.87	14.09	5	64
2 (08.SW)	497	36.26	14.83	0	64	333	37.91	14.86	0	64
3 (12.SW)	442	39.26	14.82	5	64	331	40.24	14.82	0	64
4 (16.SW)	535	43.50	14.27	0	64	435	45.23	14.11	0	64

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 41 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Zahlenlesen im Zwanzigerraum“ der Kinder der Region Nordvorpommern über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres grafisch dargestellt. Links sind die Verteilungen der Version 1, rechts die Verteilungen der parallelen Version 2 abgebildet. In den Grafiken lässt sich die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte in den beiden Versionen sehr deutlich erkennen. Über die Messzeitpunkte hinweg erscheinen die Daten der Version 1 und Version 2 optisch einer Normalverteilung zu entsprechen. Über die Schulwochen hinweg ist eine starke Zunahme der Kinder zu verzeichnen, die alle der Zahlen des CBM

innerhalb der Bearbeitungszeit von einer Minute korrekt benennen konnten.

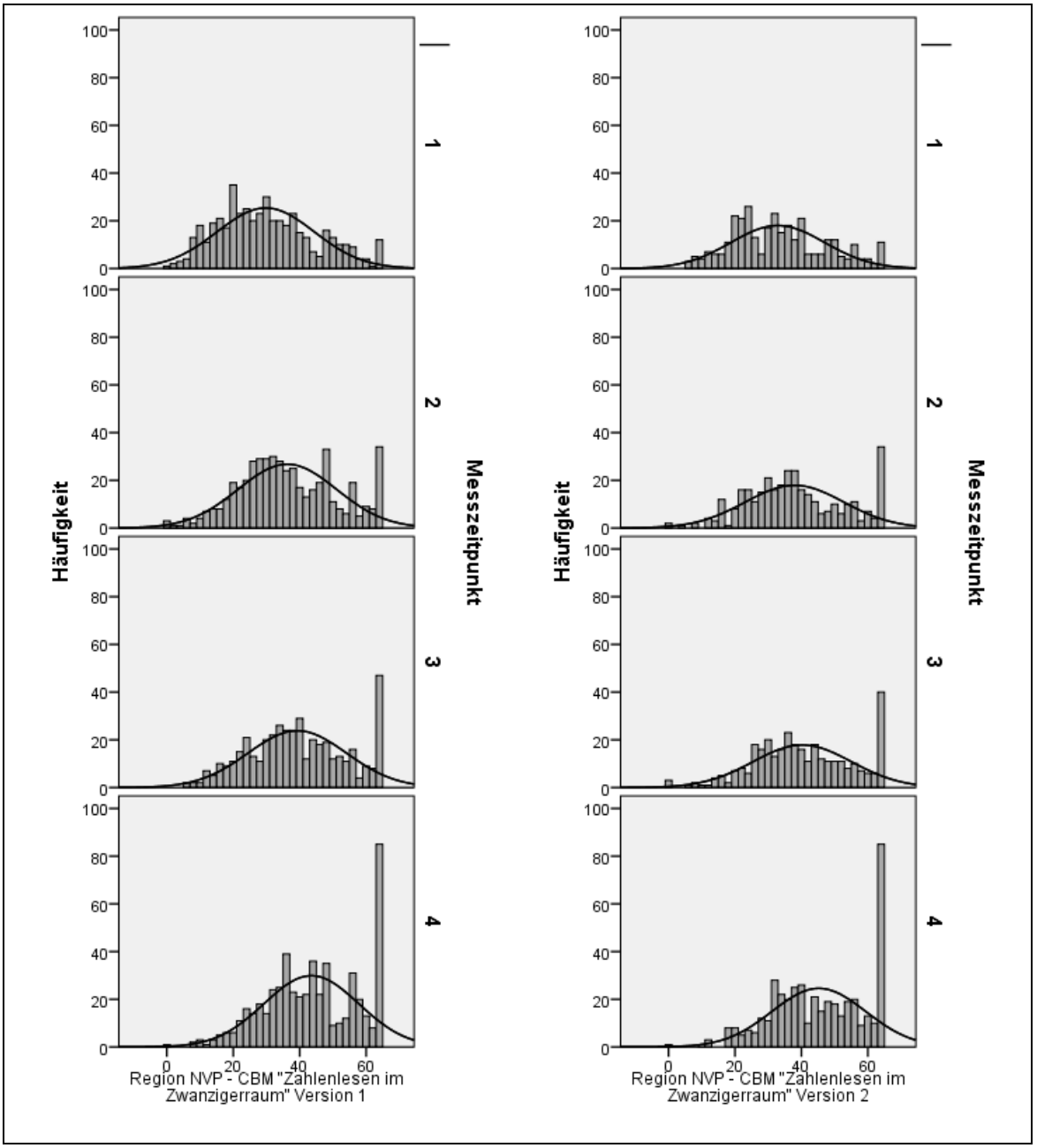


Abbildung 41: Rohwertverteilungen der Versionen 1 und 2 des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern

7.1.1.2 Untersuchungsgruppe Rügen

Für das CBM „Zahlenlesen im Zwanzigerraum“ liegen zwischen 425 bis 438 gültige Werte von Schülern der Insel Rügen vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle oder ungültige Einträge. Nachfolgend sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Rügens für das erste Schuljahr tabellarisch zusammengefasst (Tabelle 20).

Der Mittelwert der Rohwerte im CBM „Zahlenlesen im Zwanzigerraum“ liegt zum ersten Messzeitpunkt in der fünften Schulwoche bei 27.97 Punkten mit einer Standardabweichung von 14.67 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich (MZP2: 32.7 Punkte; MZP 3: 37.83 Punkte; MZP 4: 42.82 Punkte). Die Standardabweichung bleibt über alle Messungen hinweg in einem Intervall von 14.58 bis 16.17 Punkten recht stabil. Zu den Messzeitpunkten 1 sowie 4 hat mindestens ein Kind keinen Punkt erzielen können (Minimum: 0 Punkte). Zu den Messzeitpunkten 2 und 3 liegt das Minimum bei einem Punkt. Das Maximum der Verteilungen liegt über alle Zeitpunkte hinweg bei 64 Punkten, welches der Höchstpunktzahl des Verfahrens entspricht.

Tabelle 20: Deskriptive Statistiken des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen

MZP	CBM „Zahlenlesen im Zwanzigerraum“				
	N	M	SD	Min	Max
1 (05.SW)	438	27.97	14.67	0	64
2 (08.SW)	438	32.70	16.17	1	64
3 (12.SW)	433	37.83	14.85	1	64
4 (16.SW)	425	42.82	14.58	0	64

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 42 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Zahlenlesen im Zwanzigerraum“ der Kinder der Region Rügen über die vier Messzeitpunkte grafisch dargestellt. Die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte lässt sich hier sehr deutlich erkennen. Obgleich der Anteil der Kinder, welche alle 64 vorgegebenen Zahlen innerhalb der Bearbeitungszeit von einer Minute korrekt identifizieren konnten, stark ansteigt, ähneln die Daten über alle Messzeitpunkte hinweg optisch einer Normalverteilung.

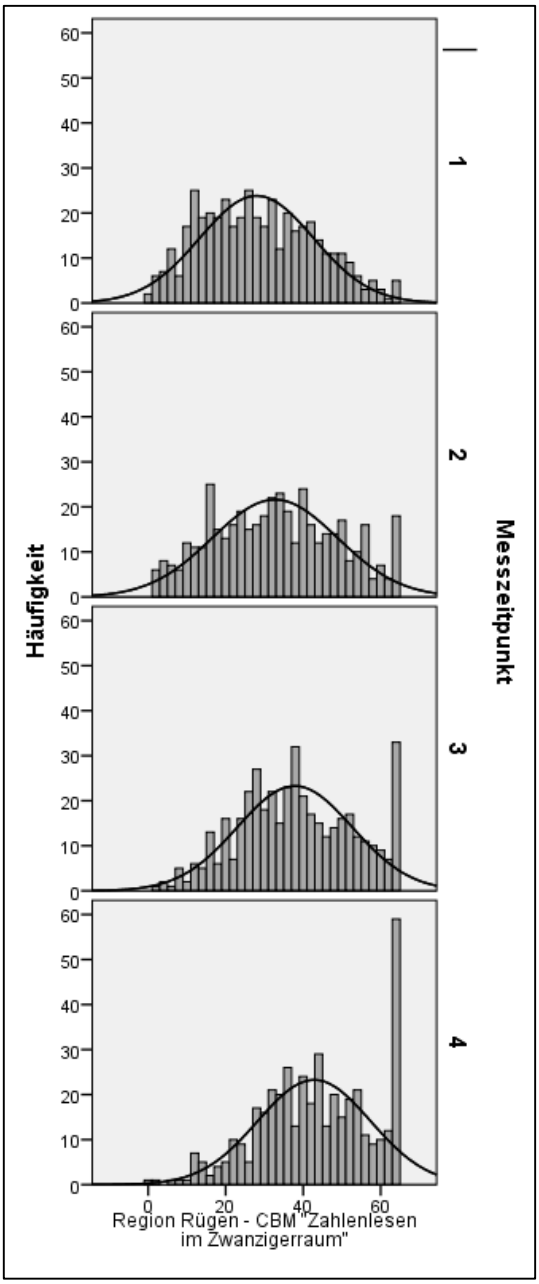


Abbildung 42: Rohwertverteilungen des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen

7.1.2 CBM „Zahlenreihe im Zwanzigerraum“

7.1.2.1 Untersuchungsgruppe Nordvorpommern

Für das CBM „Zahlenreihe im Zwanzigerraum“ liegen für Version 1 und 2 zwischen 334 bis 547 gültige Werte von Schülern aus der Region NVP vor. Auch in diesem Fall variiert die Anzahl der Schüler durch verschiedene Umstände. Durch den hohen Aufwand für die Lehrkräfte nahmen deutlich weniger Kinder an den CBM-Testungen der Version 2 teil. In Tabelle 21 sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Nordvorpommerns für das erste Schuljahr getrennt nach den Versionen tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM der Version 1 liegt zum ersten Messzeitpunkt (5. Schulwoche) bei 4.27 Punkten mit einer Standardabweichung von 3.95 Punkten. Über die Messzeitpunkte hinweg steigt auch in diesem CBM der Mittelwert kontinuierlich an (MZP 2: 7.82 Punkte; MZP 3: 10.58 Punkte; MZP 4: 11.42 Punkte). Die Standardabweichung bleibt ab dem 2. Messzeitpunkt über die Messungen hinweg recht konstant (zwischen 5.41 und 5.74 Punkten). Zum ersten Messzeitpunkt ist sie mit 3.95 Punkten deutlich niedriger. Zu jedem der Messzeitpunkte hat zumindest ein Kind keinen Punkt erzielt (Minimum: 0 Punkte), es gibt jedoch auch zu jeder Messung jeweils Schüler, die die Maximalpunktzahl von 24 Punkten erreicht haben. Die Verteilungen der Version 2 zeigen ein ähnliches Bild. Im Laufe der Schulwochen erhöht sich auch hier der Mittelwert der Verteilungen kontinuierlich (MZP 1: 5.49 Punkte; MZP 2: 8.22 Punkte; MZP 3: 9.67 Punkte; MZP 4: 11.92 Punkte). Die Standardabweichung für die Version 2 ist mit 4.18 Punkten zum ersten Messzeitpunkt deutlich kleiner als zu den restlichen Messzeitpunkten (zwischen 5.30 und 5.75 Punkten). Die Standardabweichungen der Version 1 und 2 sind zu jedem Zeitpunkt vergleichbar. Das Minimum der Verteilungen

liegt über alle Zeitpunkte hinweg bei null Punkten, das Maximum liegt jeweils bei der maximal erreichbaren Punktzahl von 24 Punkten.

Tabelle 21: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Zahlenreihe im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	CBM „Zahlenreihe im Zwanzigerraum“ Version 1					CBM „Zahlenreihe im Zwanzigerraum“ Version 2				
	N_{V1}	M_{V1}	SD_{V1}	Min_{V1}	Max_{V1}	N_{V2}	M_{V2}	SD_{V2}	Min_{V2}	Max_{V2}
1 (05.SW)	484	4.27	3.95	0	24	334	5.49	4.18	0	24
2 (08.SW)	509	7.82	5.41	0	24	391	8.22	5.30	0	24
3 (12.SW)	478	10.58	5.72	0	24	389	9.67	5.75	0	24
4 (16.SW)	547	11.42	5.74	0	24	473	11.92	5.50	0	24

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 43 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Zahlenreihe im Zwanzigerraum“ der Kinder der Region Nordvorpommern über die vier Messzeitpunkte grafisch dargestellt. Links sind die Verteilungen der Version 1, rechts die Verteilungen der parallelen Version 2 abgebildet. Auffällig ist der relativ hohe Anteil an Kindern, welche zum ersten Messzeitpunkt in beiden Versionen keine bzw. nur wenige der Aufgaben innerhalb der Bearbeitungszeit von einer Minute lösen konnten. Ein Großteil der erzielten Werte liegt im Bereich zwischen etwa zwei und sechs Punkten, nur wenige Kinder erzielen eine Punktzahl größer als 11 Punkte. Die Verteilungen sind demnach rechtsschief. Diese Häufung der Daten im unteren Bereich fällt in der Version 1 deutlich stärker aus. Während es zum ersten Messzeitpunkt deutliche Unterschiede in den Verteilungen der Daten der Version 1 und Version 2 gibt, fallen sie ab dem zweiten Messzeitpunkt sehr ähnlich aus. Über die Messzeitpunkte 2 bis 4 erscheinen die Daten beider Versionen optisch einer Normalverteilung zu entsprechen. In den Grafiken lässt sich die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte in den beiden Versionen sehr deutlich erkennen. Ob es signifikante Unterschiede zwischen den Daten der beiden Versionen des CBM „Zahlenlesen im Zwanzigerraum“ gibt, wird in Abschnitt 7.2.3.2 geprüft.

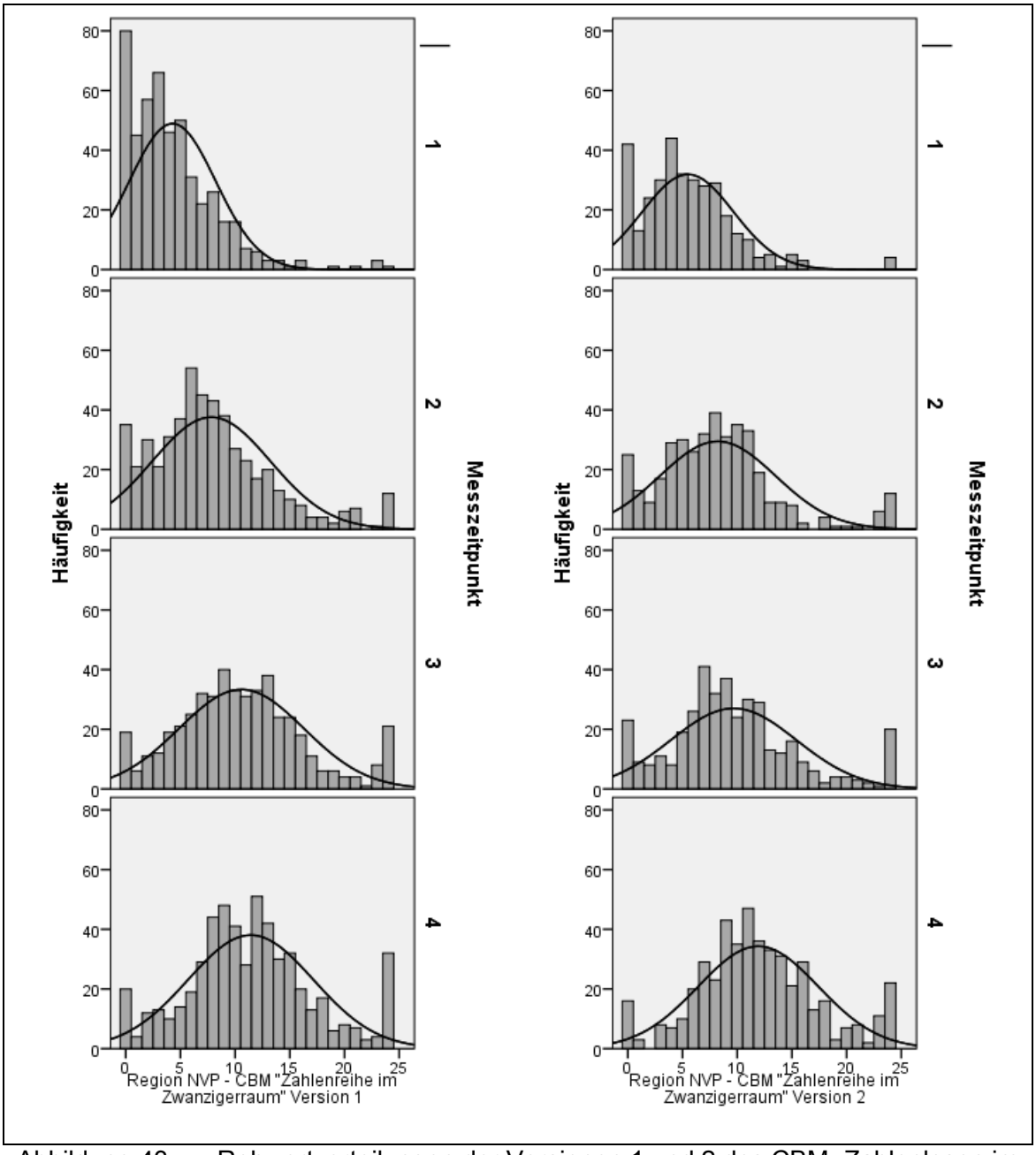


Abbildung 43: Rohwertverteilungen der Versionen 1 und 2 des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern

7.1.2.2 Untersuchungsgruppe Rügen

Für das CBM „Zahlenreihe im Zwanzigerraum“ liegen zwischen 426 bis 437 gültige Werte von Schülern der Insel Rügen vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle oder

ungültige Einträge. In Tabelle 22 sind nachfolgend die statistischen Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Rügens für das erste Schuljahr tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM „Zahlenreihe im Zwanzigerraum“ liegt zum ersten Messzeitpunkt in der fünften Schulwoche bei 4.15 Punkten mit einer Standardabweichung von 3.87 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich (MZP2: 5.85 Punkte; MZP 3: 9.33 Punkte; MZP 4: 10.13 Punkte). Die Standardabweichung bleibt über alle Messungen hinweg in einem Intervall von 3.87 bis 4.77 Punkten recht konstant. Zu allen Messzeitpunkten hat mindestens ein Kind keinen Punkt erzielen können (Minimum: 0 Punkte). Das Maximum der Verteilungen variiert über alle Zeitpunkte hinweg in einem Intervall von 21 und 24 Punkten.

Tabelle 22: Deskriptive Statistiken des CBM „Zahlenreihe im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen

MZP	CBM „Zahlenreihe im Zwanzigerraum“				
	N	M	SD	Min	Max
1 (05.SW)	437	4.15	3.87	0	21
2 (08.SW)	435	5.85	3.98	0	21
3 (12.SW)	429	9.33	4.58	0	22
4 (16.SW)	426	10.13	4.77	0	24

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 44 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Zahlenreihe im Zwanzigerraum“ der Kinder der Region Rügen über die vier Messzeitpunkte grafisch dargestellt. Auffällig ist der relativ hohe Anteil an Kindern, welche zum ersten Messzeitpunkt keine bzw. nur wenige der Aufgaben innerhalb der Bearbeitungszeit von einer Minute lösen konnten. Ein Großteil der erzielten Werte liegt im Bereich zwischen etwa zwei und sechs Punkten, nur wenige Kinder erzielen eine Punktzahl größer als 11 Punkte. Die Verteilung

ist demnach rechtsschief. Über die Messzeitpunkte 2 bis 4 erscheinen die Daten optisch einer Normalverteilung zu entsprechen. In den Grafiken lässt sich die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte erkennen.

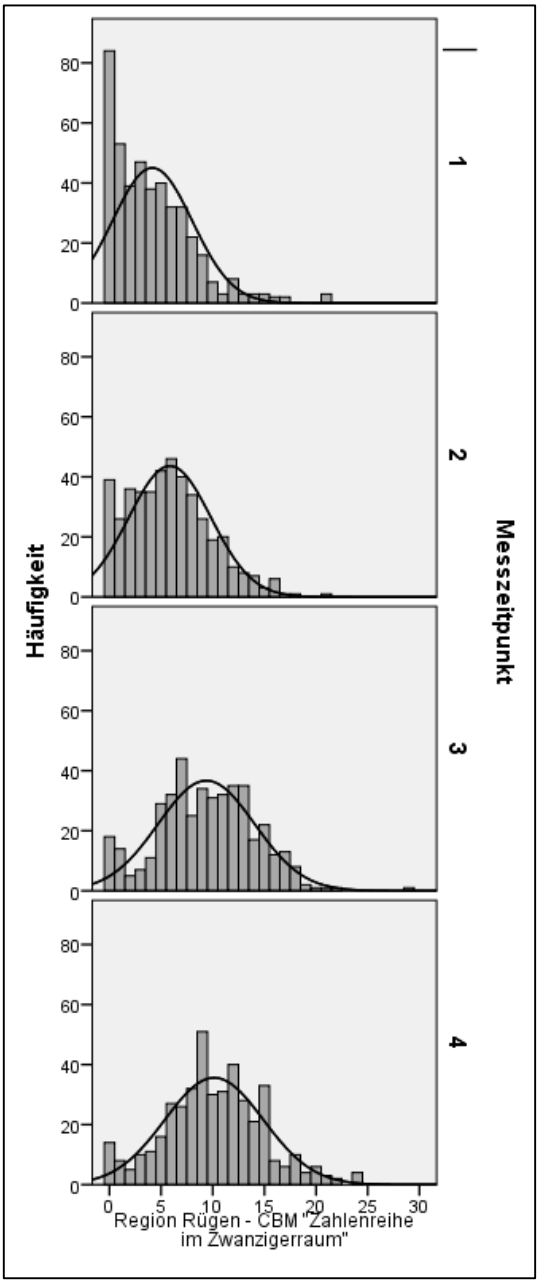


Abbildung 44: Rohwertverteilungen des CBM „Zahlenreihe im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen

7.1.3 CBM „Mengenvergleich im Zwanzigerraum“

7.1.3.1 Untersuchungsgruppe Nordvorpommern

In die Analyse der Verteilungen des CBM „Mengenvergleich im Zwanzigerraum“ gehen zwischen 353 bis 546 gültige Fälle ein. Die Anzahl der Fälle variiert auch in diesem Fall durch unterschiedlich bedingte Ausfälle (Krankheit, ungültige Einträge, schulorganisatorische Umstände, zu hoher Aufwand). Mittelwert, Standardabweichung, Minimum und Maximum der Schüler Nordvorpommerns für das erste Schuljahr sind in Tabelle 23 getrennt nach den Versionen zusammengefasst.

Im Mittel erreichen die Kinder innerhalb der Version 1 zum ersten Messzeitpunkt 16.77 Punkte, zum zweiten 22.65 Punkte, zum dritten 26.09 Punkte und zum vierten Messzeitpunkt 28.64 Punkte. Die Durchschnittswerte erhöhen sich über die Schulwochen hinweg stetig. Die Standardabweichung bleibt mit Werten von 8.6 Punkten zum ersten, 7.91 Punkten zum zweiten, 7.67 Punkten zum dritten und 5.94 Punkten zum vierten Messzeitpunkt relativ konstant. Zu jedem Messzeitpunkt liegt das Minimum der Verteilung bei 0 Punkten. Die Maximalpunktzahl von 32 wird durchweg über jeden Messzeitpunkt erreicht. Ein ähnliches Bild zeigt sich für die Version 2. Im Laufe der Schulwochen erhöht sich auch hier der Mittelwert der Verteilungen vom ersten zum vierten Messzeitpunkt kontinuierlich an (MZP 1: 20.92 Punkte; MZP 2: 25 Punkte; MZP 3: 27.96 Punkte; MZP 4: 29.24 Punkte). Die Standardabweichung für die Version 2 nimmt Werte zwischen 5.06 und 8.43 Punkten an. Das Minimum der Verteilungen liegt über alle Zeitpunkte hinweg bei null Punkten. Das Maximum bleibt konstant bei der maximal erreichbaren Punktzahl von 32.

Tabelle 23: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	CBM „Mengenvergleich im Zwanzigerraum“ Version 1					CBM „Mengenvergleich im Zwanzigerraum“ Version 2				
	N_{V1}	M_{V1}	SD_{V1}	Min_{V1}	Max_{V1}	N_{V2}	M_{V2}	SD_{V2}	Min_{V2}	Max_{V2}
1 (05.SW)	484	16.77	8.60	0	32	353	20.92	8.43	0	32
2 (08.SW)	497	22.65	7.91	0	32	391	25.00	7.21	0	32
3 (12.SW)	480	26.09	7.67	0	32	391	27.96	6.26	0	32
4 (16.SW)	546	28.64	5.94	0	32	477	29.24	5.06	0	32

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

Abbildung 45 zeigt die Verteilungen der erreichten Rohwertpunkte im CBM „Mengenvergleich im Zwanzigerraum“ in der Region Nordvorpommern über die vier Messzeitpunkte der ersten Schuljahreshälfte in Klasse eins. Links sind die Verteilungen der Version 1, rechts die Verteilungen der parallelen Version 2 abgebildet. Deutlich sichtbar ist eine Verschiebung der Mittelwerte in den beiden Versionen über die Messzeitpunkte. Zum ersten Messzeitpunkt scheinen die Daten der Version 1 optisch einer Normalverteilung zu entsprechen, die Verteilung der Version 2 hingegen erscheint zu diesem Zeitpunkt leicht linksschief. Ab dem zweiten Messzeitpunkt ähneln sich die Daten der Versionen 1 und 2. Von Messzeitpunkt zu Messzeitpunkt erhöht sich in beiden Versionen der Anteil an Kindern, die die Maximalpunktzahl von 32 erreichen. Die Verteilungen werden zunehmend linksschief, weichen immer mehr von der Normalverteilung ab. Bei den Abbildungen der CBM-Werte in Version 1 und 2 zum Messzeitpunkt 3 und 4 handelt es sich lediglich um Ausschnitte der Verteilungen. Aus Gründen der Übersichtlichkeit wurde eine maximale Skalierung bis 100 Fälle auf der Ordinate vorgenommen, die tatsächliche Anzahl der Fälle, welche die Maximalpunktzahl von 32 erreicht haben, liegt deutlich höher. So erreichten zum dritten Messzeitpunkt 201 Kinder (Version 1) bzw. 188 Kinder (Version 2) diesen Wert, zum vierten Messzeitpunkt waren es sogar 297 Kinder in Version 1 und 249 in Version 2. Ob es signifikante

Unterschiede zwischen den Daten der beiden Versionen des CBM „Mengenvergleich im Zwanzigerraum“ gibt, wird in Abschnitt 7.2.3.3 geprüft.

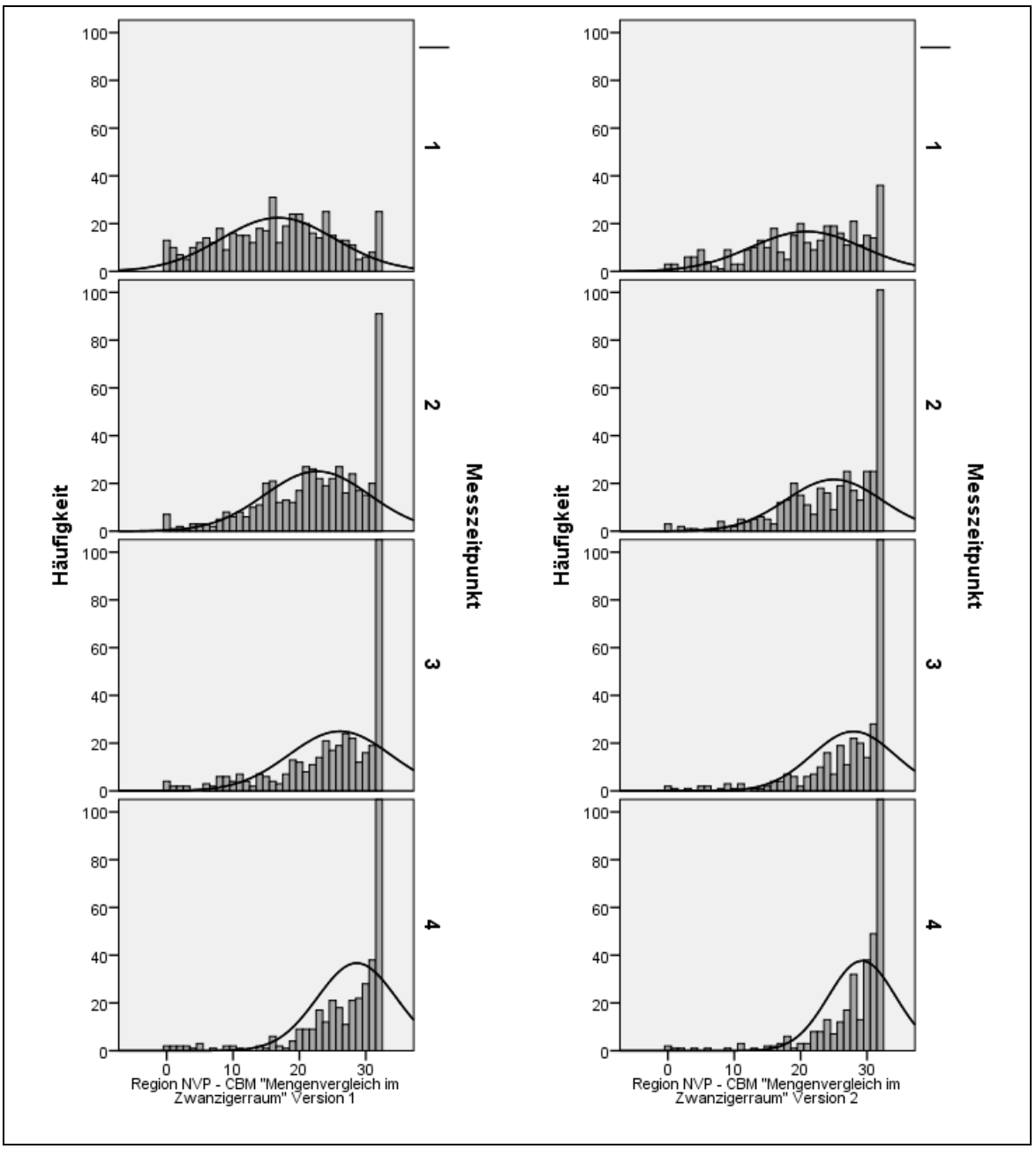


Abbildung 45: Rohwertverteilungen der Versionen 1 und 2 des CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern (Ausschnitt)

7.1.3.2 Untersuchungsgruppe Rügen

Für das CBM „Mengenvergleich im Zwanzigerraum“ liegen zwischen 425 bis 437 gültige Werte von Schülern der Insel Rügen vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle oder ungültige Einträge. Nachfolgend sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Rügens für das erste Schuljahr tabellarisch zusammengefasst (Tabelle 24).

Der Mittelwert der Rohwerte im CBM „Mengenvergleich im Zwanzigerraum“ liegt zum ersten Messzeitpunkt (5. Schulwoche) bei 19.06 Punkten mit einer Standardabweichung von 8.89 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich (MZP 2: 21.77 Punkte; MZP 3: 25.38 Punkte; MZP 4: 26.99 Punkte). Die Standardabweichung bleibt über alle Messungen hinweg in einem Intervall von 6.9 bis 8.89 Punkten recht stabil. Zu jedem der Messzeitpunkte hat mindestens ein Kind keinen Punkt erzielen können (Minimum: 0 Punkte). Das Maximum der Verteilungen liegt über alle Zeitpunkte hinweg bei 32 Punkten, welches der Höchstpunktzahl des Verfahrens entspricht.

Tabelle 24: Deskriptive Statistiken des CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen

MZP	CBM „Mengenvergleich im Zwanzigerraum“				
	N	M	SD	Min	Max
1 (05.SW)	437	19.06	8.89	0	32
2 (08.SW)	431	21.77	7.89	0	32
3 (12.SW)	429	25.38	7.85	0	32
4 (16.SW)	425	26.99	6.90	0	32

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 46 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Mengenvergleich im Zwanzigerraum“ der Kinder der Region Rügen über die vier Messzeitpunkte grafisch dargestellt. Die zuvor

erwähnte kontinuierliche Verschiebung der Mittelwerte lässt sich hier sehr deutlich erkennen. Zum ersten Messzeitpunkt ähnelt die Datenverteilung, obgleich sie sehr flachgipflig ausfällt, optisch einer Normalverteilung, weist jedoch eine leichte linksschiefe Tendenz auf. Zu diesem Zeitpunkt ist der Anteil an Kindern, welche die Maximalpunktzahl von 32 erreichen, bereits recht hoch. Die Verteilungen ab der zweiten Messung weisen eine zunehmende Ballung der Daten im oberen Bereich auf. Die Anzahl an Kindern mit erreichter Maximalpunktzahl erhöht sich über die Messzeitpunkte zunehmend. Bei den Abbildungen der CBM-Werte zum Messzeitpunkt 3 und 4 handelt es sich lediglich um Ausschnitte der Verteilungen. Aus Gründen der Übersichtlichkeit wurde eine maximale Skalierung bis 60 Fälle auf der Ordinate vorgenommen, die tatsächliche Anzahl der Fälle, welche die Maimalpunktzahl von 32 erreicht haben, liegt deutlich höher. So erreichten zum dritten Messzeitpunkt 144 Kinder diesen Wert, zum vierten Messzeitpunkt waren es sogar 174 Kinder.

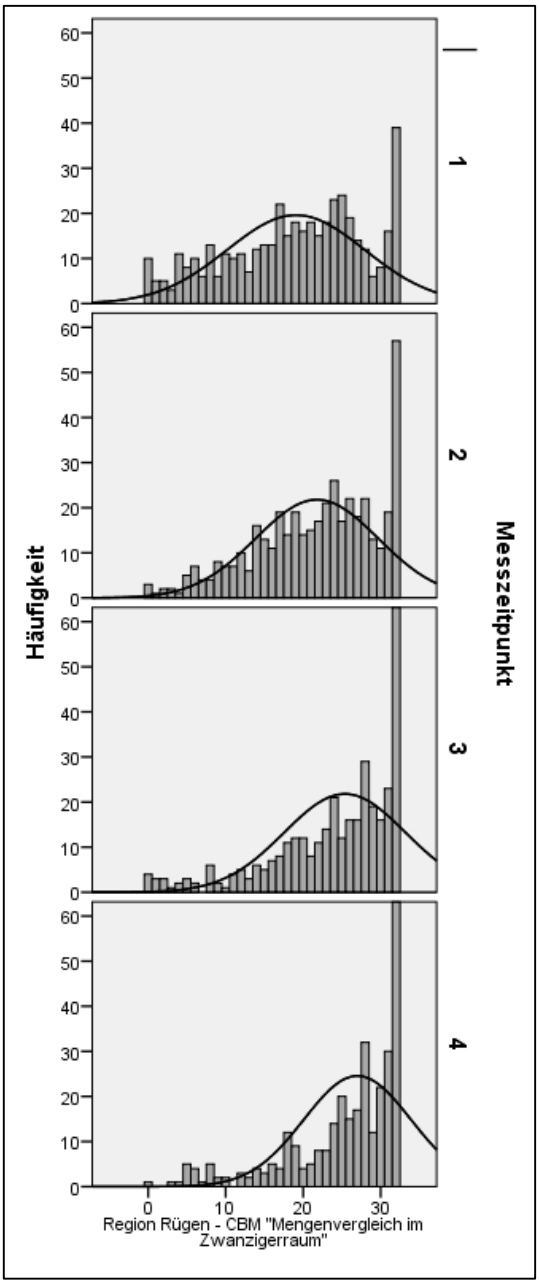


Abbildung 46: Rohwertverteilungen des CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen (Ausschnitt)

7.1.4 CBM „Arithmetische Basiskompetenzen“

7.1.4.1 Untersuchungsgruppe Nordvorpommern

Für das CBM „Arithmetische Basiskompetenzen“ liegen zwischen 293 bis 639 gültige Werte von Schülern aus Nordvorpommern vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle, ungültige Einträge und schulorganisatorische Umstände. Zum Ende des Schuljahres nahm die Bereitschaft zur Teilnahme an den curriculumbasierten Messungen in der Region Nordvorpommern deutlich ab, da die monatlichen Messungen doch deutlichen Aufwand für die Lehrkräfte bedeuteten. Dies scheint ebenso der Grund zu sein, weshalb für die Version 2 deutlich weniger Ergebnisse vorliegen. Tabelle 25 gibt nachfolgend eine Übersicht über die statistischen Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Nordvorpommerns für das erste Schuljahr getrennt nach den Versionen.

Der Mittelwert der Rohwerte im CBM der Version 1 liegt zum ersten Messzeitpunkt (5. Schulwoche) bei 3.9 Punkten mit einer Standardabweichung von 3.02 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich an (MZP 2: 5.73 Punkte; MZP 3: 6.79 Punkte; MZP 4: 8.25 Punkte; MZP 5: 9.13 Punkte; MZP 6: 10.72 Punkte; MZP 7: 10.87 Punkte; MZP 8: 11.08 Punkte). Vom achten zum neunten Messzeitpunkt ist ein Rückgang des Mittelwerts auf 10.68 Punkte zu verzeichnen. Die Standardabweichung bleibt über alle Messungen hinweg relativ konstant in einem Intervall von 3.02 bis 4.08 Punkten. Zu jedem der Messzeitpunkte hat zumindest ein Kind keinen Punkt erzielt (Minimum: 0 Punkte). Zum ersten Messzeitpunkt beträgt der Maximalwert der Verteilung 15 Punkte, ab der zweiten Messung gab es immer Schüler, die die Maximalpunktzahl von 16 erreicht haben. Ein ähnliches Bild zeigt sich in der Version 2 des CBM. Im Laufe der Schulwochen erhöht sich auch hier der Mittelwert der Verteilungen, außer vom sechsten zum

siebten sowie vom achten zum neunten Messzeitpunkt kontinuierlich (MZP 1: 3.88 Punkte; MZP 2: 6.78 Punkte; MZP 3: 7.94 Punkte; MZP 4: 8.76 Punkte; MZP 5: 8.77 Punkte; MZP 6: 9.45 Punkte; MZP 7: 9.00 Punkte; MZP 8: 9.72 Punkte; MZP 9: 9.37 Punkte). Die Standardabweichung für die Version 2 nimmt Werte zwischen 2.43 und 4.26 Punkten an, was auf eine teilweise etwas breitere Streuung der Werte innerhalb der Version 2 hinweist. Das Minimum der Verteilungen liegt über alle Zeitpunkte hinweg bei null Punkten, das Maximum liegt zu Beginn bei 13 Punkten. Ab dem zweiten Messzeitpunkt erreicht mindestens ein Kind den Maximalwert von 16 Punkten.

Tabelle 25: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr in der Region Nordvorpommern

MZP	CBM „Arithmetische Basiskompetenzen“ Version 1					CBM „Arithmetische Basiskompetenzen“ Version 2				
	N_{V1}	M_{V1}	SD_{V1}	Min_{V1}	Max_{V1}	N_{V2}	M_{V2}	SD_{V2}	Min_{V2}	Max_{V2}
1 (05.SW)	482	3.90	3.02	0	15	337	3.88	2.43	0	13
2 (08.SW)	482	5.73	3.66	0	16	386	6.78	3.86	0	16
3 (12.SW)	479	6.79	3.54	0	16	388	7.94	4.26	0	16
4 (16.SW)	532	8.25	3.92	0	16	444	8.76	4.21	0	16
5 (21.SW)	639	9.13	4.03	0	16	579	8.77	4.08	0	16
6 (26.SW)	625	10.72	3.92	0	16	558	9.45	3.73	0	16
7 (30.SW)	362	10.87	4.08	0	16	326	9.00	4.25	0	16
8 (34.SW)	507	11.08	3.79	0	16	410	9.72	3.74	0	16
9 (38.SW)	315	10.68	3.76	0	16	293	9.37	4.26	0	16

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 47 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Arithmetische Basiskompetenzen“ der Kinder der Region Nordvorpommern über die neun Messzeitpunkte in Klasse eins grafisch dargestellt. Links sind die Verteilungen der Version 1, rechts die Verteilungen der parallelen Version 2 abgebildet. Die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte in den beiden Versionen lässt sich hier sehr deutlich erkennen. Zum ersten Messzeitpunkt scheinen die Daten der Version 1 und Version 2

optisch einer Normalverteilung zu entsprechen. Ein Großteil der erzielten Werte liegt im Bereich zwischen etwa zwei und sechs Punkten, nur wenige Kinder erzielen eine Punktzahl größer als 11 Punkte. Ab der zweiten Messung werden die Verteilungen flachgipfliger, die Werte streuen breiter über das ganze Punktzahlspektrum, jedoch ähneln die Verteilungen bis zum sechsten Messzeitpunkt einer Normalverteilung. Zum siebten, achten und neunten Messzeitpunkt sind die Verteilungen in beiden Versionen sehr flach. In Version 1 gibt es einen höheren Anteil an Kindern, die die Gesamtpunktzahl des Verfahrens erreichen. Bis auf die Verteilung der Daten der achten Testung mit der Version 2 des CBM „Arithmetische Basiskompetenzen“ erscheinen die Werte nicht mehr normalverteilt zu sein. Ob es signifikante Unterschiede zwischen den Daten der beiden Versionen des CBM „Arithmetische Basiskompetenzen“ gibt, wird in Abschnitt 7.2.3.4 geprüft.

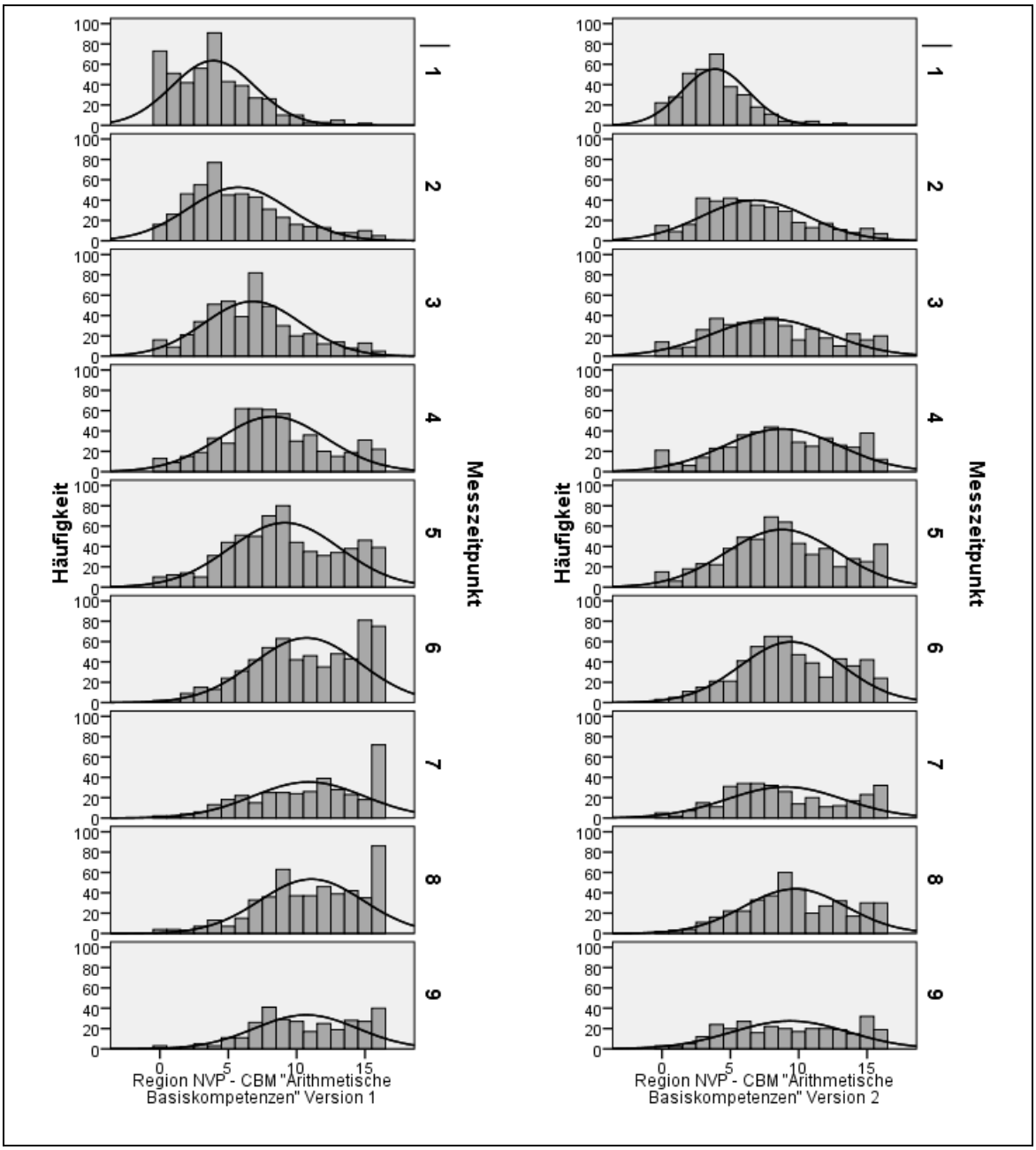


Abbildung 47: Rohwertverteilungen der Versionen 1 und 2 des CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr in der Region Nordvorpommern

7.1.4.2 Untersuchungsgruppe Rügen

Für das CBM „Arithmetische Basiskompetenzen“ liegen zwischen 423 bis 438 gültige Werte von Schülern der Insel Rügen vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle oder ungültige Einträge. Tabelle 26 gibt einen Überblick über

wesentliche statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) des CBM der Schüler Rügens für das erste Schuljahr.

Der Mittelwert der Rohwerte im CBM „Arithmetische Basiskompetenzen“ liegt zum ersten Messzeitpunkt (5. Schulwoche) bei 4.76 Punkten mit einer Standardabweichung von 3.21 Punkten. Bis auf einige Ausnahmen steigt der Mittelwert über die Messzeitpunkte hinweg kontinuierlich an (MZP2: 4.67 Punkte; MZP 3: 5.46 Punkte; MZP 4: 6.61 Punkte; MZP 5: 7.64 Punkte; MZP 6: 7.42 Punkte; MZP 7: 9.48 Punkte). Ab der 30. Schulwoche ist ein Rückgang des Mittelwertes auf 8.97 Punkte zum achten Messzeitpunkt und auf 8.74 Punkte zum neunten Messzeitpunkt zu verzeichnen. Die Standardabweichung bleibt über alle Messungen hinweg relativ konstant in einem Intervall von 2.9 bis 3.97 Punkten. Zu jedem der Messzeitpunkte hat zumindest ein Kind keinen Punkt erzielt (Minimum: 0 Punkte). Zu den ersten beiden Messzeitpunkten beträgt der Maximalwert der Verteilung jeweils 15 Punkte, ab der dritten Messung gab es immer Schüler, die die Maximalpunktzahl von 16 erreicht haben.

Tabelle 26: Deskriptive Statistiken des CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr in der Region Rügen

MZP	CBM „Arithmetische Basiskompetenzen“				
	N	M	SD	Min	Max
1 (05.SW)	438	4.76	3.21	0	15
2 (08.SW)	436	4.67	2.90	0	15
3 (12.SW)	429	5.46	3.02	0	16
4 (16.SW)	423	6.61	3.22	0	16
5 (21.SW)	426	7.64	3.43	0	16
6 (26.SW)	432	7.42	3.38	0	16
7 (30.SW)	431	9.48	3.97	0	16
8 (34.SW)	373	8.97	3.38	0	16
9 (38.SW)	428	8.74	3.38	0	16

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 48 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Arithmetische Basiskompetenzen“ der Kinder der Region Rügen über die neun Messzeitpunkte in Klasse eins grafisch dargestellt. Die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte lässt sich hier sehr deutlich erkennen. Den siebten ausgenommen, erscheinen die Daten über alle Messzeitpunkte hinweg normalverteilt zu sein. Zum siebten Messzeitpunkt in der 30. Schulwoche fällt die Verteilung sehr flachgipflig aus, die Werte der Kinder streuen zu diesem Zeitpunkt recht gleichmäßig über das gesamte Spektrum.

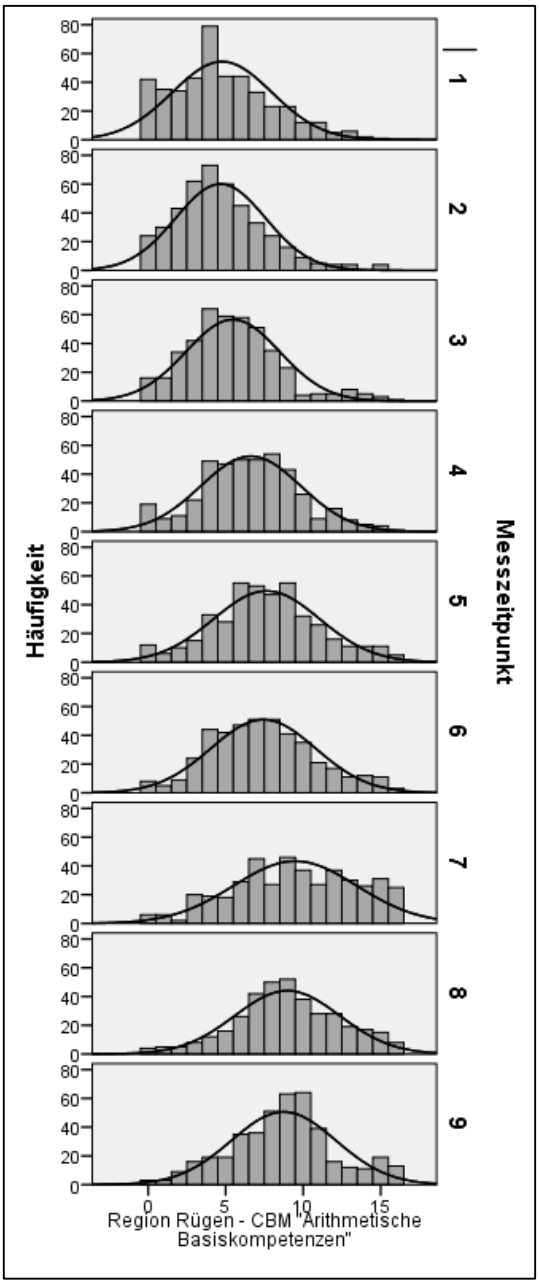


Abbildung 48: Rohwertverteilungen des CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr in der Region Rügen

7.1.5 CBM „Zahlzerlegung im Zwanzigerraum“

7.1.5.1 Untersuchungsgruppe Nordvorpommern

Für das CBM „Zahlzerlegung im Zwanzigerraum“ liegen für die beiden zu den jeweiligen Messzeitpunkten durchgeführten Versionen 1 und 2 zwischen 300 bis 645 gültige Werte von Schülern aus der Region NVP vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle, ungültige Einträge und schulorganisatorische Umstände. Zum Ende des Schuljahres nahm die Bereitschaft zur Teilnahme an den curriculumbasierten Messungen in der Region Nordvorpommern deutlich ab, da die monatlichen Messungen doch deutlichen Aufwand für die Lehrkräfte bedeuteten. Dies scheint ebenso der Grund zu sein, weshalb für die Version 2 deutlich weniger Ergebnisse vorliegen. In Tabelle 27 sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Nordvorpommerns für das erste Schuljahr getrennt nach den Versionen tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM der Version 1 liegt zum fünften Messzeitpunkt (21. Schulwoche) bei 4.65 Punkten mit einer Standardabweichung von 3.29 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich an (MZP 6: 5.27 Punkte; MZP 7: 8.02 Punkte). Zum achten Messzeitpunkt fällt der Mittelwert auf 7.23 Punkte und steigt dann wieder zum neunten Messzeitpunkt auf 7.83. Die Standardabweichung bleibt über die Messzeitpunkte hinweg mit Werten zwischen 3.29 und 4.44 Punkten recht konstant. Zu jedem der Messzeitpunkte hat mindestens ein Kind keinen Punkt erzielt (Minimum: 0 Punkte), es gibt jedoch auch zu jeder Messung jeweils Schüler, die die Maximalpunktzahl von 20 Punkten erreicht haben. Die Verteilungen der Version 2 zeigen ein ähnliches Bild. Im Laufe der Schulwochen erhöht sich auch hier der Mittelwert der Verteilungen kontinuierlich (MZP 5: 4.82 Punkte; MZP 6: 5.44 Punkte; MZP 7: 6.42 Punkte; MZP 8: 7.80 Punkte). Vom achtem zum neunten

Messzeitpunkt sinkt der Mittelwert auf 7.79 Punkte. Die Standardabweichung für die Version 2 bleibt mit Werten im Intervall von 3.06 und 4.46 über die Messzeitpunkte konstant. Zum fünften Messzeitpunkt beträgt das Maximum 18 Punkte, zu den restlichen Testungen fällt es jeweils mit der Höchstpunktzahl von 20 aus. Es gibt auch zu jeder Messung jeweils Schüler, die einen Wert entsprechend des Minimums von 0 Punkten erzielten. Insgesamt lässt sich festhalten, dass die Mittelwerte und Standardabweichungen der Versionen 1 und 2 zu jedem Zeitpunkt vergleichbar sind.

Tabelle 27: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	CBM „Zahlzerlegung im Zwanzigerraum“ Version 1					CBM „Zahlzerlegung im Zwanzigerraum“ Version 2				
	N_{V1}	M_{V1}	SD_{V1}	Min_{V1}	Max_{V1}	N_{V2}	M_{V2}	SD_{V2}	Min_{V2}	Max_{V2}
5 (21.SW)	645	4.65	3.29	0	20	594	4.82	3.06	0	18
6 (26.SW)	626	5.27	3.52	0	20	561	5.44	3.53	0	20
7 (30.SW)	363	8.02	4.11	0	20	326	6.42	4.09	0	20
8 (34.SW)	505	7.23	4.44	0	20	414	7.80	4.46	0	20
9 (38.SW)	305	7.83	4.25	0	20	300	7.79	4.11	0	20

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 49 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Zahlzerlegung im Zwanzigerraum“ der Kinder der Region Nordvorpommern über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres grafisch dargestellt. Links sind die Verteilungen der Version 1, rechts die Verteilungen der parallelen Version 2 abgebildet. In den Grafiken lässt sich die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte in den beiden Versionen sehr deutlich erkennen. Zum fünften und sechsten Messzeitpunkt ist eine Tendenz der Verteilungen zur Rechtssteilheit zu erkennen. Insgesamt erscheinen die Daten der Version 1 und Version 2 über die Messzeitpunkte hinweg optisch einer Normalverteilung zu entsprechen. Auch weisen die Verteilungen der Versionen 1 und 2

starke Übereinstimmungen auf. In Abschnitt 7.2.3.5 wird geprüft, ob diese Unterschiede tatsächlich nicht statistisch signifikant ausfallen.

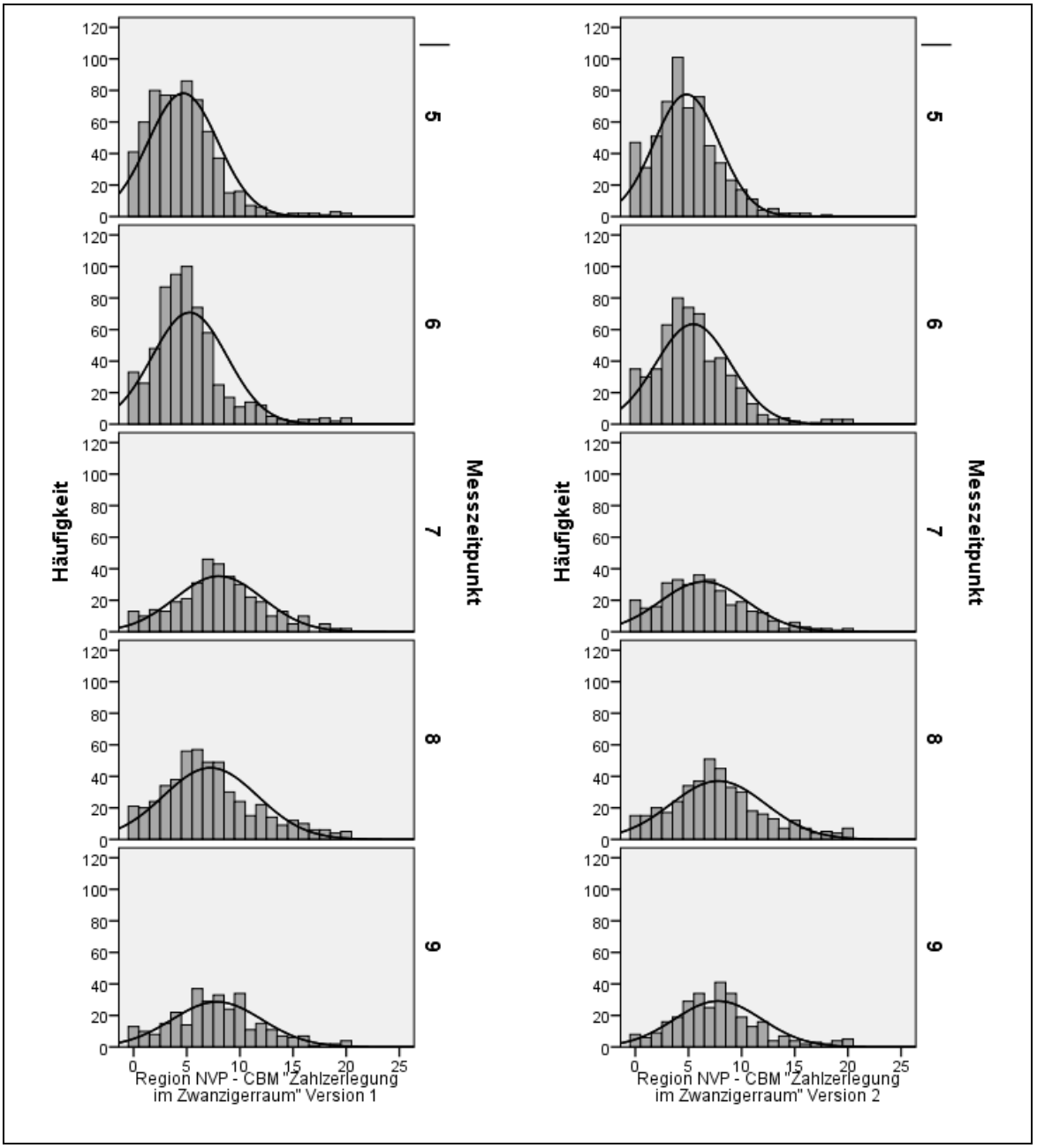


Abbildung 49: Rohwertverteilungen der Versionen 1 und 2 des CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

7.1.5.2 Untersuchungsgruppe Rügen

Für das CBM „Zahlzerlegung im Zwanzigerraum“ liegen zwischen 373 bis 433 gültige Werte von Schülern der Insel Rügen vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle oder ungültige Einträge. In Tabelle 28 sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Rügens für das erste Schuljahr tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM „Zahlzerlegung im Zwanzigerraum“ liegt zum ersten Messzeitpunkt in der fünften Schulwoche bei 3.35 Punkten mit einer Standardabweichung von 2.9 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich (MZP 6: 4.82 Punkte; MZP 7: 6.92 Punkte). Zum achten Messzeitpunkt fällt der Mittelwert auf 6.74 Punkte ab, steigt jedoch dann zum neunten Messzeitpunkt auf 6.76 Punkte. Die Standardabweichung bleibt über alle Messungen hinweg in einem Intervall von 2.9 und 3.72 Punkten recht stabil. Zu allen Messzeitpunkten hat mindestens ein Kind keinen Punkt erzielen können (Minimum: 0 Punkte). Zum Messzeitpunkt 5 liegt das Maximum bei 16 Punkten, über alle späteren Zeitpunkte hinweg liegt es bei der Höchstpunktzahl von 20 Punkten.

Tabelle 28: Deskriptive Statistiken des CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen

MZP	CBM „Zahlzerlegung im Zwanzigerraum“				
	N	M	SD	Min	Max
5 (21.SW)	420	3.35	2.90	0	16
6 (26.SW)	433	4.82	3.17	0	20
7 (30.SW)	431	6.92	3.62	0	20
8 (34.SW)	373	6.74	3.72	0	20
9 (38.SW)	429	6.76	3.65	0	20

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 50 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Zahlzerlegung im Zwanzigerraum“ der Kinder der Region Rügen über die fünf Messzeitpunkte grafisch dargestellt. Die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte lässt sich hier sehr deutlich erkennen. Obgleich der Anteil der Kinder, welche alle 20 vorgegebenen Zahlen innerhalb der Bearbeitungszeit von einer Minute korrekt zerlegen konnten über alle Messzeitpunkte hinweg stark ansteigt, gleichen alle Daten optisch jeweils einer Normalverteilung.

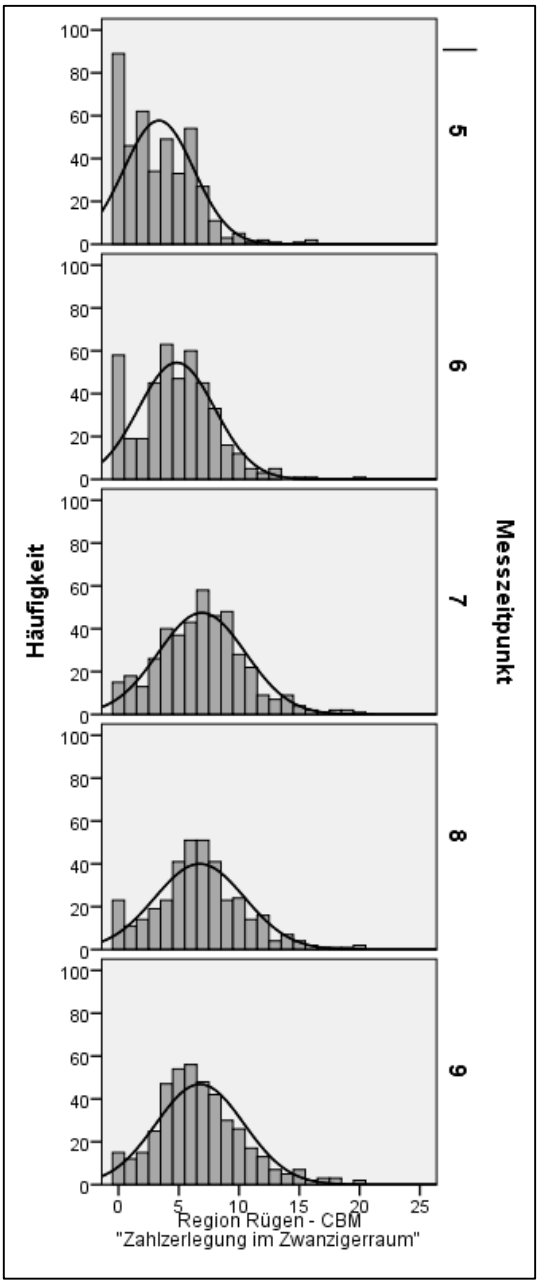


Abbildung 50: Rohwertverteilungen des CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen

7.1.6 CBM „Addition im Zwanzigerraum“

7.1.6.1 Untersuchungsgruppe Nordvorpommern

Für das CBM „Addition im Zwanzigerraum“ liegen für die beiden zu den jeweiligen Messzeitpunkten durchgeführten Versionen 1 und 2 zwischen 314 bis 646 gültige Werte von Schülern aus der Region NVP vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle, ungültige Einträge und schulorganisatorische Umstände. Zum Ende des Schuljahres nahm die Bereitschaft zur Teilnahme an den curriculumbasierten Messungen in der Region Nordvorpommern deutlich ab, da die monatlichen Messungen doch deutlichen Aufwand für die Lehrkräfte bedeuteten. Dies scheint ebenso der Grund zu sein, weshalb für die Version 2 deutlich weniger Ergebnisse vorliegen. In Tabelle 29 sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Nordvorpommerns für das erste Schuljahr getrennt nach den Versionen tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM der Version 1 liegt zum fünften Messzeitpunkt (21. Schulwoche) bei 6.78 Punkten mit einer Standardabweichung von 3.22 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich an (MZP 6: 8.02 Punkte; MZP 7: 9.4 Punkte; MZP 8: 11.65 Punkte). Zum neunten Messzeitpunkt in der 38. Schulwoche fällt der Mittelwert auf 9.92 Punkte. Die Standardabweichung bleibt über die Messzeitpunkte hinweg mit Werten zwischen 3.22 und 4.65 Punkten recht konstant. Zu jedem der Messzeitpunkte hat mindestens ein Kind keinen Punkt erzielt (Minimum: 0 Punkte). Die maximal erreichten Werte schwanken über die Messzeitpunkte in einem Bereich von 26 und 40 Punkten. Die Verteilungen der Version 2 zeigen ein ähnliches Bild. Im Laufe der Schulwochen erhöht sich auch hier der Mittelwert der Verteilungen kontinuierlich (MZP 5: 8.06 Punkte; MZP 6: 9.46 Punkte; MZP 7: 10.81 Punkte). Vom achtem zum neunten Messzeitpunkt sinkt

der Mittelwert auf 10.2 Punkte, steigt jedoch zum neunten Messzeitpunkt auf 11.63 Punkte. Die Standardabweichung für die Version 2 bleibt ebenso mit Werten im Intervall von 3.08 und 4.73 über die Messzeitpunkte konstant. Das erzielte Maximum variiert über die Messzeitpunkt in einem Intervall von 23 bis 40 Punkten. Es gibt zu jeder Messung jeweils Schüler, die einen Wert entsprechend des Minimums von 0 Punkten erzielten. Insgesamt lässt sich festhalten, dass die Mittelwerte und Standardabweichungen der Versionen 1 und 2 sich zu jedem Zeitpunkt ähneln, die Ergebnisse der Version 2 liegen jedoch durchweg höher.

Tabelle 29: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	CBM „Addition im Zwanzigerraum“ Version 1					CBM „Addition im Zwanzigerraum“ Version 2				
	N_{V1}	M_{V1}	SD_{V1}	Min_{V1}	Max_{V1}	N_{V2}	M_{V2}	SD_{V2}	Min_{V2}	Max_{V2}
5 (21.SW)	646	6.78	3.22	0	30	595	8.06	3.08	0	23
6 (26.SW)	626	8.02	3.81	0	34	577	9.46	3.93	0	38
7 (30.SW)	364	9.40	4.40	0	40	327	10.81	4.50	0	33
8 (34.SW)	504	11.65	4.65	0	40	417	10.20	4.68	0	40
9 (38.SW)	320	9.92	3.65	0	26	314	11.63	4.73	0	40

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 51 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Addition im Zwanzigerraum“ der Kinder der Region Nordvorpommern über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres grafisch dargestellt. Links sind die Verteilungen der Version 1, rechts die Verteilungen der parallelen Version 2 abgebildet. In den Grafiken lässt sich die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte in den beiden Versionen leicht erkennen. Insgesamt erscheinen die Daten der Version 1 und Version 2 über die Messzeitpunkte hinweg optisch einer Normalverteilung zu entsprechen. Auch weisen die Verteilungen der Versionen 1 und 2 starke Übereinstimmungen auf. In Abschnitt

7.2.3.6 wird geprüft, ob diese Unterschiede tatsächlich nicht statistisch signifikant ausfallen.

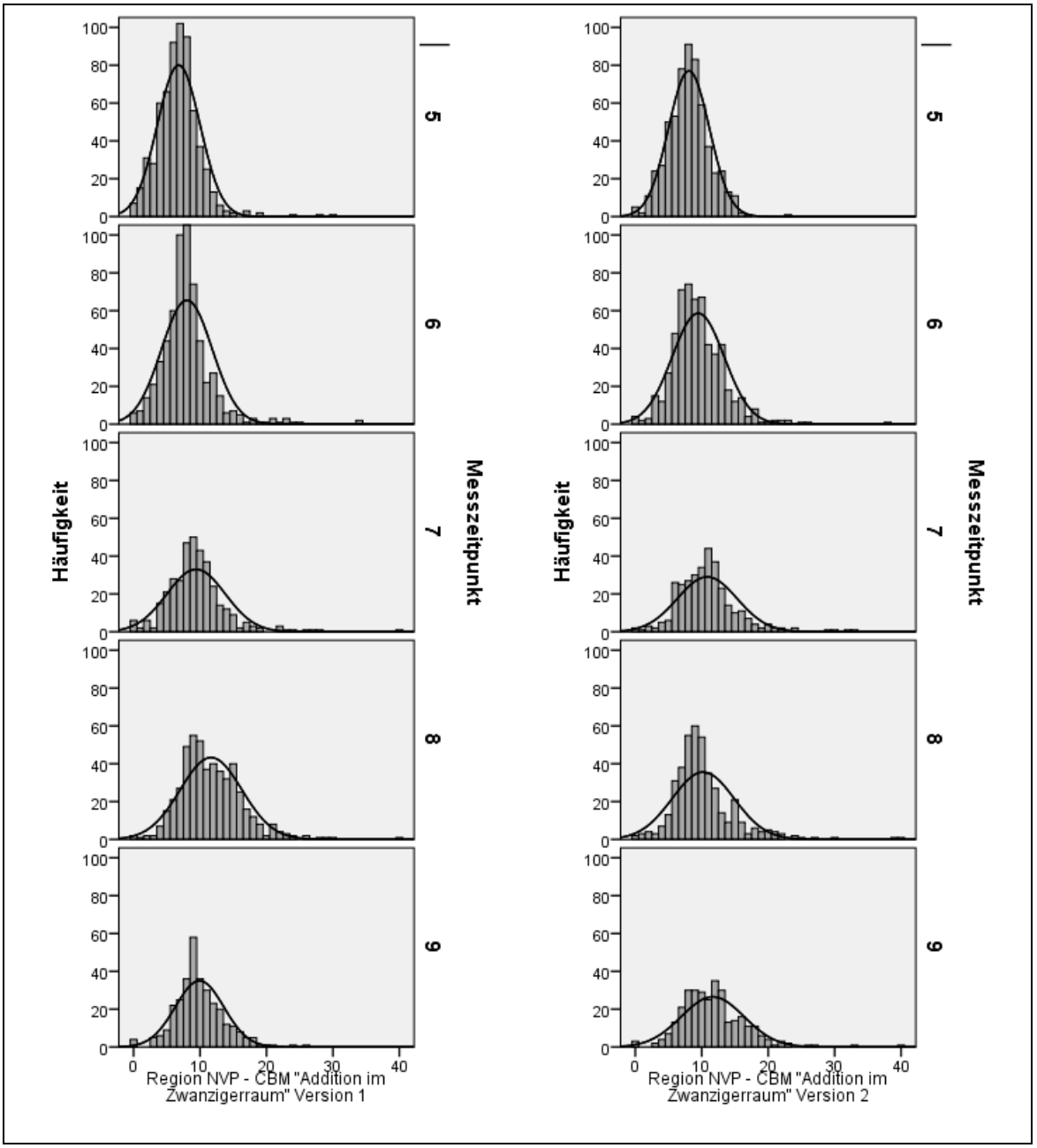


Abbildung 51: Rohwertverteilungen der Versionen 1 und 2 des CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

7.1.6.2 Untersuchungsgruppe Rügen

Für das CBM „Addition im Zwanzigerraum“ liegen zwischen 372 bis 432 gültige Werte von Schülern der Insel Rügen vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle oder ungültige Einträge. In Tabelle 30 sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Rügens für das erste Schuljahr tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM „Addition im Zwanzigerraum“ liegt zum ersten Messzeitpunkt in der fünften Schulwoche bei 5.28 Punkten mit einer Standardabweichung von 2.93 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich an (MZP 6: 7.28 Punkte; MZP 7: 8.9 Punkte; MZP 8: 8.9 Punkte; MZP 9: 9.78 Punkte). Die Standardabweichung bleibt über alle Messungen hinweg in einem Intervall von 2.93 und 3.73 Punkten recht stabil. Zu allen Messzeitpunkten hat mindestens ein Kind keinen Punkt erzielen können (Minimum: 0 Punkte). Die Werte für das Maximum variieren in einem Intervall von 19 und 25 Punkten über die Messzeitpunkte.

Tabelle 30: Deskriptive Statistiken des CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen

MZP	CBM „Addition im Zwanzigerraum“				
	N	M	SD	Min	Max
5 (21.SW)	417	5.28	2.93	0	19
6 (26.SW)	432	7.28	3.59	0	20
7 (30.SW)	431	8.90	3.29	0	20
8 (34.SW)	372	8.90	3.27	0	22
9 (38.SW)	430	9.78	3.73	0	25

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 52 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Addition im Zwanzigerraum“ der Kinder der Region Rügen über die fünf Messzeitpunkte grafisch dargestellt. Die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte lässt sich hier erkennen.

Zu jedem der Messzeitpunkte erscheinen die Daten normalverteilt zu sein.

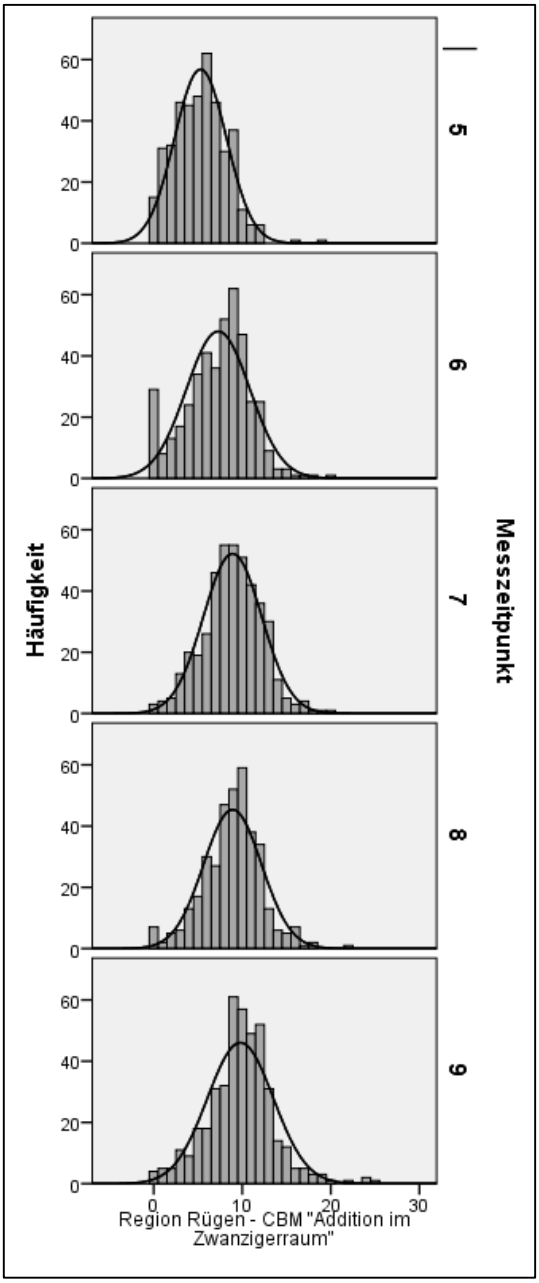


Abbildung 52: Rohwertverteilungen des CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen

7.1.7 CBM „Subtraktion im Zwanzigerraum“

7.1.7.1 Untersuchungsgruppe Nordvorpommern

Für das CBM „Subtraktion im Zwanzigerraum“ liegen für die beiden zu den jeweiligen Messzeitpunkten durchgeführten Versionen 1 und 2 zwischen 315 bis 644 gültige Werte von Schülern aus der Region NVP vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle, ungültige Einträge und schulorganisatorische Umstände. Zum Ende des Schuljahres nahm die Bereitschaft zur Teilnahme an den curriculumbasierten Messungen in der Region Nordvorpommern deutlich ab, da die monatlichen Messungen doch deutlichen Aufwand für die Lehrkräfte bedeuteten. Dies scheint ebenso der Grund zu sein, weshalb für die Version 2 deutlich weniger Ergebnisse vorliegen. In Tabelle 31 sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Nordvorpommerns für das erste Schuljahr getrennt nach den Versionen tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM der Version 1 liegt zum fünften Messzeitpunkt (21. Schulwoche) bei 4.95 Punkten mit einer Standardabweichung von 2.52 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich (MZP 6: 6.33 Punkte; MZP 7: 7.06 Punkte; MZP 8: 7.94 Punkte; MZP 9: 8.05 Punkte). Die Standardabweichung variiert über die Messzeitpunkte hinweg mit Werten zwischen 2.52 und 4.5 Punkten. Zu jedem der Messzeitpunkte hat mindestens ein Kind keinen Punkt erzielt (Minimum: 0 Punkte). Die maximal erreichten Werte schwanken über die Messzeitpunkte in einem Bereich von 15 und 34 Punkten. Die Verteilungen der Version 2 zeigen ein ähnliches Bild. Im Laufe der Schulwochen erhöht sich auch hier der Mittelwert der Verteilungen kontinuierlich (MZP 5: 5.17 Punkte; MZP 6: 6.48 Punkte; MZP 7: 7.38 Punkte; MZP 8: 8.12 Punkte; MZP 9: 8.46 Punkte). Die Standardabweichung für die Version 2 weist Werte in einem Intervall

von 2.68 und 4.64 Punkten über die Messzeitpunkte auf. Das erzielte Maximum variiert über die Messzeitpunkt in einem Intervall von 15 bis 33 Punkten. Es gibt zu jeder Messung jeweils Schüler, die einen Wert entsprechend des Minimums von 0 Punkten erzielten. Insgesamt lässt sich festhalten, dass die Mittelwerte und Standardabweichungen der Versionen 1 und 2 sich zu jedem Zeitpunkt ähneln, die Ergebnisse der Version 2 liegen jedoch durchweg etwas höher.

Tabelle 31: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	CBM „Subtraktion im Zwanzigerraum“ Version 1					CBM „Subtraktion im Zwanzigerraum“ Version 2				
	N_{V1}	M_{V1}	SD_{V1}	Min_{V1}	Max_{V1}	N_{V2}	M_{V2}	SD_{V2}	Min_{V2}	Max_{V2}
5 (21.SW)	644	4.95	2.52	0	15	596	5.17	2.68	0	15
6 (26.SW)	609	6.33	3.44	0	28	557	6.48	3.55	0	25
7 (30.SW)	362	7.06	4.24	0	24	328	7.38	4.09	0	26
8 (34.SW)	528	7.94	4.50	0	34	438	8.12	4.61	0	33
9 (38.SW)	320	8.05	4.40	0	31	315	8.46	4.64	0	32

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 53 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Subtraktion im Zwanzigerraum“ der Kinder der Region Nordvorpommern über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres grafisch dargestellt. Links sind die Verteilungen der Version 1, rechts die Verteilungen der parallelen Version 2 abgebildet. In den Grafiken lässt sich die zuvor erwähnte kontinuierliche Verschiebung der Mittelwerte in den beiden Versionen leicht erkennen. Insgesamt erscheinen die Daten der Version 1 und Version 2 über die Messzeitpunkte hinweg optisch einer Normalverteilung zu entsprechen. Auch weisen die Verteilungen der Versionen 1 und 2 starke Übereinstimmungen auf. In Abschnitt 7.2.3.7 wird geprüft, ob diese Unterschiede tatsächlich nicht statistisch signifikant ausfallen.

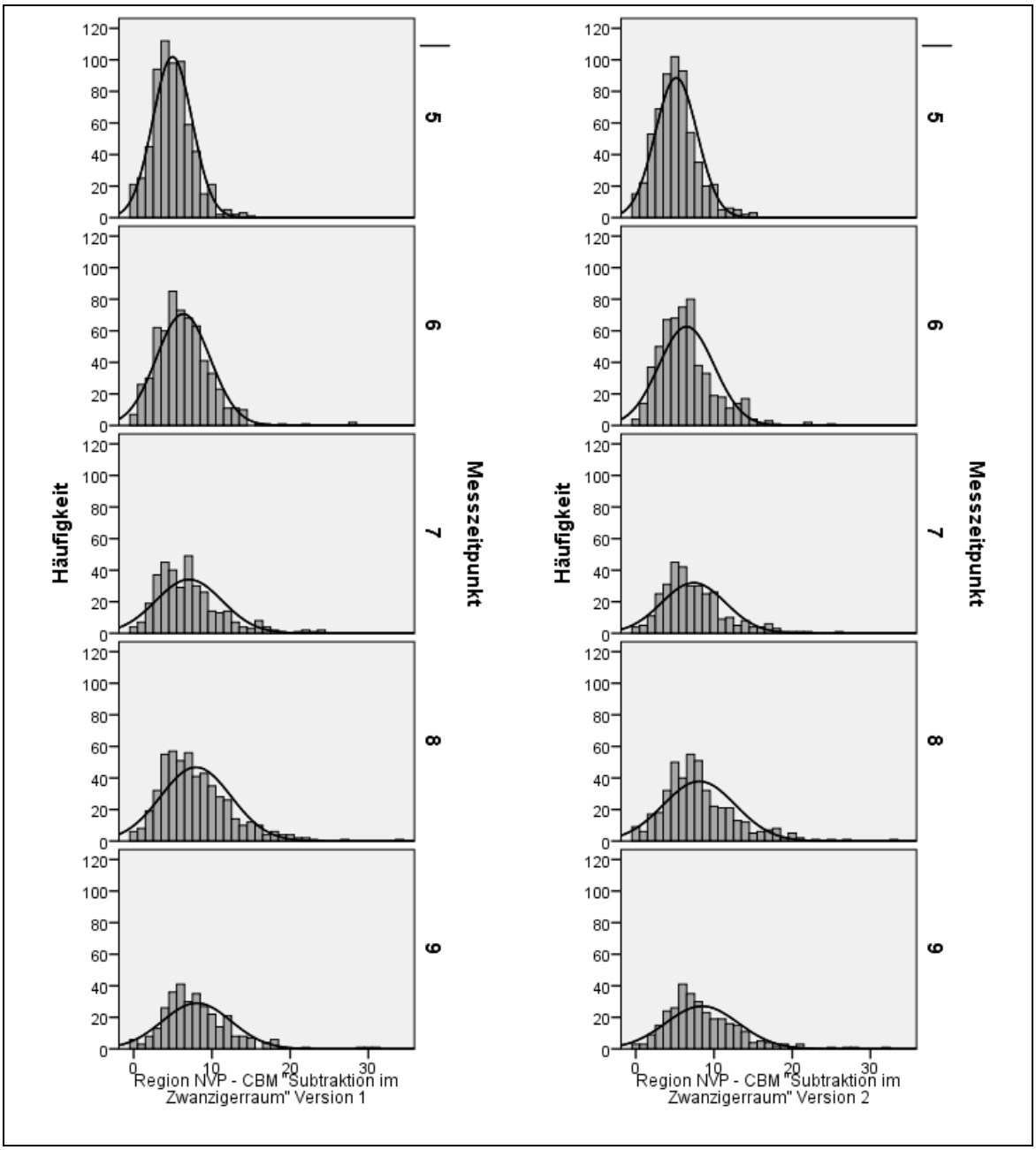


Abbildung 53: Rohwertverteilungen der Versionen 1 und 2 des CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

7.1.7.2 Untersuchungsgruppe Rügen

Für das CBM „Subtraktion im Zwanzigerraum“ liegen zwischen 372 bis 433 gültige Werte von Schülern der Insel Rügen vor. Die Anzahl der Schüler variiert durch z. B. krankheitsbedingte Ausfälle oder ungültige Einträge. In Tabelle 32 sind statistische Kennwerte (Mittelwert, Standardabweichung, Minimum und Maximum) der Schüler Rügens für das erste Schuljahr tabellarisch zusammengefasst.

Der Mittelwert der Rohwerte im CBM „Subtraktion im Zwanzigerraum“ liegt zum ersten Messzeitpunkt in der fünften Schulwoche bei 3.22 Punkten mit einer Standardabweichung von 2.73 Punkten. Über die Messzeitpunkte hinweg steigt der Mittelwert kontinuierlich (MZP 6: 4.91 Punkte; MZP 7: 6.22 Punkte; MTP 8: 6.83 Punkte; MZP 9: 6.96 Punkte). Die Standardabweichung bleibt über alle Messungen hinweg in einem Intervall von 2.73 und 3.61 Punkten recht stabil. Zu allen Messzeitpunkten hat mindestens ein Kind keinen Punkt erzielen können (Minimum: 0 Punkte). Die Werte für das Maximum variieren in einem Intervall von 20 und 26 Punkten über die Messzeitpunkte.

Tabelle 32: Deskriptive Statistiken des CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen

MZP	CBM „Subtraktion im Zwanzigerraum“				
	N	M	SD	Min	Max
5 (21.SW)	420	3.22	2.73	0	20
6 (26.SW)	433	4.91	3.14	0	26
7 (30.SW)	431	6.22	3.52	0	20
8 (34.SW)	372	6.83	3.31	0	20
9 (38.SW)	427	6.96	3.61	0	20

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max - Maximum

In Abbildung 54 sind die Verteilungen der erreichten Rohwertpunkte im CBM „Zahlzerlegung im Zwanzigerraum“ der Kinder der Region Rügen über die fünf Messzeitpunkte grafisch dargestellt. Die zuvor

erwähnte kontinuierliche Verschiebung der Mittelwerte lässt sich hier erkennen. Zu jedem der Messzeitpunkte erscheinen die Daten normalverteilt zu sein.

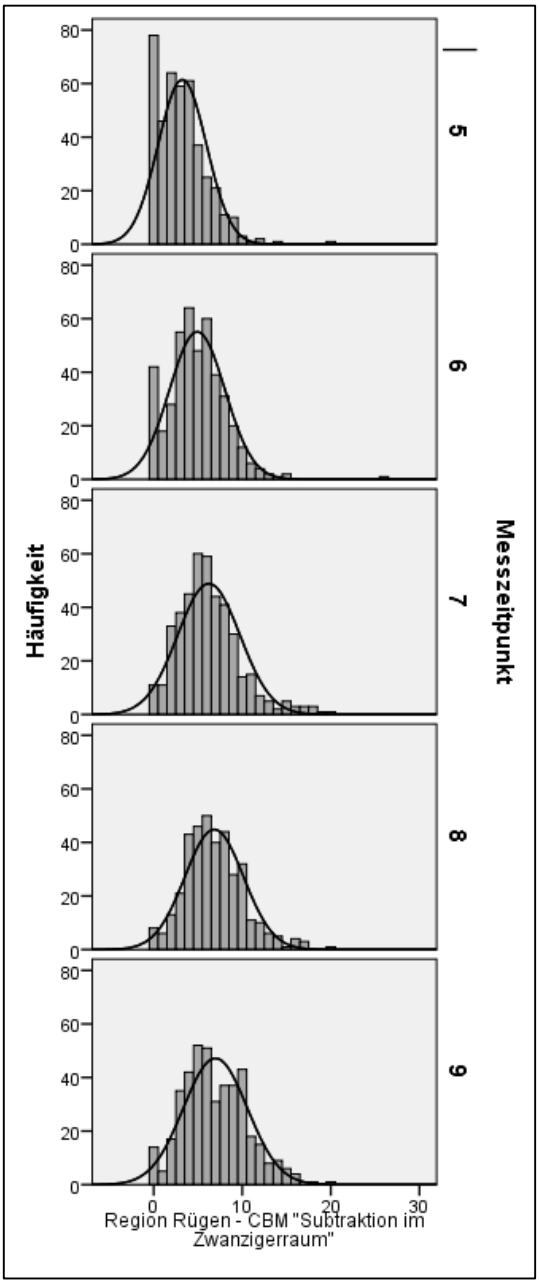


Abbildung 54: Rohwertverteilungen des CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen

7.1.8 Zusammenfassung der deskriptiven statistischen Angaben

Die Verteilungen des CBM „Zahlenlesen im Zwanzigerraum“ in den beiden Regionen Nordvorpommern und Rügen weisen augenscheinlich zunächst eine Normalverteilung der Daten auf, welche sich über die Messzeitpunkte hinweg nach rechts verschiebt, sodass zur Mitte des Schuljahres eine leichte Tendenz zu Deckeneffekten entsteht. Der Anteil an Kindern, welche die Maximalpunktzahl des CBM erreichen, ist zum letzten Messzeitpunkt in beiden Regionen recht hoch. Insgesamt nehmen die Mittelwerte der Schülergruppen über die Zeit kontinuierlich zu. Über alle Erhebungen im ersten Schulhalbjahr hinweg liegen die Ergebnisse der Rügener Schüler im Durchschnitt etwa zwei Rohwertpunkte unter den Mittelwerten der Kinder der Region Nordvorpommern.

Für das CBM „Zahlenreihe im Zwanzigerraum“ ergibt sich ein etwas anderes Bild. Während die Rügener Kinder im Durchschnitt auch hier Ergebnisse erzielen, die jeweils etwa zwei Rohwertpunkte unter den Mittelwerten der Schülergruppe Nordvorpommerns lag, zeichnet sich bei den Verteilungen in den Regionen zum ersten Messzeitpunkt zunächst ein Bodeneffekt ab. Kontinuierlich steigen auch hier die Mittelwerte über die Zeit an, sodass die Verteilungen in beiden Untersuchungskohorten zur Schuljahresmitte einer Normalverteilung ähneln.

Im CBM „Mengenvergleich im Zwanzigerraum“ erzielen die Kinder beider Regionen über alle Testzeitpunkte im Durchschnitt annähernd gleiche Werte. Zur ersten Erhebung sehen die Daten beider Untersuchungsgruppen normalverteilt aus, durch eine kontinuierliche Verschiebung der Daten nach rechts ergibt sich zur Schuljahreshälfte ein Deckeneffekt.

Die Verteilungen des CBM „Arithmetische Basiskompetenzen“ weisen auf einen Bodeneffekt zu Beginn der ersten Klasse hin, welcher sich durch Rechtsverschiebung der Daten zur Schuljahresmitte einer Normalverteilung angleicht. Zum Ende der Klasse eins deutet die Verteilung der Daten in beiden Regionen auf einen Deckeneffekt hin. Über die Zeit des Schuljahres vollzieht sich der Anstieg des Mittelwertes nicht kontinuierlich, so ist zu verschiedenen Messzeitpunkten eine Stagnation und sogar teilweise ein Rückgang des Mittelwertes zu beobachten. Im Durchschnitt erreicht die Rügener Schülerkohorte Werte die etwa zwei Punkte unter den Mittelwerten der Kinder Nordvorpommerns liegen.

Im CBM „Zahlzerlegung im Zwanzigerraum“ erzielen die Kinder beider Regionen über alle Testzeitpunkte im Durchschnitt annähernd gleiche Werte. Zur ersten Erhebung weisen die Daten beider Untersuchungsgruppen auf einen deutlichen Bodeneffekt hin, durch eine kontinuierliche Verschiebung der Daten nach rechts ergibt sich zur Schuljahreshälfte ein Deckeneffekt.

Die Verteilungen der CBM „Addition im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ deuten über alle Messzeitpunkte hinweg in beiden untersuchten Gruppen auf eine Normalverteilung der Daten hin, die sich im Laufe der Zeit kontinuierlich nach rechts verschiebt. Zu jedem Erhebungszeitpunkt gibt es in beiden Regionen Schüler, welche durch deutlich hohe Ergebnisse (Extremwerte) auffallen. In beiden CBM liegen die Durchschnittsergebnisse der Rügener Kinder durchweg um einen bis zwei Rohwerte unter denen der Kinder Nordvorpommerns.

7.2 Darstellung der Ergebnisse hinsichtlich der Reliabilität

Zur Schätzung der Reliabilität der innerhalb der vorliegenden Arbeit entwickelten CBM wird eine Analyse der Splithalf-Reliabilität

(Abschnitt 7.2.1), der Retest-Reliabilität (Abschnitt 7.2.2) sowie der Paralleltest-Reliabilität (Abschnitt 7.2.3) vorgenommen. Die Einschätzung der Reliabilität anhand der Bestimmung von Cronbachs α erscheint bei Tests, die eine zeitlich begrenzte Bearbeitungsdauer vorschreiben (Speedtest) eher nicht sinnvoll, da in der Regel nicht alle Items innerhalb des Bearbeitungszeitraumes gelöst werden können. Auf eine Bestimmung von Cronbachs α wird deshalb innerhalb der vorliegenden Arbeit verzichtet.

7.2.1 Splithalf-Reliabilität

In den folgenden Abschnitten werden die Ergebnisse hinsichtlich der Splithalf-Reliabilität der hier vorgestellten CBM der Niveaustufe 1 dargestellt. In diesem Rahmen wurden jeweils die Bearbeitungszeiten für die CBM halbiert. Die Kinder hatten somit für die Bearbeitung eines CBM zweimal eine halbe Minute Zeit. Es wurde jeweils die Anzahl richtiger Lösungen dokumentiert. Die Einschätzung der Splithalf-Reliabilitätskoeffizienten erfolgt auf Grundlage der Korrelation der Daten der beiden Bearbeitungszeiträume.

7.2.1.1 CBM „Zahlenlesen im Zwanzigerraum“

Die Ergebnisse der Analysen zur Splithalf-Reliabilität des CBM „Zahlenlesen im Zwanzigerraum“ sind in Tabelle 33 dargestellt. Es gingen Ergebnisse von 75 Schülern der Untersuchungsgruppe 4 (Erstklässler Rostocker Grundschulen) in die Berechnungen ein. Die Ergebnisse der Schüler innerhalb der beiden Bearbeitungszeiträume von jeweils 30 Sekunden ergeben hoch signifikante Korrelationen in einem Bereich von r = .88 bis r = .95. Die höchste Korrelation mit r = .95 ergibt sich für die Version 2 des CBM „Zahlenlesen im Zwanzigerraum“, die geringste Korrelation mit r = .88 weist die Version 4 auf. Die restlichen Versionen weisen Splithalf-Koeffizienten

innerhalb dieses Intervalls auf. Laut Bühl (2006) sind die sich ergebenden Koeffizienten als hoch bis sehr hoch einzuschätzen.

Tabelle 33: Ergebnisse der Analysen hinsichtlich der Splithalf-Reliabilität des CBM „Zahlenlesen im Zwanzigerraum“ in der Region Rostock

Version	N	M_1	SD_1	M_2	SD_2	r
1	75	15.91	7.69	14.31	7.30	.89**
2	75	17.43	7.30	14.55	6.67	.95**
3	75	16.72	8.21	15.27	6.70	.92**
4	75	16.63	7.29	14.53	7.53	.88**
5	75	17.03	7.77	14.77	7.42	.93**
6	75	16.85	7.29	15.64	7.34	.94**
7	75	16.35	7.22	14.48	6.45	.90**
8	75	17.31	7.23	14.79	6.49	.88**

Erläuterungen: N – Stichprobenumfang; M_1 – Mittelwert richtiger Ergebnisse der ersten 30s Bearbeitungszeit; SD_1 – Standardabweichung richtiger Ergebnisse der ersten 30s Bearbeitungszeit; M_2 - Mittelwert richtiger Ergebnisse der zweiten 30s Bearbeitungszeit; SD_2 - Standardabweichung richtiger Ergebnisse der zweiten 30s Bearbeitungszeit; r – Korrelationskoeffizient nach Spearman korrigiert nach Spearman-Brown; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig)

7.2.1.2 CBM „Zahlenreihe im Zwanzigerraum“

Die Ergebnisse der Analysen zur Splithalf-Reliabilität des CBM „Zahlenreihe im Zwanzigerraum“ sind in Tabelle 34 dargestellt. Es gingen Ergebnisse von 48 bis 68 Rostocker Erstklässler in die Berechnungen ein. Die Ergebnisse der Schüler innerhalb der beiden Bearbeitungszeiträume von jeweils 30 Sekunden ergeben hoch signifikante Korrelationen in einem Bereich von r = .77 bis r = .88. Die höchste Korrelation ergibt sich für die Version 4 des CBM „Zahlenreihe im Zwanzigerraum“, die geringste Korrelation weist die Version 5 auf. Die restlichen Versionen weisen Splithalf-Koeffizienten innerhalb dieses Intervalls auf. Laut Bühl (2006) sind die sich ergebenden Koeffizienten als hoch einzuschätzen.

Tabelle 34: Ergebnisse der Analysen hinsichtlich der Splithalf-Reliabilität des CBM „Zahlenreihe im Zwanzigerraum“ in der Region Rostock

Version	N	M_1	SD_1	M_2	SD_2	r
1	48	3.10	3.22	3.08	2.22	.79**
2	55	3.69	2.15	2.76	2.06	.84**
3	68	3.40	2.78	2.79	2.75	.83**
4	65	3.51	2.55	2.91	2.63	.88**
5	54	3.17	2.94	2.74	2.07	.77**
6	63	2.56	1.83	2.22	1.99	.77**
7	68	3.57	2.58	2.88	2.18	.78**
8	64	3.42	2.42	3.45	2.52	.86**

Erläuterungen: N – Stichprobenumfang; M_1 – Mittelwert richtiger Ergebnisse der ersten 30s Bearbeitungszeit; SD_1 – Standardabweichung richtiger Ergebnisse der ersten 30s Bearbeitungszeit; M_2 - Mittelwert richtiger Ergebnisse der zweiten 30s Bearbeitungszeit; SD_2 - Standardabweichung richtiger Ergebnisse der zweiten 30s Bearbeitungszeit; r – Korrelationskoeffizient nach Spearman korrigiert nach Spearman-Brown; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig)

7.2.1.3 CBM „Mengenvergleich im Zwanzigerraum“

Die Ergebnisse der Analysen zur Splithalf-Reliabilität des CBM „Mengenvergleich im Zwanzigerraum“ sind in Tabelle 35 dargestellt. Es gingen die Ergebnisse von 45 bis 69 Rostocker Erstklässler in die Berechnungen ein. Aufgrund eines Druckfehlers sind die Versionen 2 und 6 des CBM „Mengenvergleich im Zwanzigerraum“ nicht auswertbar, Analysen zur Splithalf-Reliabilität dieser Verfahren entfallen an dieser Stelle. Die Ergebnisse der Schüler innerhalb der beiden Bearbeitungszeiträume von jeweils 30 Sekunden ergeben fast ausschließlich hoch signifikante Korrelationen in einem Bereich von r = .44 bis r = .76. Die höchste Korrelation ergibt sich für die Version 5 des CBM „Mengenvergleich im Zwanzigerraum“ die geringste Korrelation weist die Version 8 auf. Die restlichen Versionen weisen Splithalf-Koeffizienten innerhalb dieses Intervalls auf. Laut Bühl (2006) streuen die sich ergebenden Koeffizienten in einem Bereich von gering bis hoch, wobei die Mehrheit der Koeffizienten im mittleren Bereich liegt. Lediglich für die Version 8 ergibt sich ein als gering einzustufender Splithalf-Koeffizient.

Tabelle 35: Ergebnisse der Analysen hinsichtlich der Splithalf-Reliabilität des CBM „Mengenvergleich im Zwanzigerraum“ in der Region Rostock

Version	N	M_1	SD_1	M_2	SD_2	r
1	45	8.60	3.65	9.96	4.39	.65**
2	-	-	-	-	-	-
3	68	11.28	5.24	10.53	5.20	.55**
4	61	12.13	5.74	10.89	4.40	.51**
5	55	8.00	4.55	9.44	5.05	.73**
6	-	-	-	-	-	-
7	69	11.75	4.22	12.93	5.17	.52**
8	58	14.28	3.99	12.53	3.33	.44*

Erläuterungen: N – Stichprobenumfang; M_1 – Mittelwert richtiger Ergebnisse der ersten 30s Bearbeitungszeit; SD_1 – Standardabweichung richtiger Ergebnisse der ersten 30s Bearbeitungszeit; M_2 - Mittelwert richtiger Ergebnisse der zweiten 30s Bearbeitungszeit; SD_2 - Standardabweichung richtiger Ergebnisse der zweiten 30s Bearbeitungszeit; r – Korrelationskoeffizient nach Spearman korrigiert nach Spearman-Brown; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig); * – Korrelation ist signifikant auf dem Niveau von 0.05 (2-seitig)

7.2.1.4 CBM „Arithmetische Basiskompetenzen“

Die Ergebnisse der Analysen zur Splithalf-Reliabilität des CBM „Arithmetische Basiskompetenzen“ sind in Tabelle 36 dargestellt. Es gingen die Ergebnisse von 33 bis 52 Rostocker Erstklässlern in die Berechnungen ein. Die Ergebnisse der Schüler innerhalb der beiden Bearbeitungszeiträume von jeweils 30 Sekunden ergeben fast ausschließlich hoch signifikante Korrelationen in einem Bereich von r = .35 bis r = .89. Die höchste Korrelation ergibt sich für die Version 9 des CBM „Arithmetische Basiskompetenzen“, die geringste Korrelation weist die Version 2 auf. Die restlichen Versionen weisen Splithalf-Koeffizienten innerhalb dieses Intervalls auf. Laut Bühl (2006) streuen die sich ergebenden Koeffizienten in einem Bereich von gering bis hoch, wobei die Mehrheit der Koeffizienten im mittleren Bereich liegt. Lediglich für die Version 2 ergibt sich ein als gering einzustufender Splithalf-Koeffizient.

Tabelle 36: Ergebnisse der Analysen hinsichtlich der Splithalf-Reliabilität des CBM „Arithmetische Basiskompetenzen“ in der Region Rostock

Version	N	M_1	SD_1	M_2	SD_2	r
1	42	2.67	1.95	1.69	1.52	.67**
2	42	2.24	1.41	1.43	1.76	.35
3	36	2.28	1.41	2.11	1.83	.52*
4	45	3.09	2.26	2.64	2.13	.58**
5	44	2.80	2.06	1.70	1.81	.80**
6	41	2.12	1.94	1.54	1.60	.56**
7	44	2.61	1.73	1.75	1.46	.63**
8	33	2.27	1.21	1.30	1.42	.71**
9	52	3.27	2.31	1.96	1.78	.89**
10	47	2.47	1.93	1.77	1.59	.72**

Erläuterungen: N – Stichprobenumfang; M_1 – Mittelwert richtiger Ergebnisse der ersten 30s Bearbeitungszeit; SD_1 – Standardabweichung richtiger Ergebnisse der ersten 30s Bearbeitungszeit; M_2 - Mittelwert richtiger Ergebnisse der zweiten 30s Bearbeitungszeit; SD_2 - Standardabweichung richtiger Ergebnisse der zweiten 30s Bearbeitungszeit; r – Korrelationskoeffizient nach Spearman korrigiert nach Spearman-Brown; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig); * – Korrelation ist signifikant auf dem Niveau von 0.05 (2-seitig)

7.2.2 Retest-Reliabilität

In den nachfolgenden Abschnitten werden die Ergebnisse hinsichtlich der Retest-Reliabilität der innerhalb der vorliegenden Arbeit konzipierten CBM dargestellt. In diesem Zusammenhang wurden einige Versionen der hier beschriebenen CBM wiederholt eingesetzt. Die Einschätzung der Retest-Reliabilitätskoeffizienten erfolgt auf Grundlage der Korrelation der Daten zu den verschiedenen Messzeitpunkten.

7.2.2.1 CBM „Zahlenlesen im Zwanzigerraum“

Da zum ersten als auch zum vierten Messzeitpunkt die gleichen Versionen des CBM „Zahlenlesen im Zwanzigerraum“ in der Region Nordvorpommern eingesetzt wurden, können Koeffizienten hinsichtlich der Retest-Reliabilität geschätzt werden. Aufgrund eines listenweisen Fallausschlusses fehlender Schülerdaten zu den beiden Messzeitpunkten (durch bspw. Krankheitsgründen) bilden die

Ergebnisse von 413 Schülern die Grundlage zur Errechnung der Retest-Reliabilität des CBM „Zahlenlesen im Zwanzigerraum“. Zwischen den CBM-Daten der beiden 11 Schulwochen auseinander liegenden Erhebungszeitpunkten ergibt sich ein als mittel einzuschätzender Korrelationskoeffizient von r = .58.

7.2.2.2 CBM „Zahlenreihe im Zwanzigerraum“

Zum ersten sowie zum vierten Messzeitpunkt wurden gleiche Versionen des CBM „Zahlenreihe im Zwanzigerraum“ in der UG 2 eingesetzt. Auf Grundlage der Ergebnisse zu den beiden Zeitpunkten können Koeffizienten hinsichtlich der Retest-Reliabilität für dieses Verfahren geschätzt werden. Aufgrund eines listenweisen Fallausschlusses fehlender Schülerergebnisse zu den beiden Messzeitpunkten bilden lediglich die Daten von 424 Schülern die Grundlage zur Errechnung der Retest-Reliabilität des CBM „Zahlenreihe im Zwanzigerraum“. Zwischen den CBM-Daten der beiden 11 Schulwochen auseinander liegenden Erhebungszeitpunkten ergibt sich ein als mittel einzuschätzender Korrelationskoeffizient von r = .53.

7.2.2.3 CBM „Mengenvergleich im Zwanzigerraum“

Auch vom CBM „Mengenvergleich im Zwanzigerraum“ wurden zum ersten und vierten Messzeitpunkt die gleichen Versionen in der Region Nordvorpommern eingesetzt, wodurch Koeffizienten hinsichtlich der Retest-Reliabilität geschätzt werden können. Aufgrund eines listenweisen Fallausschlusses fehlender Schülerdaten zu den beiden Messzeitpunkten bilden lediglich die Ergebnisse von 424 Schülern die Grundlage zur Errechnung der Retest-Reliabilität des CBM „Mengenvergleich im Zwanzigerraum“. Zwischen den CBM-Daten der beiden 11 Schulwochen auseinander liegenden Erhebungszeitpunkten ergibt sich ein geringer Korrelationskoeffizient von r = .33.

7.2.2.4 CBM „Arithmetische Basiskompetenzen“

Da zu verschiedenen Messzeitpunkten jeweils die gleichen Versionen des CBM „Arithmetische Basiskompetenzen“ eingesetzt wurden, können verschiedene Koeffizienten hinsichtlich der Retest-Reliabilität geschätzt werden. In Tabelle 37 sind Ergebnisse der Analysen hinsichtlich der Retest-Reliabilität des CBM „Arithmetische Basiskompetenzen“ in der Region Nordvorpommern zusammengefasst. Aufgrund des listenweisen Fallausschlusses fehlender Schülerergebnisse bilden die Daten von jeweils verschieden mächtigen Schülergruppen die Grundlage zur Errechnung der Retest-Reliabilität des CBM. Es ergeben sich durchweg als mittel bis eher hoch einzuschätzende Korrelationskoeffizienten im Bereich zwischen r = .54 und r = .69, wobei auch die zeitlichen Abstände zwischen den Messungen in einem Intervall von 5 bis 25 Schulwochen variieren.

Tabelle 37: Ergebnisse der Analysen hinsichtlich der Retest-Reliabilität des CBM „Arithmetische Basiskompetenzen“ in der Region Nordvorpommern

eingesetzte CBM-Version in der Region NVP	eingesetzt zu den MZP	zeitlicher Abstand zwischen den MZP in SW	N	r
CBM1	1↔4	11	424	.51**
	1↔5	16	449	.57**
	4↔5	5	417	.67**
CBM2	1↔7	25	213	.59**
CBM3	2↔5	13	395	.69**
CBM4	2↔6	16	336	.68**
CBM5	3↔6	14	378	.67**
CBM6	3↔7	18	269	.52**
CBM7	4↔8	22	326	.61**

Erläuterungen: MZP – Messzeitpunkt; NVP – Nordvorpommern; SW – Schulwochen; N – Stichprobenumfang; r – Korrelationskoeffizient nach Pearson; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig)

7.2.2.5 CBM „Zahlzerlegung im Zwanzigerraum“

Jeweils zum fünften und neunten (jeweils Version 1) als auch zum sechsten und neunten (jeweils Version 2) Messzeitpunkt wurden in der Region Nordvorpommern gleiche Versionen des CBM „Zahlzerlegung im Zwanzigerraum“ eingesetzt, wodurch Koeffizienten hinsichtlich der Retest-Reliabilität geschätzt werden können. Aufgrund eines listenweisen Fallausschlusses fehlender Schülerergebnisse bilden lediglich die Daten von 277 bzw. 284 Schülern die Grundlage zur Errechnung der Retest-Reliabilität des CBM „Mengenvergleich im Zwanzigerraum“. Zwischen den CBM-Daten der 17 bzw. 12 Schulwochen auseinander liegenden Erhebungszeitpunkten ergeben sich jeweils als mittel einzuschätzende Korrelationskoeffizienten von r = .60 bzw. r = .54.

7.2.2.6 CBM „Addition im Zwanzigerraum“

Da auch vom CBM „Addition im Zwanzigerraum“ zum fünften und neunten (jeweils Version 1) sowie zum sechsten und neunten (jeweils Version 2) Messzeitpunkt in der UG 2 gleiche Versionen eingesetzt

wurden, können Koeffizienten hinsichtlich der Retest-Reliabilität geschätzt werden. Da nur Daten von Kindern mit vollständigen Fällen in die Analyse eingingen, bilden die Ergebnisse von 291 bzw. 294 Schülern die Grundlage zur Errechnung der Retest-Reliabilität des CBM „Addition im Zwanzigerraum“. Zwischen den CBM-Daten der 17 bzw. 12 Schulwochen auseinander liegenden Erhebungszeitpunkten ergeben sich jeweils als mittel einzuschätzende Korrelationskoeffizienten von r = .54 bzw. r = .60.

7.2.2.7 CBM „Subtraktion im Zwanzigerraum“

In der Region Nordvorpommern wurden zum fünften und neunten (jeweils Version 1) als auch zum sechsten und neunten (jeweils Version 2) Messzeitpunkt jeweils gleiche Versionen des CBM „Zahlzerlegung im Zwanzigerraum“ eingesetzt. Aufgrund dessen können Koeffizienten hinsichtlich der Retest-Reliabilität für das Verfahren geschätzt werden. Durch den listenweisen Fallausschlusses fehlender Schülerdaten bilden lediglich die Ergebnisse von 289 bzw. 296 Schülern die Grundlage zur Errechnung der Retest-Reliabilität des CBM „Mengenvergleich im Zwanzigerraum“. Zwischen den CBM-Daten der 17 bzw. 12 Schulwochen auseinander liegenden Erhebungszeitpunkten ergeben sich als mittel bzw. hoch einzuschätzende Korrelationskoeffizienten von r = .53 bzw. r = .72.

7.2.3 Paralleltest-Reliabilität

In den nachfolgenden Abschnitten werden Kennwerte der Parallelität der innerhalb der vorliegenden Arbeit konzipierten CBM dargestellt. In diesem Zusammenhang wurden in der Region Nordvorpommern zu jedem Messzeitpunkt jeweils zwei verschiedene, als parallel angenommene Versionen eines jeden CBM durchgeführt.

7.2.3.1 CBM „Zahlenlesen im Zwanzigerraum“

Um die Paralleltest-Reliabilität der CBM „Zahlenlesen im Zwanzigerraum“ abzuschätzen, wurden zunächst die Korrelationen der jeweils zeitgleich zu den Messzeitpunkten durchgeführten CBM-Versionen berechnet. Die Ergebnisse sind in Tabelle 38 dargestellt. Durch einen listenweisen Fallausschluss der fehlenden Werte variiert die Anzahl gültiger Fälle für die Rechnungen zwischen 318 und 431 Schülern. Es zeigen sich durchweg hohe bis sehr hohe Korrelationen in einem Intervall von r_{tt} = .86 und r_{tt} = .94. Darüber hinaus wurden t-Tests bei verbundenen Stichproben über die Messzeitpunkte hinweg durchgeführt. Die Ergebnisse der Tests sind ebenso in Tabelle 38 zusammengefasst. Trotz hoher Korrelationen zwischen den CBM-Versionen zeigen sich, den dritten Messzeitpunkt ausgenommen, jeweils signifikante Unterschiede ($p < .001$ bzw. $p < .05$ zum MZP 4). Die Mittelwerte der jeweils zweiten eingesetzten Version fielen bei jeder Erhebung höher aus. Um zu prüfen, welche praktische Relevanz die signifikanten Mittelwertunterschiede haben, werden die Effektstärken (Cohens d) berechnet, welche in der letzten Spalte der Tabelle aufgeführt sind. Zum ersten und zweiten Messzeitpunkt haben die sich ergebenden signifikanten Mittelwertdifferenzen einen kleinen Effekt (MZP 1: d = 0.41; MZP 2: d = 0.24). Zum letzten Messzeitpunkt ist nach der Einteilung von Cohen (1988) keine praktische Relevanz gegeben (d = 0.15).

Tabelle 38: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Zahlenlesen im Zwanzigerraum“ zu den vier Messzeitpunkten in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	N	CBM „Zahlenlesen im Zwanzigerraum“ Version 1		CBM „Zahlenlesen im Zwanzigerraum“ Version 2		r_{tt}	t-Test p (2-seitig)	d
		M_{V1}	SD_{V1}	M_{V2}	SD_{V2}			
1 (5.SW)	318	30.00	14.17	32.87	14.09	.88**	.000	0.41
2 (8.SW)	333	36.58	14.96	37.91	14.86	.93**	.000	0.24
3 (12.SW)	329	40.13	14.70	40.31	14.82	.86**	.677	-
4 (16.SW)	431	44.50	14.16	45.25	14.14	.94**	.002	0.15

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; r_{tt} – Koeffizient zur Schätzung der Paralleltest-Reliabilität; d – Effektstärke; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig).

7.2.3.2 CBM „Zahlenreihe im Zwanzigerraum“

Auch zur Abschätzung der Paralleltest-Reliabilität für das CBM „Zahlenreihe im Zwanzigerraum“ wurden zunächst die Korrelationen der jeweils zeitgleich durchgeführten CBM-Versionen berechnet. Die Ergebnisse sind in Tabelle 39 dargestellt. Durch einen listenweisen Fallausschluss der fehlenden Werte verringert sich die Anzahl gültiger Fälle für die Rechnungen (N zwischen 333 und 469). Es zeigen sich durchweg hohe bis sehr hohe Korrelationen in einem Intervall von $r_{tt} = .80$ und $r_{tt} = .90$. Zur Überprüfung, inwieweit die beiden Verteilungen vergleichbar sind, wurden t-Tests bei verbundenen Stichproben über die Messzeitpunkte hinweg durchgeführt. Aufgrund der fehlenden Normalverteilung der Daten beider Versionen zum ersten Messzeitpunkt wurde hier der u-Test nach Mann und Whitney angewendet. Die Ergebnisse der Tests sind ebenso in Tabelle 39 zusammengefasst. Trotz hoher Korrelationen zwischen den CBM-Versionen zeigen sich, den dritten Messzeitpunkt ausgenommen, jeweils signifikante Unterschiede ($p < .001$). Die Mittelwerte der jeweils zweiten eingesetzten Version fielen, bis auf den dritten Messzeitpunkt, bei den Erhebungen höher aus. Um zu prüfen, welche praktische Relevanz die signifikanten Mittelwertunterschiede haben, werden auch in diesem Fall die

Effektstärken (Cohens d) berechnet. Vom ersten bis dritten Messzeitpunkt haben diese Mittelwertdifferenzen einen nach Cohen (1992) als gering einzustufenden Effekt (MZP 1: d = 0.48; MZP 2: d = 0.24; MZP 3: d = 0.29). Zum letzten Messzeitpunkt ist nach der Einteilung von Cohen (1988) keine praktische Relevanz gegeben (d = 0.17).

Tabelle 39: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Zahlenreihe im Zwanzigerraum“ zu den vier Messzeitpunkten in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	N	CBM „Zahlenreihe im Zwanzigerraum“ Version 1		CBM „Zahlenreihe im Zwanzigerraum“ Version 2		r_{tt}	t-Test/u-Test p (2-seitig)	d
		M_{V1}	SD_{V1}	M_{V2}	SD_{V2}			
1 (5.SW)	333	4.22	4.12	5.49	4.18	.80**	.000[1)]	0.48
2 (8.SW)	391	7.65	5.40	8.22	5.30	.90**	.000	0.24
3 (12.SW)	387	10.57	5.69	9.70	5.74	.86**	.000	0.28
4 (16.SW)	469	11.41	5.65	11.94	5.51	.85**	.000	0.17

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; r_{tt} – Koeffizient zur Schätzung der Paralleltest-Reliabilität; d – Effektstärke; [1)] - Ergebnis des u-Tests; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig).

7.2.3.3 CBM „Mengenvergleich im Zwanzigerraum“

Die Korrelationen der jeweils zeitgleich zu den Messzeitpunkten durchgeführten Versionen des CBM „Mengenvergleich im Zwanzigerraum“ zur Einschätzung der Paralleltest-Reliabilität sind in Tabelle 40 dargestellt. Durch Ausschluss von unvollständigen Schülerdatensätzen über die Testzeitpunkte variiert die Anzahl untersuchter Fälle zwischen 351 und 473. Es ergeben sich durchweg hohe Korrelationen in einem Intervall zwischen r_{tt} = .74 und r_{tt} = .76. Ebenso wurden t-Tests bei verbundenen Stichproben über die Messzeitpunkte bzw. ein u-Test zum ersten Messzeitpunkt (aufgrund nicht normalverteilter Daten beider Versionen) durchgeführt, deren Ergebnisse ebenso in Tabelle 40 zusammengefasst sind. Trotz hoher Korrelationen zwischen den CBM-Versionen zeigen sich jeweils

signifikante Unterschiede ($p < .001$ bzw. $p < .05$ zum MZP 4). Zum ersten Messzeitpunkt fällt dieser Unterschied sehr deutlich aus. Während die Effektstärke zum ersten Messzeitpunkt mit $d = 0.74$ auf einen mittleren Effekt hinweist, besteht für die sich ergebenden signifikanten Mittelwertdifferenzen zum zweiten und dritten Messzeitpunkt ein kleiner Effekt (MZP 1: $d = 0.41$; MZP 2: $d = 0.26$). Zum letzten Messzeitpunkt ist nach der Einteilung von Cohen (1992) keine praktische Relevanz gegeben ($d = 0.11$).

Tabelle 40: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Mengenvergleich im Zwanzigerraum“ zu den vier Messzeitpunkten in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	N	CBM „Mengenvergleich im Zwanzigerraum“ Version 1		CBM „Mengenvergleich im Zwanzigerraum“ Version 2		r_{tt}	t-Test/ u-Test p (2-seitig)	d
		M_{V1}	SD_{V1}	M_{V2}	SD_{V2}			
1 (5.SW)	351	16.28	8.89	20.93	8.43	.74**	.000	0.74
2 (8.SW)	391	22.83	7.85	25.00	7.21	.76**	.000	0.41
3 (12.SW)	389	26.78	7.44	28.01	6.25	.77**	.000[1)]	0.26
4 (16.SW)	473	28.82	5.541	29.22	5.07	.76**	.010[1)]	0.11

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; r_{tt} – Koeffizient zur Schätzung der Paralleltest-Reliabilität; d – Effektstärke; [1)] - Ergebnis des u-Tests; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig).

7.2.3.4 CBM „Arithmetische Basiskompetenzen“

In Tabelle 41 sind die Ergebnisse der Analysen hinsichtlich der Paralleltest-Reliabilität des CBM „Arithmetische Basiskompetenzen“ dargestellt. In einem ersten Schritt wurden die Korrelationen der jeweils zeitgleich zu den Messzeitpunkten durchgeführten CBM-Versionen berechnet. Dabei variiert die Anzahl der in die Analysen einbezogenen Schülerdaten zwischen 292 und 576 aufgrund eines listenweisen Fallausschlusses fehlender Werte. Es zeigen sich durchweg hohe bis sehr hohe Korrelationen in einem Intervall von

r_{tt} = .65 und r_{tt} = .82. Darüber hinaus wurden t-Tests bei verbundenen Stichproben über die Messzeitpunkte hinweg durchgeführt. Die Ergebnisse der Tests sind ebenso in Tabelle 41 aufgeführt. Trotz hoher Korrelationen zwischen den CBM-Versionen zeigen sich, ausgenommen zum ersten Messzeitpunkt, jeweils signifikante Unterschiede ($p < .001$ bzw. $p < .05$ zum MZP 5). Auffällig ist, dass die Schüler zu den Messzeitpunkten in der ersten Schuljahreshälfte durchweg höhere Mittelwerte in der jeweils zweiten eingesetzten Version erzielten. In der zweiten Schuljahreshälfte wechselt sich das Verhältnis, hier erreichen die Kinder jeweils in der ersten eingesetzten Version die höheren Mittelwerte. Bei den sich ergebenden signifikanten Mittelwertunterschieden vom zweiten bis zum neunten Messzeitpunkt weisen die Effektstärken zum zweiten, dritten, achten und neunten Messzeitpunkt mit d = 0.40 und d = 0.47 nach Cohen (1992) auf geringe Effekte der Unterschiede hin. Zum sechsten und siebten Messzeitpunkt ergeben sich mit d = 0.58 und d = 0.52 mittlere Effekte. Zum vierten und fünften Messzeitpunkt ist nach der Einteilung von Cohen (1992) keine praktische Relevanz der Mittelwertunterschiede gegeben (d = 0.18 bzw. d = 0.11).

Tabelle 41: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Arithmetische Basiskompetenzen“ zu den neun Messzeitpunkten im ersten Schuljahr in der Region Nordvorpommern

MZP	N	CBM „Arithmetische Basiskompetenzen“ Version 1		CBM „Arithmetische Basiskompetenzen“ Version 2		r_{tt}	t-Test p (2-seitig)	d
		M_{V1}	SD_{V1}	M_{V2}	SD_{V2}			
1 (5.SW)	335	3.68	2.93	3.89	2.43	.73**	.065	-
2 (8.SW)	386	5.69	3.59	6.78	3.86	.79**	.000	0.45
3 (12.SW)	385	6.90	3.57	7.94	4.25	.81**	.000	0.42
4 (16.SW)	442	8.17	3.84	8.77	4.22	.65**	.000	0.18
5 (21.SW)	576	9.14	4.09	8.77	4.90	.74**	.003	0.11
6 (26.SW)	551	10.83	3.97	9.48	3.71	.82**	.000	0.58
7 (30.SW)	321	10.68	4.20	8.98	4.22	.70*	.000	0.52
8 (34.SW)	409	11.06	3.91	9.73	3.73	.73**	.000	0.47
9 (38.SW)	292	10.54	3.80	9.39	4.25	.76**	.000	0.40

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; r_{tt} – Koeffizient zur Schätzung der Paralleltest-Reliabilität; d – Effektstärke; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig).

7.2.3.5 CBM „Zahlzerlegung im Zwanzigerraum“

Um die Paralleltest-Reliabilität des CBM „Zahlzerlegung im Zwanzigerraum“ abzuschätzen, wurden zunächst die Korrelationen der jeweils zeitgleich zu den Messzeitpunkten durchgeführten CBM-Versionen berechnet, welche in Tabelle 42 dargestellt sind. Die Anzahl der in die Analysen einbezogenen Fälle variiert zwischen 300 und 591 Schülern. Es konnten durchweg hohe bis sehr hohe Korrelationen in einem Intervall von $r_{tt} = .67$ und $r_{tt} = .78$ ermittelt werden. Die Ergebnisse der t-Tests deuten bis auf zum siebten und achten Messzeitpunkt auf keine signifikanten Unterschiede zwischen den Mittelwerten der jeweiligen CBM-Versionen hin. Während die Schüler zum siebten Messzeitpunkt höhere Mittelwerte in der ersten eingesetzten Version erzielten, wechselt sich das Verhältnis zum achten Messzeitpunkt. Hier erreichen die Kinder in der zweiten eingesetzten Version die höheren Mittelwerte. Der signifikante Mittelwertunterschied zum siebten Messzeitpunkt deutet mit d = 0.53 auf einen mittleren Effekt hin. Der Unterschied zwischen den CBM-

Versionen zum achten Messzeitpunkt weist mit der Effektstärke von d = 0.21 einen geringen Effekt auf.

Tabelle 42: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Zahlzerlegung im Zwanzigerraum“ zu den fünf Messzeitpunkten in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	N	CBM „Zahlzerlegung im Zwanzigerraum“ Version 1		CBM „Zahlzerlegung im Zwanzigerraum“ Version 2		r_{tt}	t-Test p (2-seitig)	d
		M_{V1}	SD_{V1}	M_{V2}	SD_{V2}			
5 (21.SW)	591	4.67	3.34	4.83	3.06	.67**	.142	-
6 (26.SW)	556	5.32	3.59	5.47	5.53	.74**	.147	-
7 (30.SW)	323	7.84	4.13	6.39	4.09	.78**	.000	0.53
8 (34.SW)	413	7.11	4.42	7.82	4.46	.72**	.000	0.21
9 (38.SW)	300	7.84	4.28	7.79	4.11	.73**	.766	-

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; r_{tt} – Koeffizient zur Schätzung der Paralleltest-Reliabilität; d – Effektstärke; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig).

7.2.3.6 CBM „Addition im Zwanzigerraum“

Auch zur Abschätzung der Paralleltest-Reliabilität für das CBM „Addition im Zwanzigerraum“ wurden zunächst die Korrelationen der jeweils zeitgleich durchgeführten CBM-Versionen berechnet. Die Ergebnisse sind in Tabelle 43 Tabelle 39dargestellt. Durch einen listenweisen Fallausschluss der fehlenden Werte verringert sich die Anzahl gültiger Fälle für die Rechnungen (N zwischen 314 und 591). Es zeigen sich durchweg hohe Korrelationen in einem Intervall von r_{tt} = .66 und r_{tt} = .79. Dennoch deuten die Ergebnisse der t-Tests jeweils auf signifikante Unterschiede (p < .001) zwischen den Verteilungen jeweils zeitgleich erhobener CBM-Versionen hin. Die Mittelwerte der jeweils zweiten eingesetzten Version fielen, mit Ausnahme zum achten Messzeitpunkt, bei jeder Erhebung höher aus. Bei dem sich ergebenden signifikanten Mittelwertunterschied zum sechsten und achten Messzeitpunkt ergeben sich mit d = 0.49 bzw.

d = 0.42 nach Cohen (1992) geringe Effekte. Die Unterschiede zwischen den CBM-Versionen zum fünften, siebten und neunten Messzeitpunkt weisen mit Effektstärken von d = 0.50, d = 0.54 sowie d = 0.50 auf mittlere Effekte hin.

Tabelle 43: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Addition im Zwanzigerraum“ zu den fünf Messzeitpunkten in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	N	CBM „Addition im Zwanzigerraum“ Version 1		CBM „Addition im Zwanzigerraum“ Version 2		r_{tt}	t-Test p (2-seitig)	d
		M_{V1}	SD_{V1}	M_{V2}	SD_{V2}			
5 (21.SW)	591	6.75	3.30	8.07	3.09	.66**	.000	0.50
6 (26.SW)	572	8.04	3.91	9.48	3.93	.72**	.000	0.49
7 (30.SW)	324	9.21	4.52	10.83	4.51	.78**	.000	0.54
8 (34.SW)	415	11.50	4.77	10.22	4.68	.79**	.000	0.42
9 (38.SW)	314	9.96	3.65	11.63	4.73	.71**	.000	0.50

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; r_{tt} – Koeffizient zur Schätzung der Paralleltest-Reliabilität; d – Effektstärke; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig).

7.2.3.7 CBM „Subtraktion im Zwanzigerraum“

Die Korrelationen der jeweils zeitgleich zu den Messzeitpunkten durchgeführten Versionen des CBM „Subtraktion im Zwanzigerraum“ zur Einschätzung der Paralleltest-Reliabilität sind in Tabelle 44 Tabelle 40dargestellt. Durch einen listenweisen Fallausschluss fehlender Schülerwerte über die Messzeitpunkte schwankt die Anzahl gültiger Fälle zwischen 315 und 591. Es zeigen sich durchweg hohe Korrelationen in einem Intervall von $r_{tt} = .67$ und $r_{tt} = .80$. Die Ergebnisse der t-Tests bei verbundenen Stichproben über die Messzeitpunkte hinweg sind ebenso in Tabelle 44 zusammengefasst. Trotz hoher Korrelationen zwischen den CBM-Versionen zeigen sich jeweils signifikante Unterschiede ($p < .05$ bzw. $p < .001$ zum MZP 7). Die Mittelwerte der jeweils zweiten eingesetzten Version fielen bei jeder Erhebung höher aus. Bei den sich ergebenden signifikanten

Mittelwertunterschieden über die Messzeitpunkte hinweg ergibt sich lediglich zum siebten Messzeitpunt mit d = 0.22 ein geringer Effekt, die Unterschiede zwischen den CBM-Versionen zu den anderen Messzeitpunkten erweisen sich mit Effektstärken im Bereich zwischen d = 0.10 und d = 0.17 als nicht relevant.

Tabelle 44: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Subtraktion im Zwanzigerraum“ zu den fünf Messzeitpunkten in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern

MZP	N	CBM „Subtraktion im Zwanzigerraum“ Version 1		CBM „Subtraktion im Zwanzigerraum“ Version 2		r_{tt}	t-Test p (2-seitig)	d
		M_{V1}	SD_{V1}	M_{V2}	SD_{V2}			
5 (21.SW)	591	4.96	2.53	5.17	2.68	.67**	.014	0.10
6 (26.SW)	537	6.26	3.43	6.55	3.57	.81**	.003	0.13
7 (30.SW)	323	6.75	4.11	7.35	4.08	.77**	.000	0.22
8 (34.SW)	437	7.64	4.42	8.13	4.61	.80**	.001	0.17
9 (38.SW)	315	8.09	4.40	8.46	4.64	.80**	.022	0.13

Erläuterungen: MZP – Messzeitpunkt; SW – Schulwoche; N – Stichprobenumfang; r_{tt} – Koeffizient zur Schätzung der Paralleltest-Reliabilität; d – Effektstärke; ** – Korrelation ist signifikant auf dem Niveau von 0.01 (2-seitig).

7.3 Darstellung der Ergebnisse hinsichtlich der Validität

Im nachfolgenden Abschnitt werden Aspekte der Validität der innerhalb der vorliegenden Arbeit konzipierten CBM beleuchtet. Dabei sollen alle Facetten der Validität, wie sie Moosbrugger und Kevala (2012) sowie Bühner (2006) beschreiben, analysiert werden. Zunächst folgen Angaben der Ergebnisse zur Augenscheinvalidität (Abschnitt 7.3.1), dann folgen die Ausführungen der Ergebnisse zur Konstruktvalidität (Abschnitt 7.3.2) sowie zur Kriteriumsvalidität (Abschnitt 7.3.3). Angaben zur Inhaltsvalidität sind der Beschreibung der CBM im Kapitel 4 zu entnehmen.

7.3.1 Augenscheinvalidität

Anhand der durch die untersuchten Lehrkräfte getroffenen Bewertungen der Aussagen des unter Abschnitt 6.3.7 beschriebenen Aussagenkatalogs können erste, wenn auch wenig differenzierte Anhaltspunkte hinsichtlich der Augenscheinvalidität der innerhalb vorliegender Arbeit erarbeiteter CBM gewonnen werden. Dabei sind insbesondere die Bewertungsergebnisse der Aussagen 1) und 3) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz relevant.

Die Ergebnisse der Einschätzung der Aussage 1) „Den Einsatz der von der Universität Rostock erarbeiteten CBM zur Dokumentation des Lernverlaufs meiner Schüler finde ich sinnvoll.“ sind in Abbildung 55 grafisch dargestellt. Von den insgesamt 37 Umfrageteilnehmern bewerteten 16 (43.2 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 16 der Lehrpersonen (43.2 %) wählten die Kategorie „trifft eher zu“. Hingegen hielten vier Personen (10.8 %) die entwickelten CBM für eher nicht sinnvoll und wählten die Antwort „trifft eher nicht zu“. Eine der Lehrkräfte (2.7 %) wählte die Option „trifft nicht zu“. Zusammenfassend lässt sich festhalten, dass die innerhalb der vorliegenden Arbeit konzipierten CBM von der Mehrheit der befragten Lehrkräfte (86.4 %) als sinnvoll erachtet wurden.

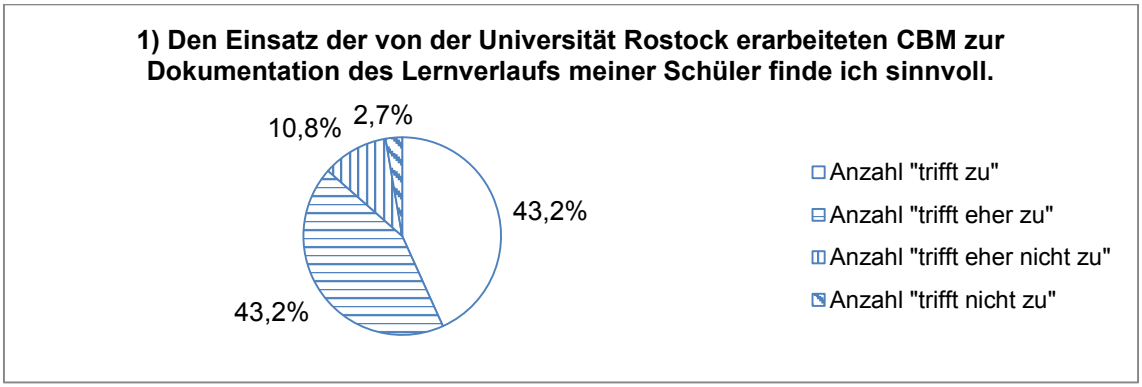


Abbildung 55: Ergebnisse der Beantwortung des Items 1) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Anhand der Aussage 3) „Eine Dokumentation des Lernverlaufs meiner Schüler finde ich sinnvoll.“ soll ermittelt werden, inwieweit die Lehrkräfte die Lernverlaufsdokumentation überhaupt als wichtig erachteten. Die Ergebnisse der Einschätzung sind in Abbildung 56 visualisiert. Von den insgesamt 37 Umfrageteilnehmern bewerteten 29 (78.4 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere sieben der Lehrpersonen (18.9 %) wählten die Kategorie „trifft eher zu“. Hingegen empfand eine Person (2.7 %), dass die Dokumentation des Lernverlaufs der Schüler nicht sinnvoll ist und wählte die Antwort „trifft eher nicht zu“. Keine der Lehrkräfte wählte die Option „trifft nicht zu“. Die Mehrheit der Befragten (97.3 %) sah in der Dokumentation des Lernverlaufs der Schüler eine sinnvolle Maßnahme. Werden diese Ergebnisse mit denen der Aussage 1) verglichen, ist festzustellen, dass es offensichtlich einen Teil von Personen (10.9 %) gab, welche die Lernverlaufsdokumentation zwar generell als richtig erachteten, jedoch die in vorliegender Arbeit konzipierten Verfahren als zu diesem Zweck nicht sinnvoll ansahen.

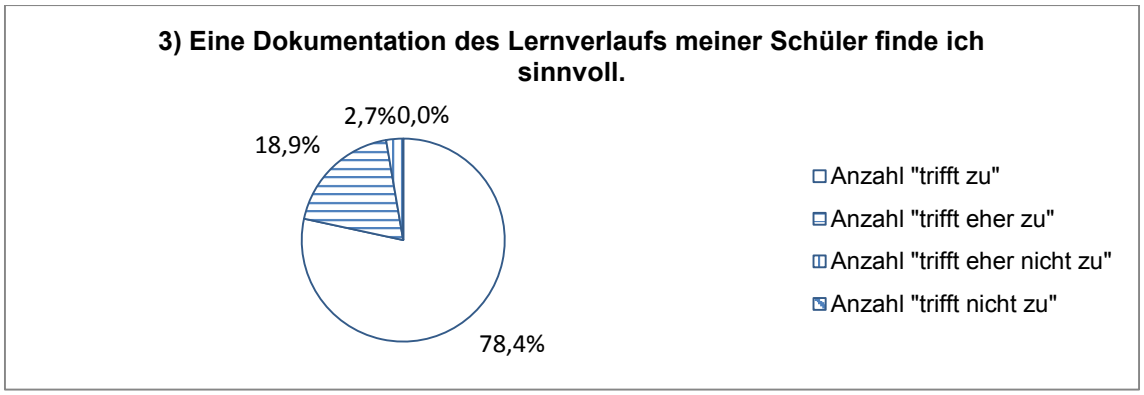


Abbildung 56: Ergebnisse der Beantwortung des Items 3) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

7.3.2 Konstruktvalidität

Für den Nachweis der Konstruktvalidität der konzipierten CBM werden in einem ersten Schritt die Interkorrelationen der Schülerergebnisse in den CBM der Region Rügen über die

Messzeitpunkte analysiert. Dabei werden mittlere Zusammenhänge zwischen den Ergebnissen der einzelnen Verfahren erwartet. Dies würde dafür sprechen, dass die CBM zum einen ein gemeinsames bzw. ähnliches Merkmal messen, in diesem Fall das Ausmaß verfügbarer arithmetischer Kompetenzen, und dass darüber hinaus jedoch auch verfahrensspezifische Aspekte, bspw. die Kompetenz zur Seriation von Zahlen, erfasst werden. In Tabelle 45 sind die Interkorrelation der CBM der Niveaustufe 1 zum ersten Messzeitpunkt in der fünften Schulwoche aufgeführt. Es zeigen sich geringe bis mittlere Korrelationen zwischen den einzelnen Verfahren. Die höchste Korrelation besteht mit r = .52 zwischen dem CBM „Zahlenreihe im Zwanzigerraum“ und dem CBM „Mengenvergleich im Zwanzigerraum“. Die geringste Korrelation weisen das CBM „Zahlenlesen im Zwanzigerraum“ und das CBM „Arithmetische Basiskompetenzen“ auf, mit r = .42 weicht diese nicht wesentlich von der Grenze zum mittleren Zusammenhang bei r ≥ .5 ab.

Tabelle 45: Interkorrelation (nach Pearson) der CBM der Niveaustufe 1 zum MZP1 in der fünften Schulwoche

	CBM ZR	CBM MV	CBM AB
CBM ZL	.44** (N = 436)	.46** (N = 436)	.42** (N = 437)
CBM ZR		.52** (N = 437)	.46** (N = 437)
CBM MV			.44** (N = 437)

Erläuterungen: CBM ZL – CBM „Zahlenlesen im Zwanzigerraum“, CBM ZR – CBM „Zahlenreihe im Zwanzigerraum“, CBM MV – CBM „Mengenvergleich im Zwanzigerraum“, CBM AB – CBM „Arithmetische Basiskompetenzen“, N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

Die Interkorrelationen der CBM der Niveaustufe 1 zum zweiten Messzeitpunkt in der achten Schulwoche sind in Tabelle 46 dargestellt. Auch hier zeigen sich geringe bis mittlere Korrelationen zwischen den einzelnen Verfahren. Die höchste Korrelation besteht mit r = .56 erneut zwischen dem CBM „Zahlenreihe im Zwanzigerraum“ und dem CBM „Mengenvergleich im Zwanzigerraum“. Das CBM „Arithmetische Basiskompetenzen“ weist durchweg Korrelationen kleiner r = .4 mit den anderen Verfahren auf.

Der geringste Zusammenhang ergibt sich dabei mit den CBM „Mengenvergleich im Zwanzigerraum“ (r = .36).

Tabelle 46: Interkorrelation (nach Pearson) der CBM der Niveaustufe 1 zum MZP 2 in der achten Schulwoche

	CBM ZR	CBM MV	CBM AB
CBM ZL	.50** (N = 432)	.52** (N = 429)	.38** (N = 433)
CBM ZR		.56** (N = 430)	.38** (N = 434)
CBM MV			.36** (N = 430)

Erläuterungen: CBM ZL – CBM „Zahlenlesen im Zwanzigerraum“, CBM ZR – CBM „Zahlenreihe im Zwanzigerraum“, CBM MV – CBM „Mengenvergleich im Zwanzigerraum“, CBM AB – CBM „Arithmetische Basiskompetenzen“, N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

Zum dritten Messzeitpunkt in der zwölften Schulwoche ergeben sich für die CBM der Niveaustufe 1 lediglich geringe Interkorrelationen zwischen den einzelnen Verfahren (Tabelle 47). Die höchste Korrelation besteht mit r = .47 erneut zwischen dem CBM „Zahlenlesen im Zwanzigerraum“ und dem CBM „Zahlenreihe im Zwanzigerraum“. Der geringste Zusammenhang herrscht zwischen den CBM „Arithmetische Basiskompetenzen“ und „Mengenvergleich im Zwanzigerraum“ (r = .31).

Tabelle 47: Interkorrelation (nach Pearson) der CBM der Niveaustufe 1 zum MZP 3 in der zwölften Schulwoche

	CBM ZR	CBM MV	CBM AB
CBM ZL	.47** (N = 428)	.423** (N = 427)	.36** (N = 427)
CBM ZR		.44** (N = 429)	.43** (N = 429)
CBM MV			.31** (N = 428)

Erläuterungen: CBM ZL – CBM „Zahlenlesen im Zwanzigerraum“, CBM ZR – CBM „Zahlenreihe im Zwanzigerraum“, CBM MV – CBM „Mengenvergleich im Zwanzigerraum“, CBM AB – CBM „Arithmetische Basiskompetenzen“, N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

Es zeigen sich geringe bis mittlere Interkorrelationen zwischen den einzelnen Verfahren der CBM der Niveaustufe 1 zum vierten Messzeitpunkt in der 16. Schulwoche. Die Kennwerte sind Tabelle 48 zu entnehmen. Die höchste Korrelation besteht dabei mit r = .56 zwischen dem CBM „Zahlenreihe im Zwanzigerraum“ und dem CBM

„Arithmetische Basiskompetenzen“. Der geringste Zusammenhang mit r = .32 lässt sich zwischen den CBM „Zahlenlesen im Zwanzigerraum“ und „Arithmetische Basiskompetenzen“ feststellen.

Tabelle 48: Interkorrelation (nach Pearson) der CBM der Niveaustufe 1 zum MZP 4 in der 16. Schulwoche

	CBM ZR	CBM MV	CBM AB
CBM ZL	.47** (N = 421)	.43** (N = 420)	.32** (N = 418)
CBM ZR		.50** (N = 425)	.56** (N = 423)
CBM MV			.43** (N = 423)

Erläuterungen: CBM ZL – CBM „Zahlenlesen im Zwanzigerraum“, CBM ZR – CBM „Zahlenreihe im Zwanzigerraum“, CBM MV – CBM „Mengenvergleich im Zwanzigerraum“, CBM AB – CBM „Arithmetische Basiskompetenzen“, N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

In Tabelle 49 sind die Interkorrelationen der CBM der Niveaustufe 2 zum fünften Messzeitpunkt in der 21. Schulwoche aufgeführt. Zwischen den Verfahren ergeben sich geringe bis eher hohe Korrelationen zwischen den einzelnen Verfahren. Die höchste Korrelation besteht mit r = .67 zwischen dem CBM „Addition im Zwanzigerraum“ und dem CBM „Subtraktion im Zwanzigerraum“. Das CBM „Arithmetische Basiskompetenzen“ weist durchweg die geringsten Korrelationen mit den anderen Verfahren zwischen r = .46 und r = .51 auf.

Tabelle 49: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP 5 in der 21. Schulwoche

	CBM ZZ	CBM AD	CBM SUB
CBM AB	.46** (N = 419)	.48** (N = 416)	.51** (N = 418)
CBM ZZ		.67** (N = 417)	.60** (N = 419)
CBM AD			.67** (N = 417)

Erläuterungen: CBM AB – CBM „Arithmetische Basiskompetenzen“, CBM ZZ – CBM „Zahlzerlegung im Zwanzigerraum“, CBM AD – CBM „Addition im Zwanzigerraum“, CBM SUB – CBM „Subtraktion im Zwanzigerraum“; N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

Bei den Interkorrelationen der CBM der Niveaustufe 2 zum sechsten Messzeitpunkt in der 26. Schulwoche ergeben sich geringe bis hohe Koeffizienten zwischen den einzelnen Verfahren (Tabelle 50). Die

höchste Korrelation besteht mit r = .75 zwischen dem CBM „Zahlzerlegung im Zwanzigerraum“ und dem CBM „Subtraktion im Zwanzigerraum“. Dieser Zusammenhang ist laut Bühl (2006) als hoch einzuordnen. Erneut weist das CBM „Arithmetische Basiskompetenzen“ mit Koeffizienten zwischen r = .43 und r = .46 die geringsten Zusammenhänge mit den anderen CBM auf. Nach Bühl (2006) handelt es sich dabei um geringe Korrelationen.

Tabelle 50: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP 6 in der 26. Schulwoche

	CBM ZZ	CBM AD	CBM SUB
CBM AB	.43** (N = 429)	.46** (N = 428)	.46** (N = 429)
CBM ZZ		.70** (N = 430)	.72** (N = 431)
CBM AD			.75** (N = 432)

Erläuterungen: CBM AB – CBM „Arithmetische Basiskompetenzen“, CBM ZZ – CBM „Zahlzerlegung im Zwanzigerraum“, CBM AD – CBM „Addition im Zwanzigerraum“, CBM SUB – CBM „Subtraktion im Zwanzigerraum“; N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

In Tabelle 51 sind die Interkorrelationen der CBM der Niveaustufe 2 zum siebten Messzeitpunkt in der 30. Schulwoche aufgeführt. Es zeigen sich, abhängig von den untersuchten Verfahrenskonstellationen geringe bis mittlere Korrelationen. Die höchste Korrelation besteht mit r = .63 zwischen dem CBM „Zahlzerlegung im Zwanzigerraum“ und dem CBM „Subtraktion im Zwanzigerraum“. Das CBM „Arithmetische Basiskompetenzen“ weist mit Korrelationen zwischen r = .42 und r = .48 durchweg geringe Zusammenhänge mit den weiteren CBM auf.

Tabelle 51: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP7 in der 30. Schulwoche

	CBM ZZ	CBM AD	CBM SUB
CBM AB	.48** (N = 426)	.47** (N = 426)	.42** (N = 426)
CBM ZZ		.58** (N = 431)	.63** (N = 431)
CBM AD			.61** (N = 431)

Erläuterungen: CBM AB – CBM „Arithmetische Basiskompetenzen“, CBM ZZ – CBM „Zahlzerlegung im Zwanzigerraum“, CBM AD – CBM „Addition im Zwanzigerraum“, CBM SUB – CBM „Subtraktion im Zwanzigerraum“; N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

Die Interkorrelationen der CBM der Niveaustufe 2 zum achten Messzeitpunkt in der 34. Schulwoche sind in Tabelle 52 aufgeführt. Zwischen den Verfahren ergeben sich erneut geringe bis hohe Korrelationen zwischen den einzelnen Verfahren. Der größte Zusammenhang besteht mit r = .72 zwischen dem CBM „Addition im Zwanzigerraum“ und dem CBM „Subtraktion im Zwanzigerraum“. Laut Bühl (2006) handelt es sich dabei um eine hohe Korrelation. Das CBM „Arithmetische Basiskompetenzen“ weist durchweg die kleinsten Zusammenhänge mit den anderen Verfahren zwischen r = .41 und r = .47 auf. Bühl (2006) folgend, handelt es sich dabei um geringe Korrelationskoeffizienten.

Tabelle 52: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP 8 in der 34. Schulwoche

	CBM ZZ	CBM AD	CBM SUB
CBM AB	.41** (N = 371)	.45** (N = 370)	.47** (N = 370)
CBM ZZ		.70** (N = 372)	.66** (N = 372)
CBM AD			.72** (N = 371)

Erläuterungen: CBM AB – CBM „Arithmetische Basiskompetenzen“, CBM ZZ – CBM „Zahlzerlegung im Zwanzigerraum“, CBM AD – CBM „Addition im Zwanzigerraum“, CBM SUB – CBM „Subtraktion im Zwanzigerraum“; N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

Für die CBM der Niveaustufe 2 zum neunten Messzeitpunkt in der 38. Schulwoche ergeben sich geringe bis hohe Interkorrelationen zwischen den einzelnen Verfahren (Tabelle 53). Zwischen dem CBM „Addition im Zwanzigerraum“ und dem CBM „Zahlzerlegung im

Zwanzigerraum“ besteht der größte Zusammenhang, welcher mit r = .74 als hoch eingeschätzt werden kann. Alle weiteren Korrelationskoeffizienten deuten mit Koeffizienten zwischen r = .55 und r = 70 auf mittlere Zusammenhänge zwischen den CBM hin. Der geringste Zusammenhang herrscht zwischen den CBM „Arithmetische Basiskompetenzen“ und „Zahlzerlegung im Zwanzigerraum“ (r = .55).

Tabelle 53: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP 9 in der 38. Schulwoche

	CBM ZZ	CBM AD	CBM SUB
CBM AB	.55** (N = 427)	.56** (N = 428)	.55** (N = 424)
CBM ZZ		.74** (N = 429)	.70** (N = 425)
CBM AD			.70** (N = 426)

Erläuterungen: CBM AB – CBM „Arithmetische Basiskompetenzen“, CBM ZZ – CBM „Zahlzerlegung im Zwanzigerraum“, CBM AD – CBM „Addition im Zwanzigerraum“, CBM SUB – CBM „Subtraktion im Zwanzigerraum“; N – Stichprobenumfang; ** - Korrelationen ist auf dem Niveau von 0.01 (2-seitig) signifikant.

7.3.2.1 Diskriminante und konvergente Validität

Wie in Abschnitt 6.4.1 beschrieben wird zur Überprüfung der Konstruktvalidität auf die von Campbell und Fiske (1959) beschriebene Multitrait-Multimethod-Methode zurückgegriffen. Dabei werden jeweils alle Ergebnisse der in der vorliegenden Arbeit konzipierten CBM der Niveaustufe 1 in der fünften Schulwoche mit den Ergebnissen der Schüler im CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), dem Münsteraner Screening (MÜSC, Mannhaupt, 2005) und dem Teil 1 des Diagnoseprogramms Kalkulie (Fritz et al., 2007) korreliert. Das CBM „Silbenlesen“ als auch das Münsteraner Screening fragen jeweils Vorläuferfähigkeiten zum Schriftspracherwerb ab. Die Ergebnisse der CBM der Niveaustufe 2, erhoben in der 21. Schulwoche, werden mit den Ergebnissen der Schüler im CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b), der Würzburger Leise Leseprobe (WLLP, Küspert & Schneider, 1998)

und dem Deutschen Mathematiktest für erste Klassen (DEMAT 1+; Krajewski et al., 2002) korreliert. In den Monotrait-Heteromethod-Blöcken wird jeweils lediglich eine mittlere bis hohe Korrelation erwartet, da CBM aufgrund ihrer Durchführungsspezifika neben dem Vorhandensein mathematischer Einsichten auch deren zügige Anwendbarkeit innerhalb der zeitlich begrenzten Bearbeitungszeit abfragen. So geben auch Salvia, Ysseldyke und Bolt (2007) sowie Foegen et al. (2007) an, dass die Reliabilitätskoeffizienten von CBM und Außenkriterien in Form von standardisierten Leistungstests eher mäßig, im Bereich von r zwischen .50 bis .70, ausfallen.

In Tabelle 54 ist die Multitrait-Multimethod-Matrix des CBM „Zahlenlesen im Zwanzigerraum“ zum ersten Messzeitpunkt in der fünften Schulwoche abgebildet (N = 436). Der Korrelationskoeffizient des Monotrait-Heteromethod-Blockes zwischen dem CBM „Zahlenlesen im Zwanzigerraum“ und dem Diagnoseprogramm Kalkulie Teil 1 (Fritz et al., 2007) fällt betragsmäßig mit einer mittleren Korrelation von r = .55 am größten aus, was für die konvergente Validität des CBM spricht. Die Korrelationen mit dem CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a) und dem Münsteraner Screening (Mannhaupt, 2006) fallen geringer aus, was als Anzeichen für vorliegende diskriminante Validität gedeutet werden kann.

Tabelle 54: Multitrait-Multimethod-Matrix des CBM „Zahlenlesen im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), des Diagnoseverfahrens Kalkulie Teil 1 (Fritz et al., 2007) sowie des Münsteraner Screenings (MÜSC; Mannhaupt, 2005) zum ersten Messzeitpunkt in der 5. Schulwoche (N = 436)

	CBM ZL	CBM SL	Kalkulie Teil 1 RW	MÜSC RP
CBM ZL	1			
CBM SL	.26**	1		
Kalkulie Teil 1 RW	.55**	.21**	1	
MÜSC RP	-.35**	-.14**	-.48**	1

Erläuterungen: CBM ZL – CBM „Zahlenlesen im Zwanzigerraum“, CBM SL – CBM „Silbenlesen“, RW – Rohwerte, MÜSC – Münsteraner Screening (Mannhaupt, 2005), RP – Risikopunkte, ** – die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

Tabelle 55 ist die Multitrait-Multimethod-Matrix des CBM „Zahlenreihe im Zwanzigerraum“ zum ersten Messzeitpunkt in der fünften Schulwoche zu entnehmen. Grundlage dafür bilden die Daten von 435 Schülern. Der Korrelationskoeffizient des Monotrait-Heteromethod-Blockes zwischen dem CBM „Zahlenreihe im Zwanzigerraum“ und dem Diagnoseprogramm Kalkulie Teil 1 (Fritz et al., 2007) fällt mit einer knapp mittleren Korrelation von r = .47 am größten aus. Die Korrelationen mit dem CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a) und dem Münsteraner Screening (Mannhaupt, 2006) fallen auch in diesem Fall geringer aus.

Tabelle 55: Multitrait-Multimethod-Matrix des CBM „Zahlenreihe im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), des Diagnoseverfahrens Kalkulie Teil 1 (Fritz et al., 2007) sowie des Münsteraner Screenings (MÜSC; Mannhaupt, 2005) zum ersten Messzeitpunkt in der 5. Schulwoche (N = 435)

	CBM ZR	CBM SL	Kalkulie Teil 1 RW	MÜSC RP
CBM ZR	1			
CBM SL	.19**	1		
Kalkulie Teil 1 RW	.47**	.21**	1	
MÜSC RP	-.30**	-.14**	-.48**	1

Erläuterungen: CBM ZR – CBM „Zahlenreihe im Zwanzigerraum“, CBM SL – CBM „Silbenlesen“, RW – Rohwerte, MÜSC – Münsteraner Screening (Mannhaupt, 2005), RP - Risikopunkte, ** – die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

In Tabelle 56 ist die Multitrait-Multimethod-Matrix des CBM „Mengenvergleich im Zwanzigerraum“ zum ersten Messzeitpunkt in der fünften Schulwoche abgebildet (N = 435). Der Korrelationskoeffizient des Monotrait-Heteromethod-Blockes zwischen dem CBM „Mengenvergleich im Zwanzigerraum“ und dem Diagnoseprogramm Kalkulie Teil 1 (Fritz et al., 2007) fällt betragsmäßig mit einer mittleren Korrelation von r = .51 am größten aus. Die Korrelationen mit dem CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a) und dem Münsteraner Screening (Mannhaupt, 2006) fallen geringer aus.

Tabelle 56: Multitrait-Multimethod-Matrix des CBM „Mengenvergleich im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), des Diagnoseverfahrens Kalkulie Teil 1 (Fritz et al., 2007) sowie des Münsteraner Screenings (MÜSC; Mannhaupt, 2005) zum ersten Messzeitpunkt in der 5. Schulwoche (N = 435)

	CBM MV	CBM SL	Kalkulie Teil 1 RW	MÜSC RP
CBM MV	1			
CBM SL	.17**	1		
Kalkulie Teil 1 RW	.51**	.21**	1	
MÜSC RP	-.35**	-.14**	-.48**	1

Erläuterungen: CBM MV – CBM „Mengenvergleich im Zwanzigerraum“, CBM SL – CBM „Silbenlesen“, RW – Rohwerte, MÜSC – Münsteraner Screening (Mannhaupt, 2005), RP - Risikopunkte, ** – die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

In Tabelle 57 ist die Multitrait-Multimethod-Matrix des CBM „Arithmetische Basiskompetenzen“ zum ersten Messzeitpunkt in der fünften Schulwoche dargestellt. Als Berechnungsgrundlage gehen die Schülerergebnisse von 435 Kindern ein. Der Korrelationskoeffizient des Monotrait-Heteromethod-Blockes zwischen dem CBM „Arithmetische Basiskompetenzen“ und dem Diagnoseprogramm Kalkulie Teil1 (Fritz et al., 2007) fällt mit einer geringen Korrelation von r = .41 am größten aus. Die Korrelationen mit dem CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a) und dem Münsteraner Screening (Mannhaupt, 2006) fallen auch in diesem Fall geringer aus.

Tabelle 57: Multitrait-Multimethod-Matrix des CBM „Arithmetische Basiskompetenzen“, des CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), des Diagnoseverfahrens Kalkulie Teil 1 (Fritz et al., 2007) sowie des Münsteraner Screenings (MÜSC; Mannhaupt, 2005) zum ersten Messzeitpunkt in der 5. Schulwoche (N = 435)

	CBM AB	CBM SL	Kalkulie Teil 1 RW	MÜSC RP
CBM AB	1			
CBM SL	.11*	1		
Kalkulie Teil 1 RW	.41**	.21**	1	
MÜSC RP	-.22**	-.14**	-.48**	1

Erläuterungen: CBM AB – CBM „Arithmetische Basiskompetenzen“, CBM SL – CBM „Silbenlesen“, RW – Rohwerte, MÜSC – Münsteraner Screening (Mannhaupt, 2005), RP – Risikopunkte, ** – die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

Tabelle 58 gibt die Multitrait-Multimethod-Matrix des CBM „Arithmetische Basiskompetenzen“ auf Grundlage der Daten von 364 Schülern zum neunten Messzeitpunkt in der 38. Schulwoche wider. Der Korrelationskoeffizient des Monotrait-Heteromethod-Blockes zwischen dem CBM „Arithmetische Basiskompetenzen“ und dem Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) zeigt mit r = .45 einen knapp mittleren Zusammenhang an und fällt verglichen mit den Korrelationen zu den Lesetests am höchsten aus. Die Korrelationen mit dem CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b) und der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) fallen mit jeweils r = .30 geringer aus.

Tabelle 58: Multitrait-Multimethod-Matrix des CBM „Arithmetische Basiskompetenzen“, des CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b), des Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) sowie der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) zum neunten Messzeitpunkt in der 38. Schulwoche (N = 364)

	CBM AB	CBM SL	DEMAT 1+ RW	WLLP RW
CBM AB	1			
CBM SL	.30**	1		
DEMAT 1+ RW	.45**	.40**	1	
WLLP RW	.30**	.71**	.44**	1

Erläuterungen: CBM AB – CBM „Arithmetische Basiskompetenzen“, CBM SL – CBM „Silbenlesen“, DEMAT 1+ - Deutscher Mathematiktest für erste Klassen (Krajewski et al., 2002), RW – Rohwerte, ** – die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

In Tabelle 59 ist die Multitrait-Multimethod-Matrix des CBM „Zahlzerlegung im Zwanzigerraum“ zum neunten Messzeitpunkt in der 38. Schulwoche abgebildet (N = 365). Der Korrelationskoeffizient des Monotrait-Heteromethod-Blockes zwischen dem CBM „Zahlzerlegung im Zwanzigerraum“ und dem Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) fällt mit einer mittleren Korrelation von r = .64 verglichen mit den Korrelationen zu den Lesetests am höchsten aus. Dabei fällt die Korrelation mit dem CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b) mit r = .44 geringer aus als der Zusammenhang mit der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) von r = .45.

Tabelle 59: Multitrait-Multimethod-Matrix des CBM „Zahlzerlegung im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b), des Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) sowie der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) zum neunten Messzeitpunkt in der 38. Schulwoche (N = 365)

	CBM ZZ	CBM SL	DEMAT 1+ RW	WLLP RW
CBM ZZ	1			
CBM SL	.44**	1		
DEMAT 1+ RW	.64**	.40**	1	
WLLP RW	.45**	.71**	.44**	1

Erläuterungen: CBM ZZ – CBM „Zahlzerlegung im Zwanzigerraum“, CBM SL – CBM „Silbenlesen“, DEMAT 1+ - Deutscher Mathematiktest für erste Klassen (Krajewski et al., 2002), RW – Rohwerte, ** – die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

Zur Einschätzung der Konstruktvalidität des CBM „Addition im Zwanzigerraum“ zum neunten Messzeitpunkt in der 38. Schulwoche wurden die Daten von 366 Schülern analysiert. In Tabelle 60 ist die entsprechende Multitrait-Multimethod-Matrix abgebildet. Der Korrelationskoeffizient des Monotrait-Heteromethod-Blockes zwischen dem CBM „Addition im Zwanzigerraum“ und dem Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) zeigt mit r = .64 einen mittleren Zusammenhang an und fällt verglichen mit den Korrelationen zu den Lesetests am höchsten aus. Die Korrelationen mit dem CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b) und der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) fallen mit r = .43. geringer aus.

Tabelle 60: Multitrait-Multimethod-Matrix des CBM „Addition im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b), des Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) sowie der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) zum neunten Messzeitpunkt in der 38. Schulwoche (N = 366)

	CBM AD	CBM SL	DEMAT 1+ RW	WLLP RW
CBM AD	1			
CBM SL	.43**	1		
DEMAT 1+ RW	.64**	.40**	1	
WLLP RW	.43**	.71**	.44**	1

Erläuterungen: CBM AD – CBM „Addition im Zwanzigerraum“, CBM SL – CBM „Silbenlesen“, DEMAT 1+ - Deutscher Mathematiktest für erste Klassen (Krajewski et al., 2002), RW – Rohwerte, ** – die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

In Tabelle 61 ist die Multitrait-Multimethod-Matrix des CBM „Subtraktion im Zwanzigerraum“ zum neunten Messzeitpunkt in der 38. Schulwoche abgebildet (N = 362). Der Korrelationskoeffizient des Monotrait-Heteromethod-Blockes zwischen dem CBM „Subtraktion im Zwanzigerraum“ und dem Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) fällt mit einer mittleren Korrelation von r = .56 verglichen mit den Korrelationen zu den Lesetests am höchsten aus. Dabei fällt die Korrelation mit dem CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b) mit r = .35 geringer aus als der Zusammenhang mit der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) von r = .40.

Tabelle 61: Multitrait-Multimethod-Matrix des CBM „Subtraktion im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b), des Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) sowie der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) zum neunten Messzeitpunkt in der 38. Schulwoche (N = 362)

	CBM SUB	CBM SL	DEMAT 1+ RW	WLLP RW
CBM SUB	1			
CBM SL	.35**	1		
DEMAT 1+ RW	.56**	.40**	1	
WLLP RW	.40**	.71**	.44**	1

Erläuterungen: CBM SUB – CBM „Subtraktion im Zwanzigerraum“, CBM SL – CBM „Silbenlesen“, DEMAT 1+ - Deutscher Mathematiktest für erste Klassen (Krajewski et al., 2002), RW – Rohwerte, ** – die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

Ergänzend zu den zuvor angewendeten Analysen zur Bestimmung der konvergenten und diskriminanten Validität anhand der Multitrait-Multimethod-Methode von innerhalb der vorliegenden Studie erarbeiteten CBM werden nachfolgend zusätzliche Analysen zur Bestimmung der konvergenten Validität durchgeführt. Dabei werden die Korrelationen der CBM mit dem von Wagner und Hartke (2006) entwickelten Inventar „Rechenfische“ ermittelt. Da die „Rechenfische“ (Wagner & Hartke, 2006) wiederholt in der zweiten Hälfte des ersten Schuljahres durchgeführt wurden, sind korrelative Analysen zu verschiedenen Messzeitpunkten möglich.

In Tabelle 62 sind die Korrelationen des Inventars „Rechenfische“ mit den CBM der Niveaustufe 1 und 2 zum vierten bzw. fünften Messzeitpunkt aufgeführt. Aufgrund des paarweisen Fallausschlusses fehlender Werte bilden verschiedene Stichprobengrößen die Analysegrundlage. Die Korrelationskoeffizienten fallen in unterschiedlicher Höhe aus, so zeigen sich geringe bis mittlere Zusammenhänge. Zwischen dem CBM „Zahlenlesen im Zwanzigerraum“ und den „Rechenfischen“ besteht mit r = .44 der kleinste Zusammenhang. Gemeinsam mit dem CBM „Mengenvergleich im Zwanzigerraum“ ergeben sich bei den

Verfahren nur geringe Korrelationen. Zwischen den anderen CBM und den „Rechenfischen“ ergeben sich durchweg mittlere Korrelationen. Das CBM „Addition im Zwanzigerraum“ weist dabei mit r = .695 den höchsten Koeffizienten auf.

Tabelle 62: Korrelationsmatrix nach Pearson der CBM der Niveaustufe 1 und 2 mit dem Inventar Rechenfische (Wagner & Hartke, 2006) zum vierten bzw. fünften Messzeitpunkt in der 21. Schulwoche

	Durchgeführt zum MZP	N	Korrelation mit RW der Rechenfische zum MZP 5
CBM „Zahlenlesen im Zwanzigerraum“	4	399	.44
CBM „Zahlenreihe im Zwanzigerraum“	4	400	.51
CBM „Mengenvergleich im Zwanzigerraum“	4	399	.50
CBM „Arithmetische Basiskompetenzen“	5	408	.51
CBM „Zahlzerlegung im Zwanzigerraum“	5	403	.64
CBM „Addition im Zwanzigerraum“	5	400	.70
CBM „Subtraktion im Zwanzigerraum“	5	403	.66

Erläuterungen: N – Stichprobenumfang; MZP – Messzeitpunkt; RW – Rohwert
Anmerkung: Alle Korrelationen sind auf dem Niveau von 0.01 (2-seitig) signifikant.

Die ermittelten Korrelationen des Inventars „Rechenfische“ (Wagner & Hartke, 2006) mit den CBM der Niveaustufe 2 zum achten Messzeitpunkt in der 34. Schulwoche sind in Tabelle 63 zusammengefasst. Auch in diesem Fall variiert die Stichprobengröße für die Analysen zwischen N = 245 und N = 246. Abhängig vom CBM können geringe bis mittlere Korrelationen ermittelt werden. Zwischen dem CBM „Arithmetische Basiskompetenzen“ und den „Rechenfischen“ besteht mit r = .43 der kleinste Zusammenhang, es handelt sich um eine geringe Korrelationen. Zwischen den CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ und den „Rechenfischen“ ergeben sich durchweg mittlere Korrelationen. Das CBM „Addition im Zwanzigerraum“ weist dabei mit r = .68 den höchsten Koeffizienten auf.

Tabelle 63: Korrelationsmatrix nach Pearson der CBM der Niveaustufe 2 mit dem Inventar Rechenfische (Wagner & Hartke, 2006) zum achten Messzeitpunkt in der 34. Schulwoche

	Durchgeführt zum MZP	Gültige N	Korrelation mit RW der Rechenfische zum MZP 8
CBM „Arithmetische Basiskompetenzen“	8	245	.43
CBM „Zahlzerlegung im Zwanzigerraum“	8	246	.58
CBM „Addition im Zwanzigerraum“	8	245	.68
CBM „Subtraktion im Zwanzigerraum“	8	245	.55

Erläuterungen: N – Stichprobenumfang; MZP – Messzeitpunkt; RW – Rohwert
Anmerkung: Alle Korrelationen sind auf dem Niveau von 0.01 (2-seitig) signifikant.

7.3.3 Kriteriumsvaliditat

In den nachfolgenden Abschnitten werden die Ergebnisse der Analysen hinsichtlich der Übereinstimmungs- (Abschnitt 7.3.3.1), der prognostischen (Abschnitt 7.3.3.2) sowie der inkrementellen Validität (Abschnitt 7.3.3.3) als Teilaspekte der Kriteriumsvalidität der einzelnen CBM dargestellt.

7.3.3.1 Übereinstimmungsvalidität

Als Grundlage zur Bestimmung der Übereinstimmungsvalidität der innerhalb der vorliegenden Arbeit konzipierten CBM dienen die Analysen hinsichtlich der konvergenten Validität, welche im Abschnitt 7.3.2 dargestellt wurden. Die MTMM-Analysen haben dabei mittlere bis knapp mittlere Zusammenhänge der CBM und den gewählten Außenkriterien (Kalkulie Teil 1, Fritz et al., 2007; DEMAT 1+, Krajewski et al., 2002) nachgewiesen. Dabei ist festzustellen, dass die Korrelationen der CBM-Ergebnisse der Niveaustufe 1 mit den Ergebnissen des Diagnoseprogramms Kalkulie (Fritz et al., 2007) insgesamt gering ausfallen. Das CBM „Arithmetische Basiskompetenzen“ weist hier mit $r = .41$ den kleinsten Korrelationskoeffizienten im Rahmen der Untersuchungen zur

konvergenten Validität auf. Dieser Wert ist nach Bühl (2006) als gering einzustufen.

In zusätzlichen Untersuchungen wurde die konvergente Validität der hier entwickelten CBM anhand von Analysen des Zusammenhangs zum Inventar „Rechenfische“ (Wagner & Hartke, 2006) durchgeführt. Zum fünften Messzeitpunkt in der 21. Schulwoche wurden dabei Korrelationen im Bereich von r = .44 und r = .70 zwischen den CBM der Niveaustufen 1 und 2 und dem Inventar „Rechenfische“ (Wagner & Hartke, 2006) ermittelt (vgl. Tabelle 62), wobei das CBM „Arithmetische Basiskompetenzen“ den geringsten und das CBM „Addition im Zwanzigerraum“ hingegen den höchsten Koeffizienten erzielte. Beim erneuten Einsatz des Inventars „Rechenfische“ (Wagner & Hartke, 2006) zum achten Messzeitpunkt in der 34. Schulwoche ergaben sich Korrelationen im Bereich von r = .43 und r= .68 zwischen den CBM der Niveaustufe 2 und dem Inventar „Rechenfische“ (Wagner & Hartke, 2006) (vgl. Tabelle 63). Auch zu diesem Zeitpunkt erreicht das CBM „Arithmetische Basiskompetenzen“ den geringsten Korrelationskoeffizienten, das CBM „Addition im Zwanzigerraum“ hingegen den höchsten.

Als eine weitere Möglichkeit zur Bestimmung der Übereinstimmungsvalidität schlagen Bortz und Döring (2009) die sog. „Technik der bekannten Gruppen“ (S. 201) vor. Dabei wird überprüft, ob zuvor festgelegte zu erwartende Unterschiede in bestimmten Probandengruppen sich auch tatsächlich im zu messenden Konstrukt widerspiegeln. Bezogen auf die CBM der vorliegenden Arbeit wird, ausgehend von den theoretischen Ausführungen in Abschnitt 2.10, angenommen, dass Schüler mit lediglich gering ausgeprägtem mathematischen Vorwissen zu Beginn des ersten Schuljahres einen gewissen Risikostatus inne haben, welcher sich bei späteren Gruppenvergleichen ungünstig auf die Mathematikwerte auswirken, sich also in geringeren Punktwerten in den einzelnen CBM im

Vergleich zu den Peers ausdrücken. Operationalisiert wird dieser Risikostatus eines Schülers hier durch einen Prozentrang kleiner bzw. gleich 15 im Diagnoseprogramm Kalkulie (Fritz et al., 2007). In Abschnitt 7.4 wird der Einfluss dieses Risikostatus auf das Ausgangsniveau sowie den Leistungsanstieg in den CBM-Daten der untersuchten Kinder mehrebenenanalytisch im Rahmen der Erforschung der Änderungssensibilität der CBM untersucht. Die Ergebnisse dieser Analysen liefern ebenso Indizien hinsichtlich der Übereinstimmungsvalidität nach der von Bortz und Döring (2009) geschilderten Technik der bekannten Gruppen.

7.3.3.2 Prognostische Validität

Im Rahmen der Überprüfung der prognostischen Validität der innerhalb der vorliegenden Arbeit entwickelten CBM werden klassifikatorische Analysen angewendet, wie sie u. a. bei Hasselhorn et al. (2005), Marx (1992) sowie Marx und Lenhard (2010) beschrieben sind (vgl. Abschnitt 6.4.2).

CBM „Zahlenlesen im Zwanzigerraum“

In Tabelle 64 sind die Ergebnisse des CBM „Zahlenlesen im Zwanzigerraum“ zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins aufgeführt. Getrennt nach Schulwochen sind dort zunächst die aufgrund der festgesetzten Cutoff-Werte resultierenden Klassifikationen sowie anschließend die wesentlichen Kennwerte zur Beurteilung der Vorhersagegüte des CBM „Zahlenlesen im Zwanzigerraum“ notiert.

Die Sensitivität ergibt sich als das Verhältnis von richtig positiv klassifizierten Kindern bezogen auf die Gesamtheit der Kinder mit niedrigen Rechenleistungen (PR ≤ 15) zum Ende der Klasse eins. Über die Messzeitpunkte hinweg variieren die Werte der Sensitivität

in einem Intervall zwischen 0.60 und 0.69, d. h. dass das CBM „Zahlenlesen im Zwanzigerraum“ über die Schulwochen einen Prozentsatz von 60 % bis 69 % der Kinder mit Minderleistungen im Rechnen zum Ende der Klasse eins korrekt identifizieren kann. Die β-Fehlerquote gibt als Komplement zur Sensitivität den prozentualen Anteil falsch negativ klassifizierter Fälle an. Für das CBM „Zahlenlesen im Zwanzigerraum“ fällt sie mit Werten zwischen ca. 31 und 40 % über die Messzeitpunkte relativ hoch aus. Insgesamt weisen die Ergebnisse darauf hin, dass über ein Drittel der Kinder auf Grundlage der Daten des CBM „Zahlenlesen im Zwanzigerraum“ fälschlicherweise als unauffällig eingeschätzt wird.

Die Spezifität gibt Aufschluss darüber, wie hoch das Verhältnis von richtig negativ klassifizierten Kindern bezogen auf die Gesamtheit der Kinder ohne schulische Minderleistungen im Rechen (PR > 15) zum Ende der Klasse eins ist. Für das CBM „Zahlenlesen im Zwanzigerraum“ ergeben sich über die verschiedenen Messzeitpunkte in der ersten Hälfte des ersten Schuljahres Spezifitätswerte im Bereich von 0.66 und 0.74. Das bedeutet, der Prozentsatz der Kinder ohne festgestellte Minderleistungen im Rechnen zum Ende der Klasse eins, die durch den jeweiligen Prädiktor auch als unauffällig eingestuft wurden, bezogen auf alle Kinder mit geringen Rechenleistungen zum Ende der Klassenstufe 1, beträgt zwischen 66 % und 74 %. Die α-Fehlerquote bildet als Komplement zur Spezifität den prozentualen Anteil falsch positiver Klassifikationen ab. Die α-Fehlerquote ist beim CBM „Zahlenlesen im Zwanzigerraum“ mit Werten zwischen etwa 26 % und 34 % über die Messzeitpunkte recht hoch. Insgesamt weisen die Ergebnisse darauf hin, dass ein Viertel bis ein Drittel der Kinder auf Grundlage der Daten des CBM „Zahlenlesen im Zwanzigerraum“ fälschlicherweise als auffällig eingeschätzt wird.

Der positive prädiktive Wert, auch Prädiktortrefferquote genannt, gibt an, mit welcher Wahrscheinlichkeit ein anhand des Prädiktors als auffällig klassifiziertes Kind später auch tatsächlich innerhalb der Kriteriumsmessung auffällig wird. Die Prädiktortrefferquote liegt für das CBM „Zahlenlesen im Zwanzigerraum“ über die Messzeitpunkte in einem Intervall von 0.24 bis 0.27. Dies weist darauf hin, dass 24 % bis 27 % der Kinder, welche anhand des CBM als auffällig eingestuft wurden, zum Ende der Klasse eins tatsächlich Minderleistungen im Rechnen aufweisen. Der negative prädiktive Wert gibt an, mit welcher Wahrscheinlichkeit ein anhand des Prädiktors als unauffällig klassifiziertes Kind später auch tatsächlich innerhalb der Kriteriumsmessung unauffällig bleibt. Mit Werten zwischen 0.92 und 0.93 über die Messzeitpunkte hinweg gibt er an, dass der Anteil der im CBM „Zahlenlesen im Zwanzigerraum“ korrekterweise als unauffällig klassifizierten Kinder über die Schulwochen über 90 % liegt.

Als einer der bedeutsamsten Kennwerte zur Vorhersagegüte von Testverfahren wurde abschließend der RATZ-Index ermittelt, welcher den relativen Anstieg gegenüber einer Zufallstrefferquote unter Berücksichtigung der maximalen Trefferquote angibt. Die RATZ-Indices zu den verschiedenen Messzeitpunkten liegen in einem Bereich zwischen 41 und 50 %, d. h. dass die Vorhersagen auf Basis des CBM „Zahlenlesen im Zwanzigerraum“ sich leicht gegenüber einer zufälligen Zuordnung verbessern und das Verfahren eine gute Klassifikationsleistung aufweist.

Tabelle 64: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Zahlenlesen im Zwanzigerraum“ (PR ≤ 25)

		Klassifikation				Kennwerte						
Schulwoche	Cutoff-Wert in RW	richtig positiv	falsch positiv	falsch negativ	richtig negativ	Sensitivität	Spezifität	positiver prädiktiver Wert	negativer prädiktiver Wert	Anteil falsch Positiver (α-Fehlerquote) in %	Anteil falsch Negativer (β-Fehlerquote) in %	RATZ-Index
5	19	31	98	21	239	0.60	0.71	0.24	0.92	29.1	40.4	0.40
8	26	36	115	16	221	0.69	0.66	0.24	0.93	34.2	30.8	0.50
12	29	30	92	20	240	0.60	0.72	0.25	0.92	27.7	40.0	0.41
16	34	31	84	19	239	0.62	0.74	0.27	0.93	26.0	38.0	0.45

Erläuterungen: RW – Rohwert

CBM „Zahlenreihe im Zwanzigerraum“

Die Ergebnisse zur Analyse der prognostischen Validität des CBM „Zahlenreihe im Zwanzigerraum“ hinsichtlich auftretender Minderleistungen im Rechnen zum Ende der Klasse eins sind in Tabelle 65 dargestellt. Neben den benutzten Cutoff-Werten und den damit einhergehenden Klassifikationen werden wesentliche Kennwerte zur Beurteilung der Vorhersagegüte des CBM „Zahlenreihe im Zwanzigerraum“ getrennt nach Schulwochen aufgeführt.

Über die Messzeitpunkte hinweg variieren die Werte der Sensitivität in einem Intervall zwischen 0.60 und 0.79, d. h. dass das CBM „Zahlenreihe im Zwanzigerraum“ über die Schulwochen einen Prozentsatz von 60 % bis 79 % der Kinder mit Minderleistungen im Rechnen zum Ende der Klasse eins korrekt identifizieren kann. Die β-Fehlerquote fällt für das CBM „Zahlenreihe im Zwanzigerraum“ mit Werten zwischen 30.8 % und 40 % über die Messzeitpunkte relativ

hoch aus. Insgesamt weisen die Ergebnisse darauf hin, dass über ein Drittel der Kinder auf Grundlage der Daten des CBM „Zahlenreihe im Zwanzigerraum“ fälschlicherweise als unauffällig eingeschätzt wird.

Für das CBM „Zahlenreihe im Zwanzigerraum“ ergeben sich über die verschiedenen Messzeitpunkte in der ersten Hälfte des ersten Schuljahres Spezifitätswerte im Bereich von 0.68 und 0.74. Das bedeutet, der Prozentsatz der Kinder ohne festgestellte Minderleistungen im Rechnen zum Ende der Klasse eins, die durch den jeweiligen Prädiktor auch als unauffällig eingestuft wurden, bezogen auf alle Kinder mit geringen Rechenleistungen zum Ende der Klassenstufe 1, beträgt zwischen 67.9 % und 74.4 %. Die α-Fehlerquote ist beim CBM „Zahlenreihe im Zwanzigerraum“ mit Werten zwischen 25.6 % und 32.1 % über die Messzeitpunkte recht hoch. Insgesamt weisen die Ergebnisse darauf hin, dass ein Viertel bis ein Drittel der Kinder auf Grundlage der Daten des CBM „Zahlenreihe im Zwanzigerraum“ fälschlicherweise als auffällig eingeschätzt wird.

Die Prädiktortrefferquote liegt für das CBM „Zahlenreihe im Zwanzigerraum“ über die Messzeitpunkte in einem Intervall von 0.23 bis 0.30. Dies weist darauf hin, dass 23.0 % bis 30.0 % der Kinder, welche anhand des CBM als auffällig eingestuft wurden, zum Ende der Klasse eins tatsächlich Minderleistungen im Rechnen aufweisen. Der negative prädiktive Wert deutet mit Werten zwischen 0.92 und 0.95 über die Messzeitpunkte hinweg darauf hin, dass der Anteil der im CBM „Zahlenreihe im Zwanzigerraum“ korrekterweise als unauffällig klassifizierten Kinder über die Schulwochen über 90 % liegt.

Die RATZ-Indices zu den verschiedenen Messzeitpunkten liegen in einem Bereich zwischen 38 und 66 %, d. h. dass die Vorhersagen auf Basis des CBM „Zahlenreihe im Zwanzigerraum“ sich leicht bis

deutlich gegenüber einer zufälligen Zuordnung verbessern und das Verfahren eine gute bis nahezu sehr gute Klassifikationsleistung aufweist.

Tabelle 65: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Zahlenreihe im Zwanzigerraum“ (PR ≤ 25)

		Klassifikation				Kennwerte						
Schulwoche	Cutoff-Wert in RW	richtig positiv	falsch positiv	falsch negativ	richtig negativ	Sensitivität	Spezifität	positiver prädiktiver Wert	negativer prädiktiver Wert	Anteil falsch Positiver (α-Fehlerquote) in %	Anteil falsch Negativer (β-Fehlerquote) in %	RATZ-Index
5	1	36	86	16	250	0.69	0.74	0.30	0.94	25.6	30.8	0.55
8	4	41	107	11	226	0.79	0.68	0.28	0.95	32.1	21.2	0.66
12	7	34	101	17	227	0.67	0.69	0.25	0.93	30.8	33.3	0.48
16	8	30	101	20	223	0.60	0.69	0.23	0.92	31.2	40.0	0.38

Erläuterungen: RW – Rohwert

CBM „Mengenvergleich im Zwanzigerraum“

Auch das CBM „Mengenvergleich im Zwanzigerraum“ wurde hinsichtlich seiner Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins geprüft. Die Ergebnisse der Analysen sind in Tabelle 66 abgebildet. Für alle Schulwochen sind die festgelegten Cutoff-Werte, die daraus resultierenden Klassifikationen sowie die wesentlichen Kennwerte zur Beurteilung der Vorhersagegüte des CBM „Mengenvergleich im Zwanzigerraum“ aufgeführt.

Über die Messzeitpunkte hinweg variieren die Werte der Sensitivität in einem Intervall zwischen 0.48 und 0.73, d. h., dass das CBM „Mengenvergleich im Zwanzigerraum“ über die Schulwochen einen Prozentsatz von 48 % bis 73 % der Kinder mit Minderleistungen im

Rechnen zum Ende der Klasse eins korrekt identifizieren kann. Die β-Fehlerquote fällt für das CBM „Mengenvergleich im Zwanzigerraum“ fällt mit Werten zwischen 27 % und 52 % über die Messzeitpunkte relativ hoch aus. Insgesamt weisen die Ergebnisse darauf hin, dass über ein Viertel bis eine Hälfte der Kinder auf Grundlage der Daten des CBM „Mengenvergleich im Zwanzigerraum“ fälschlicherweise als unauffällig eingeschätzt wird.

Für das CBM „Mengenvergleich im Zwanzigerraum“ ergeben sich über die verschiedenen Messzeitpunkte in der ersten Hälfte des ersten Schuljahres Spezifitätswerte im Bereich von 0.66 und 0.84. Das bedeutet, der Prozentsatz der Kinder ohne festgestellte Minderleistungen im Rechnen zum Ende der Klasse eins, die durch den jeweiligen Prädiktor auch als unauffällig eingestuft wurden, bezogen auf alle Kinder mit geringen Rechenleistungen zum Ende der Klassenstufe 1, beträgt zwischen 66 % und 84 %. Die α-Fehlerquote ist beim CBM „Mengenvergleich im Zwanzigerraum“ mit Werten zwischen etwa 16 % und 34 % über die Messzeitpunkte recht hoch. Insgesamt weisen die Ergebnisse darauf hin, dass ein Sechstel bis ein Drittel der Kinder auf Grundlage der Daten des CBM „Mengenvergleich im Zwanzigerraum“ fälschlicherweise als auffällig eingeschätzt wird.

Die Prädiktortrefferquote liegt für das CBM „Mengenvergleich im Zwanzigerraum“ über die Messzeitpunkte in einem Intervall von 0.23 bis 0.34. Dies weist darauf hin, dass 23 % bis 34 % der Kinder, welche anhand des CBM als auffällig eingestuft wurden, zum Ende der Klasse eins tatsächlich Minderleistungen im Rechnen aufweisen. Der negative prädiktive Wert gibt mit Werten zwischen 0.91 und 0.95 über die Messzeitpunkte hinweg an, dass der Anteil der im CBM „Mengenvergleich im Zwanzigerraum“ korrekterweise als unauffällig klassifizierten Kinder über die Schulwochen über 90 % liegt.

Die RATZ-Indices zu den verschiedenen Messzeitpunkten liegen in einem Bereich zwischen 35 und 62 %, d. h., dass die Vorhersagen auf Basis des CBM „Mengenvergleich im Zwanzigerraum“ sich leicht gegenüber einer zufälligen Zuordnung verbessern und das Verfahren eine gute Klassifikationsleistung aufweist.

Tabelle 66: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Mengenvergleich im Zwanzigerraum“ (PR ≤ 25)

		Klassifikation				Kennwerte						
Schulwoche	Cutoff-Wert in RW	richtig positiv	falsch positiv	falsch negativ	richtig negativ	Sensitivität	Spezifität	positiver prädiktiver Wert	negativer prädiktiver Wert	Anteil falsch Positiver (α-Fehlerquote) in %	Anteil falsch Negativer (β-Fehlerquote) in %	RATZ-Index
5	10	25	53	27	283	0.48	0.84	0.32	0.91	15.8	51.9	0.35
8	17	38	75	14	255	0.73	0.77	0.34	0.95	22.7	26.9	0.62
12	23	34	80	17	247	0.67	0.76	0.30	0.94	24.5	33.3	0.52
16	27	32	109	17	215	0.65	0.66	0.23	0.93	33.6	34.7	0.44

Erläuterungen: RW – Rohwert

CBM „Arithmetische Basiskompetenzen“

In Tabelle 67 sind die Ergebnisse des CBM „Arithmetische Basiskompetenzen“ zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins aufgeführt. Getrennt nach Schulwochen sind dort zunächst die aufgrund der festgesetzten Cutoff-Werte resultierenden Klassifikationen sowie anschließend die wesentlichen Kennwerte zur Beurteilung der Vorhersagegüte des CBM „Arithmetische Basiskompetenzen“ notiert.

Über die Schulwochen hinweg variieren die Werte der Sensitivität in einem Intervall zwischen 0.54 und 0.92, d. h. dass das CBM

„Arithmetische Basiskompetenzen“ über die Schulwochen einen Prozentsatz von 54 % bis 92 % der Kinder mit Minderleistungen im Rechnen zum Ende der Klasse eins korrekt identifizieren kann. Für das CBM „Arithmetische Basiskompetenzen“ schwanken die Werte für die β-Fehlerquote zwischen 8.2 und 46.2 % über die Messzeitpunkte. Insgesamt weisen die Ergebnisse darauf hin, dass über ein Zehntel bis knapp die Hälfte der Kinder auf Grundlage der Daten des CBM „Arithmetische Basiskompetenzen“ fälschlicherweise als unauffällig eingeschätzt wird.

Für das CBM „Arithmetische Basiskompetenzen“ ergeben sich über die verschiedenen Messzeitpunkte im ersten Schuljahr Spezifitätswerte im Bereich von 0.39 und 0.88. Das bedeutet, der Prozentsatz der Kinder ohne festgestellte Minderleistungen im Rechnen zum Ende der Klasse eins, die durch den jeweiligen Prädiktor auch als unauffällig eingestuft wurden, bezogen auf alle Kinder mit geringen Rechenleistungen zum Ende der Klassenstufe 1, beträgt zwischen 39 und 88 %. Die α-Fehlerquoten des CBM „Arithmetische Basiskompetenzen“ streuen relativ breit in einem Intervall zwischen 12.2 und 60.9 % über die Messzeitpunkte. Insgesamt weisen die Ergebnisse darauf hin, dass ein Achtel bis zwei Drittel der Kinder auf Grundlage der Daten des CBM „Arithmetische Basiskompetenzen“ fälschlicherweise als auffällig eingeschätzt wird.

Die Prädiktortrefferquote liegt für das CBM „Arithmetische Basiskompetenzen“ über die Messzeitpunkte in einem Intervall von 0.18 bis 0.41. Dies weist darauf hin, dass 18 % bis 41 % der Kinder, welche anhand des CBM als auffällig eingestuft wurden, zum Ende der Klasse eins tatsächlich Minderleistungen im Rechnen aufweisen. Mit Werten zwischen 0.92 und 0.97 über die Messzeitpunkte hinweg gibt der negative prädiktive Wert an, dass der Anteil der im CBM „Arithmetische Basiskompetenzen“ korrekterweise als unauffällig klassifizierten Kinder über die Schulwochen über 90 % liegt.

Die RATZ-Indices zu den verschiedenen Messzeitpunkten liegen in einem Bereich zwischen 41 % und 77 %, d. h., dass die Vorhersagen auf Basis des CBM „Arithmetische Basiskompetenzen“ sich leicht bis deutlich gegenüber einer zufälligen Zuordnung verbessern und das Verfahren eine gute bis sehr gute Klassifikationsleistung aufweist.

Tabelle 67: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Arithmetische Basiskompetenzen““ (PR ≤ 25)

		Klassifikation				Kennwerte						
Schulwoche	Cutoff-Wert in RW	richtig positiv	falsch positiv	falsch negativ	richtig negativ	Sensitivität	Spezifität	positiver prädiktiver Wert	negativer prädiktiver Wert	Anteil falsch Positiver (α-Fehlerquote) in %	Anteil falsch Negativer (β-Fehlerquote) in %	RATZ-Index
5	1	28	41	24	296	0.54	0.88	0.41	0.93	12.2	46.2	0.44
8	3	33	107	20	226	0.62	0.68	0.24	0.92	32.1	37.7	0.41
12	4	37	115	14	212	0.73	0.65	0.24	0.94	35.2	27.5	0.54
16	6	37	142	12	180	0.76	0.56	0.21	0.94	44.1	24.5	0.53
21	6	34	111	15	211	0.69	0.66	0.23	0.93	34.5	30.6	0.50
26	8	45	199	4	128	0.92	0.39	0.18	0.97	60.9	8.2	0.77
30	8	39	116	11	210	0.78	0.64	0.25	0.95	35.6	22.0	0.63
34	9	37	157	5	123	0.88	0.44	0.19	0.96	56.1	11.9	0.70

Erläuterungen: RW – Rohwert

CBM „Zahlzerlegung im Zwanzigerraum“

Die Ergebnisse zur Analyse der prognostischen Validität des CBM „Zahlzerlegung im Zwanzigerraum“ hinsichtlich auftretender Minderleistungen im Rechnen zum Ende der Klasse eins sind in Tabelle 68 dargestellt. Auch hier werden die verwendeten Cutoff-Werte, die daraus resultierenden Klassifikationen sowie die

wesentlichen Kennwerte zur Beurteilung der prognostischen Validität des CBM „Zahlenreihe im Zwanzigerraum“ getrennt nach Schulwochen angegeben.

Über die Messzeitpunkte hinweg variieren die Werte der Sensitivität in einem Intervall zwischen 0.81 und 0.96, d. h. dass das CBM „Zahlzerlegung im Zwanzigerraum“ über die Schulwochen einen Prozentsatz von 81 % bis 96 % der Kinder mit Minderleistungen im Rechnen zum Ende der Klasse eins korrekt identifizieren kann. Für das CBM „Zahlzerlegung im Zwanzigerraum“ liegt die β-Fehlerquote über die Messzeitpunkte zwischen 4.3 % und 19 %. Insgesamt weisen die Ergebnisse darauf hin, dass bis zu einem Fünftel der Kinder auf Grundlage der Daten des CBM „Zahlzerlegung im Zwanzigerraum“ fälschlicherweise als unauffällig eingeschätzt wird.

Für das CBM „Zahlzerlegung im Zwanzigerraum“ ergeben sich über die verschiedenen Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres Spezifitätswerte im Bereich von 0.59 und 0.83. Das bedeutet, der Prozentsatz der Kinder ohne festgestellte Minderleistungen im Rechnen zum Ende der Klasse eins, die durch den jeweiligen Prädiktor auch als unauffällig eingestuft wurden, bezogen auf alle Kinder mit geringen Rechenleistungen zum Ende der Klassenstufe 1, beträgt zwischen 59 % und 83 %. Die α-Fehlerquote fällt beim CBM „Zahlzerlegung im Zwanzigerraum“ mit Werten zwischen 17.1 % und 41.3 % über die Messzeitpunkte recht hoch aus. Insgesamt weisen die Ergebnisse darauf hin, dass ein bis zwei Fünftel der Kinder auf Grundlage der Daten des CBM „Zahlzerlegung im Zwanzigerraum“ fälschlicherweise als auffällig eingeschätzt werden.

Der positive prädiktive Wert liegt für das CBM „Zahlzerlegung im Zwanzigerraum“ über die Messzeitpunkte in einem Intervall von 0.25 bis 0.42. Dies weist darauf hin, dass 25 % bis 42 % der Kinder,

welche anhand des CBM als auffällig eingestuft wurden, zum Ende der Klasse eins tatsächlich Minderleistungen im Rechnen aufweisen. Mit Werten zwischen 0.97 und 0.99 über die Messzeitpunkte hinweg gibt der negative prädiktive Wert an, dass der Anteil der im CBM „Zahlzerlegung im Zwanzigerraum“ korrekterweise als unauffällig klassifizierten Kinder über die Schulwochen deutlich über 90 % liegt.

Die RATZ-Indices zu den verschiedenen Messzeitpunkten liegen in einem Bereich zwischen 74 % und 92 %, d. h., dass die Vorhersagen auf Basis des CBM „Zahlzerlegung im Zwanzigerraum“ sich deutlich gegenüber einer zufälligen Zuordnung verbessern und das Verfahren eine durchweg eine sehr gute Klassifikationsleistung aufweist.

Tabelle 68: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Zahlzerlegung im Zwanzigerraum““ (PR ≤ 25)

		Klassifikation				Kennwerte						
Schulwoche	Cutoff-Wert in RW	richtig positiv	falsch positiv	falsch negativ	richtig negativ	Sensitivität	Spezifität	positiver prädiktiver Wert	negativer prädiktiver Wert	Anteil falsch Positiver (α-Fehlerquote) in %	Anteil falsch Negativer (β-Fehlerquote) in %	RATZ-Index
21	2	45	131	2	186	0.96	0.59	0.26	0.99	41.3	4.3	0.92
26	3	41	83	7	243	0.85	0.75	0.33	0.97	25.5	14.6	0.78
30	6	44	130	6	193	0.88	0.60	0.25	0.97	40.2	12.0	0.78
34	4	34	48	8	232	0.81	0.83	0.42	0.97	17.1	19.0	0.74

Erläuterungen: RW – Rohwert

CBM „Addition im Zwanzigerraum“

Die Ergebnisse der Analysen zur prädiktiven Qualität des CBM „Addition im Zwanzigerraum“ sind in Tabelle 69 dargestellt. sind die Ergebnisse des zur Vorhersagegüte von Minderleistungen im

Rechnen zum Ende der Klasse eins aufgeführt. Neben den festgelegten Cutoff-Werten sowie den daraus resultierenden Klassifikationen werden wesentliche Kennwerte zur Beurteilung der prädiktiven Validität des CBM „Addition im Zwanzigerraum“ getrennt nach Schulwochen aufgeführt.

Über die Messzeitpunkte hinweg variieren die Werte der Sensitivität in einem Intervall zwischen 0.86 und 0.98, d. h., dass das CBM „Addition im Zwanzigerraum“ über die Schulwochen einen Prozentsatz von 86 % bis 98 % der Kinder mit Minderleistungen im Rechnen zum Ende der Klasse eins korrekt identifizieren kann. Für das CBM „Addition im Zwanzigerraum“ liegt die β-Fehlerquote über die Messzeitpunkte zwischen 2.1 und 14 %. Insgesamt weisen die Ergebnisse darauf hin, dass bis zu einem Siebtel der Kinder auf Grundlage der Daten des CBM „Addition im Zwanzigerraum“ fälschlicherweise als unauffällig eingeschätzt wird.

Für das CBM „Addition im Zwanzigerraum“ ergeben sich über die verschiedenen Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres Spezifitätswerte im Bereich von 0.53 und 0.77. Das bedeutet, der Prozentsatz der Kinder ohne festgestellte Minderleistungen im Rechnen zum Ende der Klasse eins, die durch den jeweiligen Prädiktor auch als unauffällig eingestuft wurden, bezogen auf alle Kinder mit geringen Rechenleistungen zum Ende der Klassenstufe 1, beträgt zwischen 53 % und 77 %. Die α-Fehlerquote ist beim CBM „Addition im Zwanzigerraum“ mit Werten zwischen 23.5 % und 47.3 % über die Messzeitpunkte recht hoch. Insgesamt weisen die Ergebnisse darauf hin, dass ein Viertel bis knapp die Hälfte der Kinder auf Grundlage der Daten des CBM „Addition im Zwanzigerraum“ fälschlicherweise als auffällig eingeschätzt wird.

Der positive prädiktive Wert liegt für das CBM „Addition im Zwanzigerraum“ über die Messzeitpunkte in einem Intervall von 0.24 bis 0.36. Dies weist darauf hin, dass 24 % bis 36 % der Kinder, welche anhand des CBM als auffällig eingestuft wurden, zum Ende der Klasse eins tatsächlich Minderleistungen im Rechnen aufweisen. Mit Werten zwischen 0.97 und 0.99 über die Messzeitpunkte hinweg gibt der negative prädiktive Wert an, dass der Anteil der im CBM „Addition im Zwanzigerraum“ korrekterweise als unauffällig klassifizierten Kinder über die Schulwochen deutlich über 90 % liegt.

Die RATZ-Indices zu den verschiedenen Messzeitpunkten liegen in einem Bereich zwischen 79 % und 95 %, d. h., dass die Vorhersagen auf Basis des CBM „Addition im Zwanzigerraum“ sich deutlich gegenüber einer zufälligen Zuordnung verbessern und das Verfahren eine durchweg eine sehr gute Klassifikationsleistung aufweist.

Tabelle 69: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Addition im Zwanzigerraum““ (PR ≤ 25)

		Klassifikation				Kennwerte						
Schulwoche	Cutoff-Wert in RW	richtig positiv	falsch positiv	falsch negativ	richtig negativ	Sensitivität	Spezifität	positiver prädiktiver Wert	negativer prädiktiver Wert	Anteil falsch Positiver (α-Fehlerquote) in %	Anteil falsch Negativer (β-Fehlerquote) in %	RATZ-Index
21	5	46	149	1	166	0.98	0.53	0.24	0.99	47.3	2.1	0.95
26	6	44	103	5	222	0.90	0.68	0.30	0.98	31.7	10.2	0.83
30	7	43	76	7	247	0.86	0.77	0.36	0.97	23.5	14.0	0.79
34	8	40	91	2	189	0.95	0.68	0.31	0.99	32.5	4.8	0.92

Erläuterungen: RW – Rohwert

CBM „Subtraktion im Zwanzigerraum“

In Tabelle 70 sind die Ergebnisse des CBM „Subtraktion im Zwanzigerraum“ zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins aufgeführt. Für die verschiedenen Schulwochen werden die aufgrund der festgesetzten Cutoff-Werte resultierenden Klassifikationen sowie die wesentlichen Kennwerte zur Beurteilung der Vorhersagegüte des CBM „Subtraktion im Zwanzigerraum“ angegeben.

Über die Messzeitpunkte hinweg variieren die Werte der Sensitivität in einem Intervall zwischen 0.76 und 0.96, d. h., dass das CBM „Subtraktion im Zwanzigerraum“ über die Schulwochen einen Prozentsatz von 76 % bis 96 % der Kinder mit Minderleistungen im Rechnen zum Ende der Klasse eins korrekt identifizieren kann. Für das CBM „Subtraktion im Zwanzigerraum“ liegt die β-Fehlerquote über die Messzeitpunkte zwischen 4.1 % und 24 %. Insgesamt weisen die Ergebnisse darauf hin, dass bis zu einem Viertel der Kinder auf Grundlage der Daten des CBM „Subtraktion im Zwanzigerraum“ fälschlicherweise als unauffällig eingeschätzt wird.

Für das CBM „Subtraktion im Zwanzigerraum“ ergeben sich über die verschiedenen Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres Spezifitätswerte im Bereich von 0.45 und 0.76. Das bedeutet, der Prozentsatz der Kinder ohne festgestellte Minderleistungen im Rechnen zum Ende der Klasse eins, die durch den jeweiligen Prädiktor auch als unauffällig eingestuft wurden, bezogen auf alle Kinder mit geringen Rechenleistungen zum Ende der Klassenstufe 1, beträgt zwischen 45 % und 76 %. Die α-Fehlerquote ist beim CBM „Subtraktion im Zwanzigerraum“ mit Werten zwischen 23.8 und 54.9 % über die Messzeitpunkte recht hoch. Insgesamt weisen die Ergebnisse darauf hin, dass ein Viertel bis knapp über die Hälfte der Kinder auf Grundlage der Daten des

CBM „Subtraktion im Zwanzigerraum“ fälschlicherweise als auffällig eingeschätzt wird.

Der positive prädiktive Wert liegt für das CBM „Subtraktion im Zwanzigerraum“ über die Messzeitpunkte in einem Intervall von 0.21 bis 0.33. Dies weist darauf hin, dass 21 % bis 33 % der Kinder, welche anhand des CBM als auffällig eingestuft wurden, zum Ende der Klasse eins tatsächlich Minderleistungen im Rechnen aufweisen. Der negative prädiktive Wert gibt mit Werten zwischen 0.95 und 0.99 über die Messzeitpunkte hinweg an, dass der Anteil der im CBM „Subtraktion im Zwanzigerraum“ korrekterweise als unauffällig klassifizierten Kinder über die Schulwochen deutlich über 90 % liegt.

Die RATZ-Indices zu den verschiedenen Messzeitpunkten liegen in einem Bereich zwischen 65 % und 92 %, d. h., dass die Vorhersagen auf Basis des CBM „Subtraktion im Zwanzigerraum“ sich leicht bis deutlich gegenüber einer zufälligen Zuordnung verbessern und das Verfahren eine überwiegend eine sehr gute Klassifikationsleistung aufweist.

Tabelle 70: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Subtraktion im Zwanzigerraum““ (PR ≤ 25)

		Klassifikation				Kennwerte						
Schulwoche	Cutoff-Wert in RW	richtig positiv	falsch positiv	falsch negativ	richtig negativ	Sensitivität	Spezifität	positiver prädiktiver Wert	negativer prädiktiver Wert	Anteil falsch Positiver (α-Fehlerquote) in %	Anteil falsch Negativer (β-Fehlerquote) in %	RATZ-Index
21	3	45	174	2	143	0.96	0.45	0.21	0.99	54.9	4.3	0.89
26	4	47	137	2	189	0.96	0.58	0.26	0.99	42.0	4.1	0.92
30	4	38	77	12	246	0.76	0.76	0.33	0.95	23.8	24.0	0.65
34	5	35	84	7	195	0.83	0.70	0.29	0.97	30.1	16.7	0.74

Erläuterungen: RW – Rohwert

7.3.3.3 Inkrementelle Validität

Nachfolgend werden die Ergebnisse hinsichtlich der inkrementellen Validität in der vorliegenden Studie erarbeiteten CBM dargestellt. Dabei wird die Güte zur Prognose schulischer Minderleistungen im Bereich Mathematik zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) bzw. zum Ende der Klasse zwei (PR ≤ 15 im DEMAT 2+; Krajewski et al., 2004) unter Berücksichtigung der Daten aller eingesetzter CBM der Niveaustufe 1 bzw. 2 in Kombination analysiert.

Vorhersage schulischer Minderleistungen im Rechnen zum Ende der Klasse eins anhand der CBM der Niveaustufe 1

Das berechnete binär-logistische Regressionsmodell zur Vorhersage der Daten des DEMAT 1+ (Krajewski et al., 2002) zum Ende der ersten Klasse anhand der Ergebnisse der CBM der Niveaustufe 1 zum ersten Messzeitpunkt in der fünften Schulwoche gibt mit $X^2 = 61.246$ ($df = 4$, $p < .001$) an, dass die gewählten Prädiktoren einen signifikanten Zuwachs bei der Modellanpassung liefern. Die Ergebnisse der Regressionsanalyse werden in Tabelle 71 aufgeführt. Es fällt auf, dass die Zugehörigkeit zur Risikogruppe am Ende der Klassenstufe 1 durch die Ergebnisse der CBM „Zahlenlesen im Zwanzigerraum“, „Mengenvergleich im Zwanzigerraum“ sowie „Arithmetische Basiskompetenzen“ in signifikanter Weise vorhergesagt werden kann. Für diesen Messzeitpunkt stellen diese CBM gültige Prädiktoren zur Vorhersage schulischer Minderleistungen zum Ende der Klasse eins dar. Einzig die Ergebnisse des CBM „Zahlenreihe im Zwanzigerraum“ zum ersten Messzeitpunkt in der fünften Schulwoche eignen sich, aufgrund eines nicht-signifikanten Regressionsgewichtes, nicht zur Vorhersage der Zugehörigkeit zur Risikogruppe zum Ende der Klassenstufe 1 und werden nicht weiter analysiert.

Tabelle 71: Ergebnisse der logistischen Regressionsanalyse der CBM der Niveaustufe 1 des ersten Messzeitpunkts in der fünften Schulwoche zur Vorhersage der Zugehörigkeit zur Gruppe der Risikokinder im Bereich Mathematik, operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 1+ (Krajewski et al., 2002) zum Ende der Klassenstufe 1 (N = 387)

Prädiktoren	b	p	exp(b)
CBM „Zahlenlesen im Zwanzigerraum“	-.036	.022	.964
CBM „Zahlenreihe im Zwanzigerraum“	-.085	.293	.919
CBM „Mengenvergleich im Zwanzigerraum“	-.047	.036	.954
CBM „Arithmetische Basiskompetenzen“	-.232	.004	.793

Erläuterungen: b = Regressionskoeffizient; p = Irrtumswahrscheinlichkeit; exp(b) = Odds ratio

Die negativen Vorzeichen der Regressionskoeffizienten b zeigen an, dass sich das Risiko einer Minderleistung im Bereich Rechnen mit

steigenden Leistungen in den CBM jeweils vermindert. Um die Stärke des Einflusses der Prädiktoren zu bewerten, werden die exp(b)-Koeffizienten der CBM „Zahlenlesen im Zwanzigerraum“, „Mengenvergleich im Zwanzigerraum“ und „Arithmetische Basiskompetenzen“ analysiert. Diese können als Faktor der Vervielfachung der odds ratio interpretiert werden, wenn sich die Prädiktorvariable um eine Einheit verändert. Der Koeffizient exp(b) für das CBM „Zahlenlesen im Zwanzigerraum“ beträgt exp(-.036) = .964. Dies bedeutet, dass sich bei Zunahme der Prädiktorvariablen CBM „Zahlenlesen im Zwanzigerraum“ um eine Einheit, das Verhältnis der Wahrscheinlichkeit, eine Minderleistung im Rechnen auszubilden zur Wahrscheinlichkeit, keine Minderleistung auszubilden um etwa 3.6 % (100 • (1 – 0.964)= 3.6) verringert. Mit einem exp(b) von exp(-.047) = .954 verringert sich für das CBM „Mengenvergleich im Zwanzigerraum“ das Verhältnis der Wahrscheinlichkeiten um etwa 4.6 % (100 • (1 – 0.954)= 3.6), beim CBM „Arithmetische Basiskompetenzen“ verringert es sich mit einem exp(b) von exp(-.232) = .793 um etwa 20.7 % (100 • (1 – 0.793)= 2.07). Der Wert des Nagelkerkes-R^2 = .27 gibt an, dass durch das Regressionsmodell insgesamt 27 % der Varianz in den Rechenleistungen zum Ende des ersten Schuljahres vorhergesagt werden.

Vorhersage schulischer Minderleistungen im Rechnen zum Ende der Klasse eins anhand der CBM der Niveaustufe 2

Das berechnete binär-logistische Regressionsmodell zur Vorhersage der Daten des DEMAT 1+ (Krajewski et al., 2002) zum Ende der ersten Klasse anhand der Ergebnisse der CBM der Niveaustufe 2 zum fünften Messzeitpunkt in der 21. Schulwoche gibt mit X^2 = 120.97 (df = 4, p < .001) an, dass die gewählten Prädiktoren einen signifikanten Zuwachs bei der Modellanpassung liefern. Die Ergebnisse der Regressionsanalyse werden in Tabelle 72 aufgeführt. Es fällt auf, dass die Zugehörigkeit zur Risikogruppe durch die

Ergebnisse der CBM „Zahlzerlegung im Zwanzigerraum“ und „Addition im Zwanzigerraum“ in signifikanter Weise vorhergesagt werden kann. Für diesen Messzeitpunkt stellen diese CBM gültige Prädiktoren zur Vorhersage schulischer Minderleistungen zum Ende der Klasse eins dar. Die Ergebnisse des CBM „Arithmetische Basiskompetenzen“ sowie das CBM „Subtraktion im Zwanzigerraum“ zum fünften Messzeitpunkt in der 21. Schulwoche eignen sich, aufgrund eines nicht-signifikanten Regressionsgewichtes, nicht zur Vorhersage der Zugehörigkeit zur Risikogruppe zum Ende der Klassenstufe 1 und werden nicht weiter analysiert.

Tabelle 72: Ergebnisse der logistischen Regressionsanalyse der CBM der Niveaustufe 2 des fünften Messzeitpunkts in der 21. Schulwoche zur Vorhersage der Zugehörigkeit zur Gruppe der Risikokinder im Bereich Mathematik, operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 1+ (Krajewski et al., 2002) zum Ende der Klassenstufe 1 (N = 362)

Prädiktoren	b	p	exp(b)
CBM „Arithmetische Basiskompetenzen“	.063	.39	1.065
CBM „Zahlzerlegung im Zwanzigerraum“	-.515	.008	.598
CBM „Addition im Zwanzigerraum“	-.71	.000	.492
CBM „Subtraktion im Zwanzigerraum“	-.091	.059	.913

Erläuterungen: b = Regressionskoeffizient; p = Irrtumswahrscheinlichkeit; exp(b) = Odds ratio

Die negativen Vorzeichen der Regressionskoeffizienten b für die CBM „Zahlzerlegung im Zwanzigerraum“ sowie „Addition im Zwanzigerraum“ zeigen an, dass sich das Risiko einer Minderleistung im Bereich Rechnen mit steigenden Leistungen in den CBM jeweils vermindert. Um die Stärke des Einflusses der Prädiktoren zu bewerten, werden wieder die exp(b)-Koeffizienten analysiert. Diese können als Faktor der Vervielfachung der odds ratio interpretiert werden, wenn sich die Prädiktorvariable um eine Einheit verändert. Der Koeffizient exp(b) für das CBM „Zahlzerlegung im Zwanzigerraum“ beträgt exp(-.515) = .598. Dies bedeutet, dass sich bei Zunahme der Prädiktorvariablen CBM „Zahlzerlegung im Zwanzigerraum“ um eine Einheit, das Verhältnis der Wahrscheinlichkeit, eine Minderleistung im Rechnen auszubilden zur

Wahrscheinlichkeit, keine Minderleistung auszubilden um etwa 40.2 % (100 • (1 – 0.598)= 40.2) verringert. Mit einem exp(b) von exp(-.71) = .492 verringert sich für das CBM „Addition im Zwanzigerraum“ das Verhältnis der Wahrscheinlichkeiten um etwa 50.8 % (100 • (1 – 0.492)= 50.8). Der Wert des Nagelkerkes-R^2 = .528 gibt an, dass durch das Regressionsmodell insgesamt mehr als die Hälfte der Varianz (52.8 %) in den Rechenleistungen zum Ende des ersten Schuljahres vorhergesagt werden.

Vorhersage schulischer Minderleistungen im Rechnen zum Ende der Klasse zwei anhand der CBM der Niveaustufe 1

Das berechnete binär-logistische Regressionsmodell zur Vorhersage der Daten des DEMAT 2+ zum Ende der zweiten Klasse anhand der Ergebnisse der CBM der Niveaustufe 1 zum ersten Messzeitpunkt in der fünften Schulwoche gibt mit X^2 = 86.754 (df = 4, p < .001) an, dass die gewählten Prädiktoren einen signifikanten Zuwachs bei der Modellanpassung liefern. Die Ergebnisse der Regressionsanalyse werden in Tabelle 73 aufgeführt. Es fällt auf, dass die Zugehörigkeit zur Risikogruppe (PR ≤ 15) durch die Ergebnisse der CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ sowie „Mengenvergleich im Zwanzigerraum“ in signifikanter Weise vorhergesagt werden kann. Für diesen Messzeitpunkt stellen diese CBM gültige Prädiktoren zur Vorhersage schulischer Minderleistungen zum Ende der Klasse zwei dar. Einzig die Ergebnisse des CBM „Arithmetische Basiskompetenzen“ zum ersten Messzeitpunkt in der fünften Schulwoche eignen sich, aufgrund eines nicht-signifikanten Regressionsgewichtes, nicht zur Vorhersage der Zugehörigkeit zur Risikogruppe zum Ende der Klassenstufe 2 und werden in den sich anschließenden differenzierten Analysen nicht weiter berücksichtigt.

Tabelle 73: Ergebnisse der logistischen Regressionsanalyse der CBM der Niveaustufe 1 des ersten Messzeitpunkts in der fünften Schulwoche zur Vorhersage der Zugehörigkeit zur Gruppe der Risikokinder im Bereich Mathematik, operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 2+ zum Ende der Klassenstufe 2 (N = 388)

Prädiktoren	b	p	exp(b)
CBM „Zahlenlesen im Zwanzigerraum“	-.041	.002	.96
CBM „Zahlenreihe im Zwanzigerraum“	-.175	.011	.839
CBM „Mengenvergleich im Zwanzigerraum“	-.055	.004	.946
CBM „Arithmetische Basiskompetenzen“	-.07	.251	.933

Erläuterungen: b = Regressionskoeffizient; p = Irrtumswahrscheinlichkeit; exp(b) = Odds ratio

Die negativen Vorzeichen der Regressionskoeffizienten b zeigen an, dass sich das Risiko einer Minderleistung im Bereich Rechnen mit steigenden Leistungen in den CBM jeweils vermindert. Um die Stärke des Einflusses der Prädiktoren zu bewerten, werden die exp(b)-Koeffizienten der CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ und „Mengenvergleich im Zwanzigerraum“ analysiert. Diese können als Faktor der Vervielfachung der odds ratio interpretiert werden, wenn sich die Prädiktorvariable um eine Einheit verändert. Der Koeffizient exp(b) für das CBM „Zahlenlesen im Zwanzigerraum“ beträgt exp(-.041) = .96. Dies bedeutet, dass sich bei Zunahme der Prädiktorvariablen CBM „Zahlenlesen im Zwanzigerraum“ um eine Einheit, das Verhältnis der Wahrscheinlichkeit, eine Minderleistung im Rechnen auszubilden zur Wahrscheinlichkeit, keine Minderleistung auszubilden um etwa 4 % (100 • (1 – 0.96)= 4) verringert. Mit einem exp(b) von exp(-.175) = .839 verringert sich für das CBM „Zahlenreihe im Zwanzigerraum“ das Verhältnis der Wahrscheinlichkeiten um etwa 16.1 % (100 • (1 – 0.839)= 16.1), beim CBM „Mengenvergleich im Zwanzigerraum“ verringert es sich mit einem exp(b) von exp(-.055) = .933 um etwa 6.7 % (100 • (1 – 0.933)= 6.7). Der Wert des Nagelkerkes-R^2 = .314 gibt an, dass durch das Regressionsmodell insgesamt 31.4 % der Varianz in den Rechenleistungen zum Ende des zweiten Schuljahres vorhergesagt werden.

Vorhersage schulischer Minderleistungen im Rechnen zum Ende der Klasse zwei anhand der CBM der Niveaustufe 2

Das berechnete binär-logistische Regressionsmodell zur Vorhersage der Daten des DEMAT 2+ zum Ende der zweiten Klasse anhand der Ergebnisse der CBM der Niveaustufe 2 zum fünften Messzeitpunkt in der 21. Schulwoche gibt mit $X^2 = 89.677$ ($df = 4$, $p < .001$) an, dass die gewählten Prädiktoren einen signifikanten Zuwachs bei der Modellanpassung liefern. Die Ergebnisse der Regressionsanalyse werden in Tabelle 74 aufgeführt. Es fällt auf, dass die Zugehörigkeit zur Risikogruppe auch zum Ende der Klassenstufe 2 durch die Ergebnisse der CBM „Zahlzerlegung im Zwanzigerraum“ und „Addition im Zwanzigerraum“ der 21. Schulwoche in Klasse eins in signifikanter Weise vorhergesagt werden kann. Für diesen Messzeitpunkt stellen diese CBM gültige Prädiktoren zur Vorhersage schulischer Minderleistungen zum Ende der Klasse zwei dar. Die Ergebnisse des CBM „Arithmetische Basiskompetenzen“ sowie das CBM „Subtraktion im Zwanzigerraum“ zum fünften Messzeitpunkt in der 21. Schulwoche der Klasse eins eignen sich, aufgrund eines nicht-signifikanten Regressionsgewichtes, nicht zur Vorhersage der Zugehörigkeit zur Risikogruppe zum Ende der Klassenstufe 2 und werden in den sich anschließenden differenzierten Analysen nicht weiter berücksichtigt.

Tabelle 74: Ergebnisse der logistischen Regressionsanalyse der CBM der Niveaustufe 2 des fünften Messzeitpunkts in der 21. Schulwoche zur Vorhersage der Zugehörigkeit zur Gruppe der Risikokinder im Bereich Mathematik, operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 1+ (Krajewski et al., 2002) zum Ende der Klassenstufe 2 (N = 377)

Prädiktoren	b	p	exp(b)
CBM „Arithmetische Basiskompetenzen“	-.073	.164	.93
CBM „Zahlzerlegung im Zwanzigerraum“	-.197	.032	.821
CBM „Addition im Zwanzigerraum“	-.309	.000	.734
CBM „Subtraktion im Zwanzigerraum“	-.081	.381	.922

Erläuterungen: b = Regressionskoeffizient; p = Irrtumswahrscheinlichkeit; exp(b) = Odds ratio

Die negativen Vorzeichen der Regressionskoeffizienten b für die CBM „Zahlzerlegung im Zwanzigerraum“ sowie „Addition im Zwanzigerraum“ zeigen an, dass sich das Risiko einer Minderleistung im Bereich Rechnen mit steigenden Leistungen in den CBM jeweils vermindert. Um die Stärke des Einflusses der Prädiktoren zu bewerten, werden erneut die exp(b)-Koeffizienten analysiert. Diese können als Faktor der Vervielfachung der odds ratio interpretiert werden, wenn sich die Prädiktorvariable um eine Einheit verändert. Der Koeffizient exp(b) für das CBM „Zahlzerlegung im Zwanzigerraum“ beträgt exp(-.197) = .821. Dies bedeutet, dass sich bei Zunahme der Prädiktorvariablen CBM „Zahlzerlegung im Zwanzigerraum“ um eine Einheit, das Verhältnis der Wahrscheinlichkeit, eine Minderleistung im Rechnen auszubilden zur Wahrscheinlichkeit, keine Minderleistung auszubilden sich um etwa 17.9 % (100 • (1 – 0.821)= 17.9) verringert. Mit einem exp(b) von exp(-.309) = .734 verringert sich für das CBM „Addition im Zwanzigerraum“ das Verhältnis der Wahrscheinlichkeiten um etwa 26.6 % (100 • (1 – 0.734)= 26.6). Der Wert des Nagelkerkes-R^2 = .333 gibt an, dass durch das Regressionsmodell insgesamt ein Drittel der Varianz (33.3 %) in den Rechenleistungen zum Ende des zweiten Schuljahres vorhergesagt werden.

7.4 Darstellung der Ergebnisse hinsichtlich der Änderungssensibilität

Die Analysen hinsichtlich der Änderungssensibilität der innerhalb der vorliegenden Arbeit konzipierten CBM vollziehen sich in mehreren Schritten. Zunächst sollen dazu die von Klauer (2011) postulierten Kriterien für einen veränderungssensiblen Test geprüft werden (vgl. Abschnitt 3.3.3). Dazu muss gezeigt werden, dass die Korrelationen zwischen den einzelnen Versionen der CBM kleiner werden, je größer der Zeitraum zwischen den Erhebungszeitpunkten ist (Aspekt 1). Überdies soll in der zusammenfassenden Darstellung der Ergebnisse

geprüft werden, ob die CBM nur mittlere Retest- bzw. Paralleltest-Reliabilitäten aufweisen (Aspekt 2) und die Splithalf-Reliabilitäten der CBM höher als die Retest- bzw. Paralleltest-Reliabilitäten ausfallen (Aspekt 3). Zur Untersuchung der Gewährleistung des Aspekts 1 wurden Interkorrelationsmatrizen für jedes CBM über die Messzeitpunkte erstellt und analysiert. Die in Abschnitt 7.2.1, 7.2.2 sowie 7.2.3 bestimmten Reliabilitätskoeffizienten dienen dabei als Grundlage der Analysen zu den Aspekten 2 und 3.

In einem weiteren Schritt zur Untersuchung der Änderungssensibilität der CBM sollen Mehrebenenanalysen (HLM) Aufschluss über die Leistungsentwicklung der Schüler liefern. Dazu werden zunächst unkonditionierte Modelle geschätzt. Anschließend soll anhand eines allgemeinen Modells der Einfluss eines zu Beginn der ersten Klasse erhobenen Risikostatus der Schüler im Bereich Mathematik als eine weitere Prädiktorvariable auf Schülerebene analysiert werden.

7.4.1 CBM „Zahlenlesen im Zwanzigerraum“

In einem ersten Schritt sollen die Interkorrelationen der einzelnen Versionen des CBM „Zahlenlesen im Zwanzigerraum“ analysiert werden. Die jeweiligen Korrelationskoeffizienten über die Messzeitpunkte sind in Tabelle 75 abgebildet. Die Koeffizienten fallen durchweg hoch aus und streuen in einem Intervall von r = .74 und r = .87. Bei näherer Betrachtung der Tabelle zeigt sich eine, ebenso bei Strathmann und Klauer (2010) beschriebene, Simplexstruktur der Daten, d. h., dass Ergebnisse jeweils aufeinander folgender CBM-Erhebungen höher miteinander korreliert sind als die weiter auseinander entfernt liegenden. So zeigt sich innerhalb der ersten Zeile ein Abfall der Korrelationskoeffizienten von links nach rechts. In der letzten Spalte, von oben nach unten betrachtet, ergibt sich eine steigende Tendenz der Koeffizienten.

Tabelle 75: Interkorrelationen der CBM „Zahlenlesen im Zwanzigerraum“ über die MZP 1 – 4 (listenweiser Fallausschluss fehlender Werte, N = 402)

	Korrelation r mit		
	MZP 2 CBM ZL (8.SW)	MZP 3 CBM ZL (12.SW)	MZP 4 CBM ZL (16.SW)
MZP 1 CBM ZL (5.SW)	.87**	.83**	.74**
MZP 2 CBM ZL (8.SW)		.83**	.78**
MZP 3 CBM ZL (12.SW)			.83**

Erläuterungen: MZP – Messzeitpunkt; CBM ZL – CBM „Zahlenlesen im Zwanzigerraum“; SW – Schulwoche; ** – Die Korrelation ist auf dem Niveau von 0.01 2-seitig signifikant.

In einem weiteren Schritt zur Untersuchung der Änderungssensibilität des CBM sollen Mehrebenenanalysen Aufschluss über die Leistungsentwicklung der Schüler liefern. Dazu wird zunächst das unkonditionierte Modell für das CBM „Zahlenlesen im Zwanzigerraum“ geschätzt. Anschließend soll anhand eines allgemeinen Modells der Einfluss eines zu Beginn der ersten Klasse erhobenen Risikostatus der Schüler im Bereich Mathematik als eine weitere Prädiktorvariable auf Schülerebene analysiert werden.

7.4.1.1 Unkonditioniertes Modell

Das unkonditionierte Modell wird durch folgende Regressionsgleichungen auf den Ebenen 1 und 2 definiert:

Modell der Ebene 1: $Y_{ij} = \beta_{0i} + \beta_{1i} \bullet X_{ij} + e_{ij}$

Modelle der Ebene 2: $\beta_{0i} = \gamma_{00} + u_{0i}$

$\beta_{1i} = \gamma_{10} + u_{1i}$

Dabei gibt X_{ij} die Schulwoche an, in welcher eine Version des CBM „Zahlenlesen im Zwanzigerraum“ von den Kindern bearbeitet wurde. Wie in Abschnitt 6.4.3 erwähnt, ist dieser Prädiktor zentriert, sodass β_{0i} das Ausgangsniveau der Schüler zur ersten Testung in der fünften Schulwoche angibt.

Tabelle 76 fasst die Ergebnisse der unkonditionierten Hierarchisch-linearen Modellierung für das CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres zusammen. Die Schüler weisen ein durchschnittliches Ausgangsniveau von fast fünf richtig gelösten Items zur ersten Testung in der fünften Schulwoche auf ($\gamma_{00} = 3.60$). Der mittlere Anstieg γ_{10} von 0.22 gibt das Maß der durchschnittlichen Verbesserung von Schulwoche zu Schulwoche an. Er zeigt an, dass die Kinder mit jeder weiteren vergangenen Schulwoche im Durchschnitt weitere 0.22 Items richtig lösen können, was eine durchschnittliche Verbesserung um einen Punkt nach etwa fünf Schulwochen ergibt. Die beiden geschätzten Regressionsparameter γ_{10} und γ_{00} sind signifikant ($p < .001$) von Null verschieden.

Bei Betrachtung der geschätzten zufälligen Effekte zeigt sich, dass es mit $X^2 = 4540.3$ ($df = 438$, $p < .001$) schülerspezifische Unterschiede zwischen den Regressionskonstanten gibt, welche statistisch signifikant ausfallen. Dies bedeutet, dass die Schüler mit unterschiedlichen Ausgangsniveaus im CBM „Zahlenlesen im Zwanzigerraum“ zur ersten Testung in der fünften Schulwoche starten. Die schülerspezifischen Unterschiede zwischen den Steigungskoeffizienten sind mit $X^2 = 779.02$ ($df = 438$, $p < .001$) ebenso statistisch signifikant. Dies zeigt an, dass sich die untersuchten Schüler unter der Annahme eines linearen Wachstumsmodells hinsichtlich ihres Lernzuwachses über die Schulwochen signifikant voneinander unterscheiden.

Es ergibt sich eine Reliabilität des geschätzten Ausgangsniveaus β_0 von .903, was darauf hinweist, dass etwa 90.3 % der individuellen Unterschiede im Ausgangsniveau durch die wahre Varianz zwischen den Schülern erklärt werden können und nicht durch Messfehler bedingt sind. Der Reliabilitätskoeffizient des geschätzten Anstiegs β_1 fällt mit einem Wert von .433 geringer aus. Er deutet daraufhin, dass

etwa 43.3 % der individuellen Unterschiede im Leistungsanstieg durch die wahre Varianz zwischen den Schülern erklärt werden können.

Tabelle 76: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM "Zahlenlesen im Zwanzigerraum"

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Mittleres Ausgangsniveau γ_{00}	28.27	0.73	38.90	438	< .001
Mittlerer Anstieg γ_{10}	1.32	0.05	29.07	438	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	14.49	210.01	438	4540.30	< .001
Anstieg u_{1i}	0.63	0.40	438	779.02	< .001
Level-1-Fehler e_{ij}	5.83	33.93			

Reliabilität der Regressionskoeffizienten	
Ausgangsniveau β_0	.903
Anstieg β_1	.433

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

7.4.1.2 Bedingtes intercept-and-slope-as-outcome-Modell

Die Berechnungen des unkonditionierten Modells haben signifikante Unterschiede in den Ausgangsniveaus und den Steigungskoeffizienten über die untersuchten Schüler hinweg ergeben. Demnach wird nachfolgend der Einfluss schülerspezifischer Merkmale auf die Regressionsparameter β_0 und β_1 überprüft. Es wird vermutet, dass die zuvor festgestellten signifikanten individuellen Unterschiede in dem Ausgangsniveau β_0 und den Steigungskoeffizienten β_1 durch in unterschiedlichem Maße ausgeprägtes mathematisches Vorwissen erklärbar sind. Um diese Hypothese zu prüfen, wird die Zugehörigkeit zur Gruppe von Kindern mit einem Risiko im Bereich des mathematischen Vorwissens

(operationalisiert durch einen Prozentrang PR ≤ 15 im Diagnoseprogramm Kalkulie Teil 1; Fritz et al., 2007) im intercept-and-slope-as-outcome-Modell berücksichtigt. Dieser Risikostatus wird als Prädiktor auf Ebene 2 als eine dichotome Variable, die mit den Ausprägungen 0 = „kein Risiko“ und 1 = „Risiko“ definiert ist, aufgenommen. Es ergeben sich somit folgende Modelle auf Ebene 1 und 2:

Modell der Ebene 1: $Y_i = \beta_{0i} + \beta_{1i} \bullet X_{ij} + e_{ij}$

Modelle der Ebene 2: $\beta_{0i} = \gamma_{00} + \gamma_{01} \bullet Z_i + u_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11} \bullet Z_i + u_{1i}$

Dabei gibt γ_{01} den Effekt der Zugehörigkeit zur Risikogruppe im mathematischen Vorwissen auf das Ausgangsniveau β_{0i} der Regressionsgeraden an. γ_{00} kann als das mittlere Ausgangsniveau der Kinder ohne Risiko im Bereich des mathematischen Vorwissens zu Beginn des ersten Schuljahres aufgefasst werden. Der Steigungskoeffizient γ_{01} gibt dabei die Differenz der jeweiligen Ausgangsniveaus der Gruppe von Kindern mit im Vergleich zur Gruppe von Kindern ohne Risikostatus an. Der Parameter γ_{10} gilt als Maß für den Effekt auf den Steigungskoeffizienten β_{1i} der Regressionsgeraden, wenn ein Kind nicht der Risikogruppe im mathematischen Vorwissen zugehörig ist. γ_{11} kann als der mittlere Leistungszuwachs der Kinder ohne Risiko im Bereich des mathematischen Vorwissens aufgefasst werden. Der Steigungskoeffizient γ_{11} des slope-as-outcome-Modells gibt dabei die Differenz der jeweiligen wöchentlichen Leistungsanstiege der Gruppe von Kindern mit festgestellten Risiken im mathematischen Vorwissen verglichen mit der Gruppe von Kindern ohne Risikostatus an.

In Tabelle 77 sind die Ergebnisse der allgemeinen Hierarchisch-linearen Modellierung für das CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten

Schuljahres aufgeführt. Das mittlere Ausgangsniveau γ_{00} von 31.80 deutet darauf hin, dass die Schüler, welche nicht der Risikogruppe im mathematischen Vorwissen angehören, im Durchschnitt knapp 32 der vorgegebenen Zahlen zum ersten Messzeitpunkt in der fünften Schulwoche richtig benennen konnten. Hingegen konnten Schüler, welche der Risikogruppe angehören, im Vergleich zu den anderen Kindern zum ersten Messzeitpunkt im Schnitt knapp 15 Zahlen weniger, d. h. etwa 17 Zahlen, identifizieren (γ_{01} = -15.34). Dieser Unterschied fällt statistisch signifikant aus ($p < .001$). Der mittlere Anstieg γ_{10} von 1.25 deutet darauf hin, dass Kinder ohne Risikostatus im Durchschnitt etwa eine Zahl mehr von Schulwoche zu Schulwoche richtig benennen konnten. Kinder mit einem zu Beginn der ersten Klasse festgestellten Risikostatus im mathematischen Vorwissen hatten dagegen einen um $\gamma_{11} = 0.31$ höheren wöchentlichen Lernzuwachs, d. h. sie lösen im Durchschnitt je Schulwoche 1.56 weitere Items richtig. Dieser Unterschied ist statistisch signifikant ($p < .01$). Während Kinder ohne Risikostatus im Schnitt nach vier Schulwochen eine Verbesserung um fünf weitere richtig benannte Zahlen aufweisen, erreichen Kinder mit festgestelltem Risikostatus in dieser Zeit eine Verbesserung um etwa 6 Punkte.

Tabelle 77: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM “Zahlenlesen im Zwanzigerraum”

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Modell für den intercept β_0					
Mittleres Ausgangsniveau γ_{00}	31.80	0.78	40.57	437	< .001
Risikostatus γ_{01}	-15.34	13.86	-11.07	437	< .001
Modell für den slope β_1					
Mittlerer Anstieg γ_{10}	1.25	0.05	25.37	437	< .001
Risikostatus γ_{11}	0.31	0.12	2.65	437	.008

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	12.99	168.69	437	3721.24	< .001
Anstieg u_{1i}	0.62	0.38	437	764.74	< .001
Level-1-Fehler e_{ij}	5.82	33.91			

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

Abschließend wird der Anteil der Varianz, welcher durch das allgemeine Modell im Vergleich zum unkonditionierten Modell erklärt wird, anhand der in Abschnitt 6.4.3 beschriebenen Formel berechnet, die Ergebnisse sind in Tabelle 78 notiert. Die Ergebnisse zeigen, dass die Zugehörigkeit zur Risikogruppe im Bereich des mathematischen Vorwissens 19.7 % der Varianz des Ausgangsniveaus u_{0i} und 5.0 % der Varianz des Anstiegs u_{1i} erklärt

Tabelle 78: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Zahlenlesen im Zwanzigerraum“

Modell	Varianzkomponente des Ausgangsniveaus u_{0i}	Varianzkomponente des Anstiegs u_{1i}
Unkonditioniertes Modell[a]	210.01	0.40
Allgemeines Modell[b]	168.69	0.38
Anteil der erklärten Varianz	19.7 %	5.0 %

[a] aus Tabelle 76 zu entnehmen
[b] aus Tabelle 77 zu entnehmen

7.4.2 CBM „Zahlenreihe im Zwanzigerraum“

Die Interkorrelationen der einzelnen Versionen des CBM „Zahlenreihe im Zwanzigerraum“ über die Messzeitpunkte sind in Tabelle 79 abgebildet. Es ergeben sich durchweg mittlere Korrelationen, welche in einem Intervall von r = .52 und r = .69 streuen. Bei näherer Betrachtung der Tabelle zeigt sich beim CBM „Zahlenreihe im Zwanzigerraum“ ebenso eine deutliche Ähnlichkeit zur Simplexstruktur der Daten.

Tabelle 79: Interkorrelationen der CBM „Zahlenreihe im Zwanzigerraum“ über die MZP 1 – 4 (listenweiser Fallausschluss fehlender Werte, N = 395)

	Korrelation r mit		
	MZP 2 CBM ZR (8.SW)	MZP 3 CBM ZR (12.SW)	MZP 4 CBM ZR (16.SW)
MZP 1 CBM ZR (5.SW)	.65**	.57**	.52**
MZP 2 CBM ZR (8.SW)		.62**	.63**
MZP 3 CBM ZR (12.SW)			.69**

Erläuterungen: MZP – Messzeitpunkt; CBM ZR – CBM „Zahlenreihe im Zwanzigerraum“; SW – Schulwoche; ** – Die Korrelation ist auf dem Niveau von 0.01 2-seitig signifikant.

7.4.2.1 Unkonditioniertes Modell

Analog zur Vorgehensweise beim CBM „Zahlenlesen im Zwanzigerraum“ wird zunächst das unkonditionierte Modell nach den in Kapitel 7.4.1.1 aufgeführten Regressionsgleichungen auf den Ebenen 1 und 2 berechnet. Der Parameter β_{0i} gibt dabei das Ausgangsniveau der Schüler zur ersten Messzeitpunkt in der fünften Schulwoche an. Der Parameter β_{0i} gibt dabei das Ausgangsniveau der Schüler zur ersten Messzeitpunkt in der fünften Schulwoche an. Tabelle 80 fasst die Ergebnisse der unkonditionierten Hierarchisch-linearen Modellierung für das CBM „Zahlenreihe im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres zusammen. Die Schüler weisen ein durchschnittliches Ausgangsniveau von etwa vier richtig gelösten Items zur ersten

Testung in der fünften Schulwoche auf (γ_{00} = 4.35). Der mittlere Anstieg γ_{10} von 0.58 gibt das Maß der durchschnittlichen Verbesserung von Schulwoche zu Schulwoche an. Er zeigt an, dass die Kinder mit jeder weiteren vergangenen Schulwoche im Durchschnitt weitere 0.58 Items richtig lösen konnten, was eine durchschnittliche Verbesserung um einen Punkt nach etwa zwei Schulwochen ergibt. Die beiden geschätzten Regressionsparameter γ_{10} und γ_{00} sind signifikant ($p < .001$) von Null verschieden.

Bei Betrachtung der geschätzten zufälligen Effekte zeigt sich, dass es mit $X^2 = 1338.27$ ($df = 439$, $p < .001$) schülerspezifische Unterschiede zwischen den Regressionskonstanten gibt, welche statistisch signifikant ausfallen. Dies bedeutet, dass die Schüler mit unterschiedlichen Ausgangsniveaus im CBM „Zahlenreihe im Zwanzigerraum“ zur ersten Testung in der fünften Schulwoche starteten. Die schülerspezifischen Unterschiede zwischen den Steigungskoeffizienten sind mit $X^2 = 645.02$ ($df = 439$, $p < .001$) ebenso statistisch signifikant. Dies zeigt an, dass sich die untersuchten Schüler unter der Annahme eines linearen Wachstumsmodells hinsichtlich ihres Lernzuwachses über die Schulwochen signifikant voneinander unterschieden.

Es ergibt sich eine Reliabilität des geschätzten Ausgangsniveaus $\beta_0 = .67$, was darauf hinweist, dass etwa 67 % der individuellen Unterschiede im Ausgangsniveau durch die wahre Varianz zwischen den Schülern erklärt werden können und nicht durch Messfehler bedingt sind. Der Reliabilitätskoeffizient des geschätzten Anstiegs β_1 fällt mit einem Wert von .317 geringer aus. Er deutet daraufhin, dass etwa 31.7 % der individuellen Unterschiede im Leistungsanstieg durch die wahre Varianz zwischen den Schülern erklärt werden können.

Tabelle 80: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM “Zahlenreihe im Zwanzigerraum”

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Mittleres Ausgangsniveau γ_{00}	4.35	0.18	24.72	439	< .001
Mittlerer Anstieg γ_{10}	0.58	0.02	31.41	439	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	3.03	9.16	439	1338.27	< .001
Anstieg u_{1i}	0.22	0.05	439	645.02	< .001
Level-1-Fehler e_{ij}	2.60	6.75			

Reliabilität der Regressionskoeffizienten	
Ausgangsniveau β_0	.670
Anstieg β_1	.317

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

7.4.2.2 Bedingtes intercept-and-slope-as-outcome-Modell

Die Berechnungen des unkonditionierten Modells haben signifikante Unterschiede in den Ausgangsniveaus und den Steigungskoeffizienten über die untersuchten Schüler hinweg ergeben. Ob diese individuellen Unterschiede durch in unterschiedlichem Maße ausgeprägtes mathematisches Vorwissen erklärbar sind, wird analog zum CBM „Zahlenlesen im Zwanzigerraum“ überprüft. Dazu wird die Zugehörigkeit zur Gruppe von Kindern mit einem Risiko im Bereich des mathematischen Vorwissens als Prädiktor auf Ebene 2 im intercept-and-slope-as-outcome-Modell berücksichtigt. Es ergeben sich somit die gleichen Regressionsgleichungen auf Ebene 1 und 2 wie in Kapitel 7.4.1.2 für das CBM „Zahlenlesen im Zwanzigerraum“ beschrieben. Die Ergebnisse der allgemeinen Hierarchisch-linearen Modellierung für das CBM „Zahlenreihe im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres sind in

Tabelle 81 aufgeführt. Das mittlere Ausgangsniveau γ_{00} von 5.05 deutet darauf hin, dass die Schüler, welche nicht der Risikogruppe im mathematischen Vorwissen angehörten, im Durchschnitt knapp fünf der vorgegebenen Items zum ersten Messzeitpunkt in der fünften Schulwoche richtig lösen konnten. Hingegen konnten Schüler, welche der Risikogruppe angehörten, im Vergleich zu den anderen Kindern zum ersten Messzeitpunkt im Schnitt knapp drei Items weniger, d. h. etwa zwei Items, richtig lösen (γ_{01} = -3.05). Dieser Unterschied fällt statistisch signifikant aus ($p < .001$). Der mittlere Anstieg γ_{10} = 0.59 deutet darauf hin, dass Kinder ohne Risikostatus im Durchschnitt etwa ein halbes Item mehr je Schulwoche richtig lösen konnten. Kinder mit einem zu Beginn der ersten Klasse festgestellten Risikostatus im mathematischen Vorwissen hatten dagegen einen um γ_{11} = -0.06 geringeren wöchentlichen Lernzuwachs, d. h. sie lösten je Schulwoche im Durchschnitt weitere 0.54 Items richtig. Dieser Unterschied ist statistisch nicht signifikant ($p > .05$). Alle Kinder, ob ohne oder mit Risikostatus, konnten im Schnitt nach zwei Schulwochen eine Verbesserung um ein weiteres richtig gelöstes Item aufweisen.

Tabelle 81: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM “Zahlenreihe im Zwanzigerraum”

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Modell für den intercept β_0					
Mittleres Ausgangsniveau γ_{00}	5.05	0.20	25.47	438	< .001
Risikostatus γ_{01}	-3.05	0.34	-9.09	438	< .001
Modell für den slope β_1					
Mittlerer Anstieg γ_{10}	0.59	0.02	28.69	438	< .001
Risikostatus γ_{11}	-0.06	0.05	-1.20	438	.232

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	2.74	7.52	438	1174.31	< .001
Anstieg u_{1i}	0.22	0.05	438	642.31	< .001
Level-1-Fehler e_{ij}	2.60	6.76			

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

Abschließend wird der Anteil der Varianz, welcher durch das allgemeine Modell im Vergleich zum unkonditionierten Modell erklärt wird, anhand der in Abschnitt 6.4.3 beschriebenen Formel berechnet, die Ergebnisse sind in Tabelle 82 notiert. Die Ergebnisse zeigen, dass die Zugehörigkeit zur Risikogruppe im Bereich des mathematischen Vorwissens 17.9 % der Varianz des Ausgangsniveaus u_{0i} und 0.0 % der Varianz des Anstiegs u_{1i} erklärt

Tabelle 82: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Zahlenreihe im Zwanzigerraum“

Modell	Varianzkomponente des Ausgangsniveaus u_{0i}	Varianzkomponente des Anstiegs u_{1i}
Unkonditioniertes Modell[a]	9.16	0.05
Allgemeines Modell[b]	7.52	0.05
Anteil der erklärten Varianz	17.9 %	0.0 %

[a] aus Tabelle 80 zu entnehmen
[b] aus Tabelle 81 zu entnehmen

7.4.3 CBM „Mengenvergleich im Zwanzigerraum“

In einem ersten Schritt werden die Interkorrelationen der einzelnen Versionen des CBM „Mengenvergleich im Zwanzigerraum“ analysiert. Die jeweiligen Korrelationskoeffizienten über die Messzeitpunkte sind in Tabelle 83 abgebildet. Es ergeben sich durchweg mittlere bis hohe Korrelationen, welche in einem Intervall von r = .49 und r = .74 streuen. Bei näherer Betrachtung der Tabelle zeigt sich beim CBM „Mengenvergleich im Zwanzigerraum“ ebenso eine Simplexstruktur der Daten.

Tabelle 83: Interkorrelationen der CBM „Mengenvergleich im Zwanzigerraum“ über die MZP 1 – 4 (listenweiser Fallausschluss fehlender Werte, N = 390)

	Korrelation r mit		
	MZP 2 CBM MV (8.SW)	MZP 3 CBM MV (12.SW)	MZP 4 CBM MV (16.SW)
MZP 1 CBM MV (5.SW)	.69**	.56**	.49**
MZP 2 CBM MV (8.SW)		.69**	.67**
MZP 3 CBM MV (12.SW)			.74**

Erläuterungen: MZP – Messzeitpunkt; CBM MV – CBM „Mengenvergleich im Zwanzigerraum“; SW – Schulwoche; ** – Die Korrelation ist auf dem Niveau von 0.01 2-seitig signifikant.

In einem weiteren Schritt zur Untersuchung der Änderungssensibilität des CBM sollen Mehrebenenanalysen Aufschluss über die Leistungsentwicklung der Schüler liefern. Dazu wird zunächst das unkonditionierte Modell für das CBM „Mengenvergleich im Zwanzigerraum“ geschätzt. Anschließend soll anhand eines allgemeinen Modells der Einfluss eines zu Beginn der ersten Klasse erhobenen Risikostatus der Schüler im Bereich Mathematik als eine weitere Prädiktorvariable auf Schülerebene analysiert werden.

7.4.3.1 Unkonditioniertes Modell

Analog zur Vorgehensweise beim CBM „Zahlenlesen im Zwanzigerraum“ wird zunächst das unkonditionierte Modell nach den in Kapitel 7.4.1.1 aufgeführten Regressionsgleichungen auf den Ebenen 1 und 2 berechnet. Der Parameter β_{0i} gibt dabei das Ausgangsniveau der Schüler zur ersten Messzeitpunkt in der fünften Schulwoche an. Tabelle 84 fasst die Ergebnisse der unkonditionierten Hierarchisch-linearen Modellierung für das CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres zusammen. Die Schüler wiesen ein durchschnittliches Ausgangsniveau von etwa 19 richtig gelösten Items zur ersten Testung in der fünften Schulwoche auf (γ_{00} = 19.40). Der mittlere Anstieg γ_{10} von 0.73 gibt das Maß der durchschnittlichen Verbesserung von Schulwoche zu Schulwoche an. Er zeigt an, dass die Kinder mit jeder weiteren vergangenen Schulwoche im Durchschnitt weitere 0.73 Items richtig lösen konnten, was eine durchschnittliche Verbesserung um einen Punkt nach knapp eineinhalb Schulwochen ergibt. Die beiden geschätzten Regressionsparameter γ_{10} und γ_{00} sind signifikant ($p < .001$) von Null verschieden.

Bei Betrachtung der geschätzten zufälligen Effekte zeigt sich, dass es mit $X^2 = 2490.58$ ($df = 439$, $p < .001$) schülerspezifische Unterschiede zwischen den Regressionskonstanten gibt, welche statistisch signifikant ausfallen. Dies bedeutet, dass die Schüler mit unterschiedlichen Ausgangsniveaus im CBM „Mengenvergleich im Zwanzigerraum“ zur ersten Testung in der fünften Schulwoche starteten. Die schülerspezifischen Unterschiede zwischen den Steigungskoeffizienten sind mit $X^2 = 806.15$ ($df = 439$, $p < .001$) ebenso statistisch signifikant. Dies zeigt an, dass sich die untersuchten Schüler unter der Annahme eines linearen Wachstumsmodells hinsichtlich ihres Lernzuwachses über die Schulwochen signifikant voneinander unterschieden.

Es ergibt sich eine Reliabilität des geschätzten Ausgangsniveaus β_0 von .822, was Hinweis darauf gibt, dass etwa 82.2 % der individuellen Unterschiede im Ausgangsniveau durch die wahre Varianz zwischen den Schülern erklärt werden können und nicht durch Messfehler bedingt sind. Der Reliabilitätskoeffizient des geschätzten Anstiegs β_1 fällt mit einem Wert von .453 geringer aus. Er deutet daraufhin, dass etwa 45.3 % der individuellen Unterschiede im Leistungsanstieg durch die wahre Varianz zwischen den Schülern erklärt werden können.

Tabelle 84: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM "Mengenvergleich im Zwanzigerraum"

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Mittleres Ausgangsniveau γ_{00}	19.40	0.40	48.98	439	< .001
Mittlerer Anstieg γ_{10}	0.73	0.03	21.48	439	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	7.55	56.97	439	2490.58	< .001
Anstieg u_{1i}	0.48	0.23	439	806.15	< .001
Level-1-Fehler e_{ij}	4.28	18.31			

Reliabilität der Regressionskoeffizienten	
Ausgangsniveau β_0	.822
Anstieg β_1	.453

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

7.4.3.2 Bedingtes intercept-and-slope-as-outcome-Modell

Die Berechnungen des unkonditionierten Modells haben signifikante Unterschiede in den Ausgangsniveaus und den Steigungskoeffizienten über die untersuchten Schüler hinweg ergeben. Ob diese individuellen Unterschiede durch in

unterschiedlichem Maße ausgeprägtes mathematisches Vorwissen erklärbar sind, wird analog zum CBM „Zahlenlesen im Zwanzigerraum“ überprüft. Dazu wird die Zugehörigkeit zur Gruppe von Kindern mit einem Risiko im Bereich des mathematischen Vorwissens als Prädiktor auf Ebene 2 im intercept-and-slope-as-outcome-Modell berücksichtigt. Es ergeben sich somit die gleichen Regressionsgleichungen auf Ebene 1 und 2 wie in Kapitel 7.4.1.2 für das CBM „Zahlenlesen im Zwanzigerraum“ beschrieben. In Tabelle 85 sind die Ergebnisse der allgemeinen Hierarchisch-linearen Modellierung für das CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres aufgeführt unter Berücksichtigung des Risikostatus im mathematischen Vorwissen. Das mittlere Ausgangsniveau γ_{00} von 21.31 deutet darauf hin, dass die Schüler, welche nicht der Risikogruppe im mathematischen Vorwissen angehörten, im Durchschnitt knapp 21 der vorgegebenen Items zum ersten Messzeitpunkt in der fünften Schulwoche richtig lösen konnten. Hingegen konnten Schüler, welche der Risikogruppe angehörten, im Vergleich zu den anderen Kindern zum ersten Messzeitpunkt im Schnitt knapp acht Items weniger, d. h. etwa 13 Items, richtig lösen (γ_{01} = -8.29). Dieser Unterschied fällt statistisch signifikant aus ($p < .001$). Der mittlere Anstieg γ_{10} von 0.67 deutet darauf hin, dass Kinder ohne Risikostatus im Durchschnitt etwa ein halbes Item mehr von Schulwoche zu Schulwoche richtig lösen konnten. Kinder mit einem zu Beginn der ersten Klasse festgestellten Risikostatus im mathematischen Vorwissen hatten dagegen einen um γ_{11} = -0.29 geringeren wöchentlichen Lernzuwachs, d. h. sie lösten im Durchschnitt weitere 0.38 Items richtig je Schulwoche. Dieser Unterschied ist statistisch signifikant ($p < .01$). Während Kinder ohne Risikostatus im Schnitt nach eineinhalb Schulwochen eine Verbesserung um ein weiteres richtig gelöstes Item aufweisen, benötigen Kinder mit festgestelltem Risikostatus für eine derartige Verbesserung etwa zwei Schulwochen.

Tabelle 85: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM “Mengenvergleich im Zwanzigerraum”

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Modell für den intercept β_0					
Mittleres Ausgangsniveau γ_{00}	21.31	0.40	53.54	438	< .001
Risikostatus γ_{01}	-8.29	0.91	-9.13	438	< .001
Modell für den slope β_1					
Mittlerer Anstieg γ_{10}	0.67	0.24	18.62	438	< .001
Risikostatus γ_{11}	-0.29	0.09	3.15	438	.002

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	6.70	44.85	438	2049.01	< .001
Anstieg u_{1i}	0.47	0.22	438	783.77	< .001
Level-1-Fehler e_{ij}	4.28	18.31			

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

Abschließend wird der Anteil der Varianz, welcher durch das allgemeine Modell im Vergleich zum unkonditionierten Modell erklärt wird, anhand der in Abschnitt 6.4.3 beschriebenen Formel berechnet, die Ergebnisse sind in Tabelle 86 notiert. Die Ergebnisse zeigen, dass die Zugehörigkeit zur Risikogruppe im Bereich des mathematischen Vorwissens 21.3 % der Varianz des Ausgangsniveaus u_{0i} und 4.3 % der Varianz des Anstiegs u_{1i} erklärt

Tabelle 86: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Mengenvergleich im Zwanzigerraum“

Modell	Varianzkomponente des Ausgangsniveaus u_{0i}	Varianzkomponente des Anstiegs u_{1i}
Unkonditioniertes Modell[a]	56.97	0.23
Allgemeines Modell[b]	44.85	0.22
Anteil der erklärten Varianz	21.3 %	4.3 %

[a] aus Tabelle 84 zu entnehmen
[b] aus Tabelle 85 zu entnehmen

7.4.4 CBM „Arithmetische Basiskompetenzen“

In einem ersten Schritt werden die Interkorrelationen der einzelnen Versionen des CBM „Arithmetische Basiskompetenzen“ analysiert. Die jeweiligen Korrelationskoeffizienten über die Messzeitpunkte sind in Tabelle 87 abgebildet. Es ergeben sich geringe bis hohe Korrelationen, welche in einem Intervall von r = .42 und r = .75 streuen. Bei näherer Betrachtung der Tabelle zeigt sich beim CBM „Arithmetische Basiskompetenzen“ ebenso eine Ähnlichkeit zur Simplexstruktur der Daten.

Tabelle 87: Interkorrelationen der CBM „Arithmetische Basiskompetenzen“ über die MZP 1 – 9 (listenweiser Fallausschluss fehlender Werte, N = 307)

	Korrelation r mit							
	MZP 2 CBM AB (8.SW)	MZP 3 CBM AB (12.SW)	MZP 4 CBM AB (16.SW)	MZP 5 CBM AB (21.SW)	MZP 6 CBM AB (26.SW)	MZP 7 CBM AB (30.SW)	MZP 8 CBM AB (34.SW)	MZP 9 CBM AB (38.SW)
MZP 1 CBM AB (5.SW)	.65**	.54**	.59**	.48**	.46**	.56**	.53**	.42**
MZP 2 CBM AB (8.SW)		.65**	.62**	.51**	.52**	.59**	.56**	.48**
MZP 3 CBM AB (12.SW)			.65**	.60**	.59**	.62**	.61**	.50**
MZP 4 CBM AB (16.SW)				.60**	.64**	.70**	.65**	.57**
MZP 5 CBM AB (21.SW)					.65**	.68**	.67**	.62**
MZP 6 CBM AB (26.SW)						.71**	.65**	.63**
MZP 7 CBM AB (30.SW)							.75**	.64**
MZP 8 CBM AB (34.SW)								.69**

Erläuterungen: MZP – Messzeitpunkt; CBM AB – CBM „Arithmetische Basiskompetenzen“; SW – Schulwoche; ** – Die Korrelation ist auf dem Niveau von 0.01 2-seitig signifikant.

7.4.4.1 Unkonditioniertes Modell

Analog zur Vorgehensweise beim CBM „Zahlenlesen im Zwanzigerraum“ wird zunächst das unkonditionierte Modell nach den in Kapitel 7.4.1.1 aufgeführten Regressionsgleichungen auf den Ebenen 1 und 2 berechnet. Der Parameter β_{0i} gibt dabei das Ausgangsniveau der Schüler zur ersten Messzeitpunkt in der fünften Schulwoche an. Tabelle 88 fasst die Ergebnisse der unkonditionierten Hierarchisch-linearen Modellierung für das CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr zusammen. Die Schüler weisen ein durchschnittliches Ausgangsniveau von fast fünf richtig gelösten Items zur ersten Testung in der fünften Schulwoche auf ($\gamma_{00} = 4.70$). Der mittlere Anstieg γ_{10} von 0.15 gibt das Maß der durchschnittlichen Verbesserung von Schulwoche zu Schulwoche an. Er zeigt an, dass die Kinder mit jeder weiteren vergangenen Schulwoche im Durchschnitt weitere 0.15 Items richtig lösen können, was eine durchschnittliche Verbesserung um einen Punkt nach etwa sieben Schulwochen ergibt. Die beiden geschätzten Regressionsparameter γ_{10} und γ_{00} sind signifikant ($p < .001$) von Null verschieden.

Bei Betrachtung der geschätzten zufälligen Effekte zeigt sich, dass es mit $X^2 = 2108.58$ ($df = 440$, $p < .001$) schülerspezifische Unterschiede zwischen den Regressionskonstanten gibt, welche statistisch signifikant ausfallen. Dies bedeutet, dass die Schüler mit unterschiedlichen Ausgangsniveaus im CBM „Arithmetische Basiskompetenzen“ zur ersten Testung in der fünften Schulwoche starten. Die schülerspezifischen Unterschiede zwischen den Steigungskoeffizienten sind mit $X^2 = 781.52$ ($df = 440$, $p < .001$) ebenso statistisch signifikant. Dies zeigt an, dass sich die untersuchten Schüler unter der Annahme eines linearen Wachstumsmodells hinsichtlich ihres Lernzuwachses über die Schulwochen signifikant voneinander unterscheiden.

Es ergibt sich eine Reliabilität des geschätzten Ausgangsniveaus β_0 von .788, was Hinweis darauf gibt, dass etwa 78.8 % der individuellen Unterschiede im Ausgangsniveau durch die wahre Varianz zwischen den Schülern erklärt werden können und nicht durch Messfehler bedingt sind. Der Reliabilitätskoeffizient des geschätzten Anstiegs β_1 fällt mit einem Wert von .433 geringer aus. Er deutet daraufhin, dass etwa 43.3 % der individuellen Unterschiede im Leistungsanstieg durch die wahre Varianz zwischen den Schülern erklärt werden können.

Tabelle 88: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM "Arithmetische Basiskompetenzen"

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Mittleres Ausgangsniveau γ_{00}	4.70	0.13	36.07	440	< .001
Mittlerer Anstieg γ_{10}	0.15	0.004	35.24	440	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	2.44	5.93	440	2108.58	< .001
Anstieg u_{1i}	0.058	0.00	440	781.52	< .001
Level-1-Fehler e_{ij}	2.08	4.34			

Reliabilität der Regressionskoeffizienten	
Ausgangsniveau β_0	0.788
Anstieg β_1	0.433

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

7.4.4.2 Bedingtes intercept-and-slope-as-outcome-Modell

Die Berechnungen des unkonditionierten Modells haben signifikante Unterschiede in den Ausgangsniveaus und den Steigungskoeffizienten über die untersuchten Schüler hinweg ergeben. Ob diese individuellen Unterschiede durch in unterschiedlichem Maße ausgeprägtes mathematisches Vorwissen

erklärbar sind, wird analog zum CBM „Zahlenlesen im Zwanzigerraum“ überprüft. Dazu wird die Zugehörigkeit zur Gruppe von Kindern mit einem Risiko im Bereich des mathematischen Vorwissens als Prädiktor auf Ebene 2 im intercept-and-slope-as-outcome-Modell berücksichtigt. Es ergeben sich somit die gleichen Regressionsgleichungen auf Ebene 1 und 2 wie in Kapitel 7.4.1.2 für das CBM „Zahlenlesen im Zwanzigerraum“ beschrieben. In Tabelle 89 sind die Ergebnisse der allgemeinen Hierarchisch-linearen Modellierung für das CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr aufgeführt. Das mittlere Ausgangsniveau γ_{00} von 5.12 deutet darauf hin, dass die Schüler, welche nicht der Risikogruppe im mathematischen Vorwissen angehören, im Durchschnitt knapp fünf Items zum ersten Messzeitpunkt in der fünften Schulwoche richtig gelöst haben. Hingegen konnten Schüler, welche der Risikogruppe angehören, im Vergleich zu den anderen Kindern zum ersten Messzeitpunkt im Schnitt etwa zwei Items weniger, d. h. etwa drei Items, lösen (γ_{01} = -1.80). Dieser Unterschied fällt statistisch signifikant aus ($p < .001$). Der mittlere Anstieg γ_{10} von 0.15 deutet darauf hin, dass Kinder ohne Risikostatus im Durchschnitt 0.15 Items mehr je Schulwoche richtig lösen können. Kinder mit einem zu Beginn der ersten Klasse festgestellten Risikostatus im mathematischen Vorwissen haben dagegen einen um γ_{11} = -0.01 geringeren wöchentlichen Lernzuwachs, d. h. sie lösen im Durchschnitt weitere 0.14 Items richtig je Schulwoche. Dieser Unterschied ist jedoch nicht statistisch signifikant ($p > .05$).

Tabelle 89: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM “Arithmetische Basiskompetenzen”

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Modell für den intercept β_0					
Mittleres Ausgangsniveau γ_{00}	5.12	0.15	34.97	439	< .001
Risikostatus γ_{01}	-1.80	0.28	-6.45	439	< .001
Modell für den slope β_1					
Mittlerer Anstieg γ_{10}	0.15	0.01	31.77	439	< .001
Risikostatus γ_{11}	-0.01	0.01	-1.24	439	.216

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	2.32	5.37	439	1949.57	< .001
Anstieg u_{1i}	0.06	0.00	439	778.05	< .001
Level-1-Fehler e_{ij}	2.08	4.35			

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

Abschließend wird der Anteil der Varianz, welcher durch das allgemeine Modell im Vergleich zum unkonditionierten Modell erklärt wird, anhand der in Abschnitt 6.4.3 beschriebenen Formel berechnet, die Ergebnisse sind in Tabelle 90 notiert. Die Ergebnisse zeigen, dass die Zugehörigkeit zur Risikogruppe im Bereich des mathematischen Vorwissens 9.4 % der Varianz des Ausgangsniveaus u_{0i} und 0 % der Varianz des Anstiegs u_{1i} erklärt

Tabelle 90: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Arithmetische Basiskompetenzen“

Modell	Varianzkomponente des Ausgangsniveaus u_{0i}	Varianzkomponente des Anstiegs u_{1i}
Unkonditioniertes Modell[a]	5.93	0.00
Allgemeines Modell[b]	5.37	0.00
Anteil der erklärten Varianz	9.4 %	0 %

[a] aus Tabelle 88 zu entnehmen
[b] aus Tabelle 89 zu entnehmen

7.4.5 CBM „Zahlzerlegung im Zwanzigerraum“

In einem ersten Schritt werden die Interkorrelationen der einzelnen Versionen des CBM „Zahlzerlegung im Zwanzigerraum“ analysiert. Die jeweiligen Korrelationskoeffizienten über die Messzeitpunkte sind in Tabelle 91 abgebildet. Es ergeben sich durchweg mittlere bis hohe Korrelationen, welche in einem Intervall von r = .60 und r = .75 streuen. Für das CBM „Zahlzerlegung im Zwanzigerraum“ ergibt sich keine Simplexstruktur der Daten. Im Gegenteil steigen die Korrelationen jeweils zum achten und zum neunten Messzeitpunkt an.

Tabelle 91: Interkorrelationen der CBM „Zahlzerlegung im Zwanzigerraum“ über die MZP 5 – 9 (listenweiser Fallausschluss fehlender Werte, N = 335)

	Korrelation r mit			
	MZP 6 CBM ZZ (26.SW)	MZP 7 CBM ZZ (30.SW)	MZP 8 CBM ZZ (34.SW)	MZP 9 CBM ZZ (38.SW)
MZP 5 CBM ZZ (21.SW)	.61**	.60**	.62**	.64**
MZP 6 CBM ZZ (26.SW)		.63**	.65**	.68**
MZP 7 CBM ZZ (30.SW)			.70**	.75**
MZP 8 CBM ZZ (34.SW)				.72**

Erläuterungen: MZP – Messzeitpunkt; CBM ZZ – CBM „Zahlzerlegung im Zwanzigerraum“; SW – Schulwoche; ** – Die Korrelation ist auf dem Niveau von 0.01 2-seitig signifikant.

7.4.5.1 Unkonditioniertes Modell

Analog zur Vorgehensweise beim CBM „Zahlenlesen im Zwanzigerraum“ wird zunächst das unkonditionierte Modell nach den in Kapitel 7.4.1.1 aufgeführten Regressionsgleichungen auf den Ebenen 1 und 2 berechnet. Der Parameter β_{0i} gibt dabei das Ausgangsniveau der Schüler zur ersten Testung in der 21. Schulwoche an. In Tabelle 92 sind die Ergebnisse des unkonditionierten Hierarchisch-linearen Modelles für das CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in

der zweiten Hälfte des ersten Schuljahres abgebildet. Das mittlere Ausgangsniveau γ_{00} von 3.82 deutet darauf hin, dass die Schüler im Durchschnitt knapp vier Items zum fünften Messzeitpunkt in der 21. Schulwoche richtig gelöst haben. Der mittlere Anstieg γ_{00} von 0.21 zeigt an, dass die Kinder mit jeder weiteren vergangenen Schulwoche im Durchschnitt weitere 0.21 Items richtig lösen können, was eine durchschnittliche Verbesserung um einen Punkt nach fünf Schulwochen ergibt.

Es zeigen sich signifikante Unterschiede in den individuellen Abweichungen vom mittleren Ausgangsniveau u_{0i} zum fünften Messzeitpunkt in der 21. Schulwoche: $X^2 = 1058.34$ ($df = 431$, $p < .001$). Hingegen zeigen sich keine signifikanten Unterschiede in den individuellen Abweichungen der linearen Leistungsanstiege u_{1i} der Kinder über die Schulwochen: $X^2 = 454.74$ ($df = 431$, $p > .05$). Die geschätzte Standardabweichung für die individuellen Abweichungen der linearen Leistungsanstiege u_{1i} beträgt 0.07, d. h., dass ein Kind, welches einen Leistungsanstieg aufweist, der einer Standardabweichung über dem des Durchschnitts entspricht, zeigt einen Leistungsanstieg von Schulwoche zu Schulwoche von 0.21 + 0.07 = 0.28.

Es ergibt sich eine Reliabilität des geschätzten Ausgangsniveaus β_0 von .597, was Hinweis darauf gibt, dass etwa 59.7 % der individuellen Unterschiede im Ausgangsniveau durch die wahre Varianz zwischen den Schülern erklärt werden können und nicht durch Messfehler bedingt sind. Der Reliabilitätskoeffizient des geschätzten Anstiegs β_1 fällt mit einem Wert von .132 geringer aus. Er deutet daraufhin, dass etwa 13.2 % der individuellen Unterschiede im Leistungsanstieg durch die wahre Varianz zwischen den Schülern erklärt werden können.

Tabelle 92: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM „Zahlzerlegung im Zwanzigerraum"

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Mittleres Ausgangsniveau γ_{00}	3.82	0.13	29.20	433	< .001
Mittlerer Anstieg γ_{10}	0.21	0.01	25.99	433	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	2.16	4.65	431	1058.34	< .001
Anstieg u_{1i}	0.07	0.00	431	454.74	.207
Level-1-Fehler e_{ij}	2.14	4.57			

Reliabilität der Regressionskoeffizienten	
Ausgangsniveau β_0	.597
Anstieg β_1	.132

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

7.4.5.2 Bedingtes intercept-as-outcome-Modell

Die Berechnungen des unkonditionierten Modells haben keine signifikanten Unterschiede in den Steigungskoeffizienten über die untersuchten Schüler hinweg ergeben. Demnach wird nachfolgend lediglich der Einfluss schülerspezifischer Merkmale auf die Regressionskonstante β_0 überprüft, deren Einfluss auf den Steigungskoeffizienten β_1 wird als fest angenommen. Um zu prüfen, ob die Unterschiede in dem Ausgangsniveau β_0 durch in unterschiedlichem Maße ausgeprägtes mathematisches Vorwissen erklärbar sind, wird die Zugehörigkeit zur Gruppe von Kindern mit einem Risiko im Bereich des mathematischen Vorwissens im intercept-as-outcome-Modell berücksichtigt. Es ergeben sich somit folgende Modelle auf Ebene 1 und 2:

Modell der Ebene 1: $Y_{i} = \beta_{0i} + \beta_{1i} \bullet X_{ij} + e_{ij}$

Modelle der Ebene 2: $\beta_{0i} = \gamma_{00} + \gamma_{01} \bullet Z_i + u_{0i}$

$\beta_{1i} = \gamma_{10} + u_{1i}$

In Tabelle 93 sind die Ergebnisse der allgemeinen Hierarchisch-linearen Modellierung für das CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres aufgeführt. Das mittlere Ausgangsniveau γ_{00} von 4.33 deutet darauf hin, dass die Schüler, welche nicht der Risikogruppe im mathematischen Vorwissen angehörten, im Durchschnitt knapp vier Items des CBM „Zahlzerlegung im Zwanzigerraum“ zum ersten Testzeitpunkt in der 21. Schulwoche richtig gelöst haben. Hingegen konnten Schüler, welche der Risikogruppe angehörten, im Vergleich zu den anderen Kindern zu diesem Zeitpunkt im Schnitt zwei Items weniger, d. h. etwa zwei Items, lösen (γ_{01} = -2.28). Dieser Unterschied fällt statistisch signifikant aus ($p < .001$). Der mittlere Anstieg γ_{10} von Kindern mit und ohne Risikostatus wurde als fester Effekt in das Modell gegeben und entspricht somit dem Wert des unkonditionierten Modells γ_{10} = 0.21. Der Wert weist darauf hin, dass alle Kinder im Durchschnitt 0.21 Items mehr von Schulwoche zu Schulwoche richtig lösen können, dies bedeutet eine Verbesserung um ein richtig gelöstes Item nach etwa fünf Schulwochen.

Tabelle 93: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM „Zahlzerlegung im Zwanzigerraum“

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Modell für den intercept β_0					
Mittleres Ausgangsniveau γ_{00}	4.33	0.15	29.91	432	< .001
Risikostatus γ_{01}	-2.28	0.27	-8.45	432	< .001
Modell für den slope β_1					
Mittlerer Anstieg γ_{10}	0.21	0.01	26.015	433	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	1.97	3.89	430	955.24	< .001
Anstieg u_{1i}	0.06	0.00	431	454.16	.213
Level-1-Fehler e_{ij}	2.14	4.58			

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

Abschließend wird der Anteil der Varianz, welcher durch das allgemeine Modell im Vergleich zum unkonditionierten Modell erklärt wird, anhand der in Abschnitt 6.4.3 beschriebenen Formel berechnet, die Ergebnisse sind in Tabelle 94 notiert. Die Ergebnisse zeigen, dass die Zugehörigkeit zur Risikogruppe im Bereich des mathematischen Vorwissens 16.3 % der Varianz des Ausgangsniveaus u_{0i} und 0 % der Varianz des Anstiegs u_{1i} erklärt.

Tabelle 94: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Zahlzerlegung im Zwanzigerraum“

Modell	Varianzkomponente des Ausgangsniveaus u_{0i}	Varianzkomponente des Anstiegs u_{1i}
Unkonditioniertes Modell[a]	4.65	0.00
Allgemeines Modell[b]	3.89	0.00
Anteil der erklärten Varianz	16.3 %	0 %

[a] aus Tabelle 92 zu entnehmen
[b] aus Tabelle 93 zu entnehmen

7.4.6 CBM „Addition im Zwanzigerraum“

In einem ersten Schritt werden die Interkorrelationen der einzelnen Versionen des CBM „Addition im Zwanzigerraum“ analysiert. Die jeweiligen Korrelationskoeffizienten über die Messzeitpunkte sind in Tabelle 95 abgebildet. Es ergeben sich durchweg hohe Korrelationen, welche in einem Intervall von r = .63 und r = .80 streuen. Bei näherer Betrachtung der Tabelle zeigt sich beim CBM „Addition im Zwanzigerraum“ ebenso eine deutliche Ähnlichkeit zur Simplexstruktur der Daten.

Tabelle 95: Interkorrelationen der CBM „Addition im Zwanzigerraum“ über die MZP 5 – 9 (listenweiser Fallausschluss fehlender Werte, N = 331)

	Korrelation r mit			
	MZP 6 CBM AD (26.SW)	MZP 7 CBM AD (30.SW)	MZP 8 CBM AD (34.SW)	MZP 9 CBM AD (38.SW)
MZP 5 CBM AD (21.SW)	.73**	.63**	.67**	.65**
MZP 6 CBM AD (26.SW)		.72**	.75**	.74**
MZP 7 CBM AD (30.SW)			.71**	.69**
MZP 8 CBM AD (34.SW)				.80**

Erläuterungen: MZP – Messzeitpunkt; CBM AD – CBM „Addition im Zwanzigerraum“; SW – Schulwoche; ** – Die Korrelation ist auf dem Niveau von 0.01 2-seitig signifikant.

7.4.6.1 Unkonditioniertes Modell

Analog zur Vorgehensweise beim CBM „Zahlenlesen im Zwanzigerraum“ wird zunächst das unkonditionierte Modell nach den in Kapitel 7.4.1.1 aufgeführten Regressionsgleichungen auf den Ebenen 1 und 2 berechnet. Der Parameter β_{0i} gibt dabei das Ausgangsniveau der Schüler zur ersten Testung in der 21. Schulwoche an. Die Ergebnisse des unkonditionierten Modells sind in Tabelle 96 zusammengefasst. Die Schüler weisen ein durchschnittliches Ausgangsniveau von fast sechs richtig gelösten Items zur ersten Testung in der 21. Schulwoche auf (γ_{00} = 5.73). Der

mittlere Anstieg γ_{10} von 0.26 gibt das Maß der durchschnittlichen Verbesserung von Schulwoche zu Schulwoche an. Er zeigt an, dass die Kinder mit jeder weiteren vergangenen Schulwoche im Durchschnitt weitere 0.26 Items richtig lösen können, was eine durchschnittliche Verbesserung um einen Punkt nach etwa vier Schulwochen ergibt. Die beiden geschätzten Regressionsparameter γ_{10} und γ_{00} sind signifikant ($p < .001$) von Null verschieden.

Bei Betrachtung der geschätzten zufälligen Effekte zeigt sich, dass es schülerspezifische Unterschiede zwischen den Regressionskonstanten gibt, welche statistisch signifikant ausfallen $X^2 = 1281.31$ ($df = 431$, $p < .001$). Dies bedeutet, dass die Schüler mit unterschiedlichen Ausgangsniveaus im CBM „Addition im Zwanzigerraum“ zur ersten Testung in der 21. Schulwoche starten. Die schülerspezifischen Unterschiede zwischen den Steigungskoeffizienten sind mit $X^2 = 447.34$ ($df = 431$, $p > .05$) hingegen nicht statistisch signifikant. Dies zeigt an, dass sich die untersuchten Schüler unter der Annahme eines linearen Wachstumsmodells hinsichtlich ihres Lernzuwachses über die Schulwochen nicht signifikant voneinander unterscheiden.

Es ergibt sich eine Reliabilität des geschätzten Ausgangsniveaus β_0 von .673, was Hinweis darauf gibt, dass etwa 67.3 % der individuellen Unterschiede im Ausgangsniveau durch die wahre Varianz zwischen den Schülern erklärt werden können und nicht durch Messfehler bedingt sind. Der Reliabilitätskoeffizient des geschätzten Anstiegs β_1 fällt mit einem Wert von .091 geringer aus. Er deutet daraufhin, dass etwa 9.1 % der individuellen Unterschiede im Leistungsanstieg durch die wahre Varianz zwischen den Schülern erklärt werden können.

Tabelle 96: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM “Addition im Zwanzigerraum”

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Mittleres Ausgangsniveau γ_{00}	5.73	0.14	41.449	433	< .001
Mittlerer Anstieg γ_{10}	0.29	0.01	33.619	433	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	2.39	5.70	431	1281.31	< .001
Anstieg u_{1i}	0.05	0.00	431	447.34	.283
Level-1-Fehler e_{ij}	1.99	3.98			

Reliabilität der Regressionskoeffizienten	
Ausgangsniveau β_0	.673
Anstieg β_1	.091

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

7.4.6.2 Bedingtes intercept-as-outcome-Modelll

Wie beim CBM „Zahlzerlegung im Zwanzigerraum“ haben die Berechnungen des unkonditionierten Modells signifikante Unterschiede in den Ausgangsniveaus, jedoch nicht in den Steigungskoeffizienten über die untersuchten Schüler hinweg ergeben. Ob diese individuellen Unterschiede durch in unterschiedlichem Maße ausgeprägtes mathematisches Vorwissen erklärbar sind, wird analog zum CBM „Zahlzerlegung im Zwanzigerraum“ überprüft. Dazu wird die Zugehörigkeit zur Gruppe von Kindern mit einem Risiko im Bereich des mathematischen Vorwissens als Prädiktor auf Ebene 2 im intercept-as-outcome-Modell berücksichtigt. Es ergeben sich somit die gleichen Regressionsgleichungen auf Ebene 1 und 2 wie in Kapitel 7.4.5.2 für das CBM „Zahlzerlegung im Zwanzigerraum“ beschrieben. In Tabelle 97 sind die Ergebnisse der allgemeinen Hierarchisch-linearen Modellierung für das CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres

aufgeführt. Das mittlere Ausgangsniveau γ_{00} von 6.38 deutet darauf hin, dass die Schüler, welche nicht der Risikogruppe im mathematischen Vorwissen angehörten, im Durchschnitt knapp sechs Items des CBM „Addition im Zwanzigerraum“ zum ersten Testzeitpunkt in der 21. Schulwoche richtig gelöst haben. Hingegen konnten Schüler, welche der Risikogruppe angehörten, im Vergleich zu den anderen Kindern zu diesem Zeitpunkt im Schnitt etwa drei Items, d. h. etwa drei Items, weniger lösen (γ_{01} = -2.87). Dieser Unterschied fällt statistisch signifikant aus (p < .001). Der mittlere Anstieg γ_{10} von Kindern mit und ohne Risikostatus wurde als fester Effekt in das Modell gegeben und entspricht somit dem Wert des unkonditionierten Modells γ_{10} = 0.26. Der Wert weist darauf hin, dass alle Kinder im Durchschnitt 0.26 Items mehr von Schulwoche zu Schulwoche richtig lösen können, dies bedeutet eine Verbesserung um ein richtig gelöstes Item nach etwa vier Schulwochen.

Tabelle 97: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM „Addition im Zwanzigerraum“

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Modell für den intercept β_0					
Mittleres Ausgangsniveau γ_{00}	6.38	0.15	43.79	432	< .001
Risikostatus γ_{01}	-2.87	0.30	-9.59	432	< .001
Modell für den slope β_1					
Mittlerer Anstieg γ_{10}	0.26	0.01	33.60	433	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	2.11	4.46	430	1097.65	< .001
Anstieg u_{1i}	0.05	0.00	431	446.28	.295
Level-1-Fehler e_{ij}	2.00	3.99			

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

Abschließend wird der Anteil der Varianz, welcher durch das allgemeine Modell im Vergleich zum unkonditionierten Modell erklärt

wird, anhand der in Abschnitt 6.4.3 beschriebenen Formel berechnet, die Ergebnisse sind in Tabelle 98 notiert. Die Ergebnisse zeigen, dass die Zugehörigkeit zur Risikogruppe im Bereich des mathematischen Vorwissens 21.8 % der Varianz des Ausgangsniveaus u_{0i} und 0 % der Varianz des Anstiegs u_{1i} erklärt.

Tabelle 98: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Addition im Zwanzigerraum“

Modell	Varianzkomponente des Ausgangsniveaus u_{0i}	Varianzkomponente des Anstiegs u_{1i}
Unkonditioniertes Modell[a]	5.70	0.00
Allgemeines Modell[b]	4.46	0.00
Anteil der erklärten Varianz	21.8 %	0 %

[a] aus Tabelle 96 zu entnehmen
[b] aus Tabelle 97 zu entnehmen

7.4.7 CBM „Subtraktion im Zwanzigerraum“

In einem ersten Schritt werden die Interkorrelationen der einzelnen Versionen des CBM „Subtraktion im Zwanzigerraum“ analysiert. Die jeweiligen Korrelationskoeffizienten über die Messzeitpunkte sind in Tabelle 99 abgebildet. Es ergeben sich durchweg mittlere bis hohe Korrelationen, welche in einem Intervall von r = .57 und r = .80 streuen. Bei näherer Betrachtung der Tabelle zeigt sich beim CBM „Subtraktion im Zwanzigerraum“ ebenso eine deutliche Ähnlichkeit zur Simplexstruktur der Daten.

Tabelle 99: Interkorrelationen der CBM „Subtraktion im Zwanzigerraum“ über die MZP 5 – 9 (listenweiser Fallausschluss fehlender Werte, N = 333)

	Korrelation r mit			
	MZP 6 CBM SUB (26.SW)	MZP 7 CBM SUB (30.SW)	MZP 8 CBM SUB (34.SW)	MZP 9 CBM SUB (38.SW)
MZP 5 CBM SUB (21.SW)	.71**	.68**	.64**	.57**
MZP 6 CBM SUB (26.SW)		.73**	.70**	.65**
MZP 7 CBM SUB (30.SW)			.76**	.68**
MZP 8 CBM SUB (34.SW)				.80**

Erläuterungen: MZP – Messzeitpunkt; CBM SUB – CBM „Subtraktion im Zwanzigerraum“; SW – Schulwoche; ** – Die Korrelation ist auf dem Niveau von 0.01 2-seitig signifikant.

7.4.7.1 Unkonditioniertes Modell

Analog zur Vorgehensweise beim CBM „Zahlenlesen im Zwanzigerraum“ wird zunächst das unkonditionierte Modell nach den in Kapitel 7.4.1.1 aufgeführten Regressionsgleichungen auf den Ebenen 1 und 2 berechnet. Der Parameter β_{0i} gibt dabei das Ausgangsniveau der Schüler zur ersten Testung in der 21. Schulwoche an. Tabelle 100 fasst die Ergebnisse der unkonditionierten Hierarchisch-linearen Modellierung für das CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres zusammen. Die Schüler weisen ein durchschnittliches Ausgangsniveau von fast fünf richtig gelösten Items zur ersten Testung in der 21. Schulwoche auf (γ_{00} = 3.60). Der mittlere Anstieg γ_{10} von 0.22 gibt das Maß der durchschnittlichen Verbesserung von Schulwoche zu Schulwoche an. Er zeigt an, dass die Kinder mit jeder weiteren vergangenen Schulwoche im Durchschnitt 0.22 Items mehr richtig lösen können, was eine durchschnittliche Verbesserung um einen Punkt nach etwa fünf Schulwochen ergibt. Die beiden geschätzten Regressionsparameter γ_{10} und γ_{00} sind signifikant ($p < .001$) von Null verschieden.

Bei Betrachtung der geschätzten zufälligen Effekte zeigt sich, dass es schülerspezifische Unterschiede zwischen den Regressionskonstanten gibt, welche statistisch signifikant ausfallen $X^2 = 1431.68$ ($df = 431$, $p < .001$). Dies bedeutet, dass die Schüler mit unterschiedlichen Ausgangsniveaus im CBM „Subtraktion im Zwanzigerraum“ zur ersten Testung in der 21. Schulwoche starten. Die schülerspezifischen Unterschiede zwischen den Steigungskoeffizienten sind mit $X^2 = 674.80$ ($df = 431$, $p < .001$) ebenso statistisch signifikant. Dies zeigt an, dass sich die untersuchten Schüler unter der Annahme eines linearen Wachstumsmodells hinsichtlich ihres Lernzuwachses über die Schulwochen signifikant voneinander unterscheiden.

Es ergibt sich eine Reliabilität des geschätzten Ausgangsniveaus β_0 von .669, was Hinweis darauf gibt, dass etwa 66.9 % der individuellen Unterschiede im Ausgangsniveau durch die wahre Varianz zwischen den Schülern erklärt werden können und nicht durch Messfehler bedingt sind. Der Reliabilitätskoeffizient des geschätzten Anstiegs β_1 fällt mit einem Wert von .354 geringer aus. Er deutet daraufhin, dass etwa 35.4 % der individuellen Unterschiede im Leistungsanstieg durch die wahre Varianz zwischen den Schülern erklärt werden können.

Tabelle 100: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM „Subtraktion im Zwanzigerraum“

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Mittleres Ausgangsniveau γ_{00}	3.60	0.13	27.27	433	< .001
Mittlerer Anstieg γ_{10}	0.22	0.01	25.83	433	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	2.31	5.32	431	1431.68	< .001
Anstieg u_{1i}	0.11	0.01	431	674.80	< .001
Level-1-Fehler e_{ij}	1.83	3.35			

Reliabilität der Regressionskoeffizienten	
Ausgangsniveau β_0	.669
Anstieg β_1	.354

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

7.4.7.2 Bedingtes intercept-and-slope-as-outcome-Modell

Die Berechnungen des unkonditionierten Modells haben signifikante Unterschiede in den Ausgangsniveaus und den Steigungskoeffizienten über die untersuchten Schüler hinweg ergeben. Ob diese individuellen Unterschiede durch in unterschiedlichem Maße ausgeprägtes mathematisches Vorwissen erklärbar sind, wird analog zum CBM „Zahlenlesen im Zwanzigerraum“ überprüft. Dazu wird die Zugehörigkeit zur Gruppe von Kindern mit einem Risiko im Bereich des mathematischen Vorwissens als Prädiktor auf Ebene 2 im intercept-and-slope-as-outcome-Modell berücksichtigt. Es ergeben sich somit die gleichen Regressionsgleichungen auf Ebene 1 und 2 wie in Kapitel 7.4.1.2 für das CBM „Zahlenlesen im Zwanzigerraum“ beschrieben. In Tabelle 101 sind die Ergebnisse der allgemeinen Hierarchisch-linearen Modellierung für das CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres

aufgeführt. Das mittlere Ausgangsniveau γ_{00} von 3.98 deutet darauf hin, dass die Schüler, welche nicht der Risikogruppe im mathematischen Vorwissen angehören im Durchschnitt knapp vier Items zum fünften Messzeitpunkt in der 21. Schulwoche richtig gelöst haben. Hingegen konnten Schüler, welche der Risikogruppe angehören im Vergleich zu den anderen Kindern zum fünften Messzeitpunkt im Schnitt etwa zwei Items weniger, d. h. etwa zwei Items, lösen (γ_{01} = -1.65). Dieser Unterschied fällt statistisch signifikant aus ($p < .001$). Der mittlere Anstieg γ_{10} von 0.24 deutet darauf hin, dass Kinder ohne Risikostatus im Durchschnitt weitere 0.24 Items von Schulwoche zu Schulwoche richtig lösen können. Kinder mit einem zu Beginn der ersten Klasse festgestellten Risikostatus im mathematischen Vorwissen haben dagegen einen um γ_{11} = -0.07 geringeren wöchentlichen Lernzuwachs, d. h. sie lösen im Durchschnitt weitere 0.17 Items richtig je Schulwoche. Dieser Unterschied ist statistisch signifikant ($p < .001$). Während Kinder ohne Risikostatus im Schnitt nach vier Schulwochen eine Verbesserung um ein richtig gelöstes Item mehr aufweisen, benötigen Kinder mit festgestelltem Risikostatus für eine derartige Verbesserung fast sechs Schulwochen.

Tabelle 101: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM „Subtraktion im Zwanzigerraum“

Feste Effekte	Koeffizient	Standard-fehler	t-ratio	df	p
Modell für den intercept β_0					
Mittleres Ausgangsniveau γ_{00}	3.98	0.15	26.29	432	< .001
Risikostatus γ_{01}	-1.65	0.27	-6.06	432	< .001
Modell für den slope β_1					
Mittlerer Anstieg γ_{10}	0.24	0.01	24.627	432	< .001
Risikostatus γ_{11}	-0.07	0.02	-3.25	432	< .001

Zufällige Effekte	Standard-abweichung	Varianz-komponente	df	X^2	p
Ausgangsniveau u_{0i}	2.21	4.87	430	1352.34	< .001
Anstieg u_{1i}	0.10	0.01	430	660.43	< .001
Level-1-Fehler e_{ij}	1.83	3.34			

Erläuterungen: df – Anzahl der Freiheitsgrade, p - Signifikanzwert

Abschließend wird der Anteil der Varianz, welcher durch das allgemeine Modell im Vergleich zum unkonditionierten Modell erklärt wird, anhand der in Abschnitt 6.4.3 beschriebenen Formel berechnet, die Ergebnisse sind in Tabelle 102 notiert. Die Ergebnisse zeigen, dass die Zugehörigkeit zur Risikogruppe im Bereich des mathematischen Vorwissens 8.5 % der Varianz des Ausgangsniveaus u_{0i} und 0 % der Varianz des Anstiegs u_{1i} erklärt

Tabelle 102: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Subtraktion im Zwanzigerraum“

Modell	Varianzkomponente des Ausgangsniveaus u_{0i}	Varianzkomponente des Anstiegs u_{1i}
Unkonditioniertes Modell[a]	5.32	0.01
Allgemeines Modell[b]	4.87	0.01
Anteil der erklärten Varianz	8.5 %	0 %

[a] aus Tabelle 100 zu entnehmen
[b] aus Tabelle 101 zu entnehmen

7.5 Darstellung der Ergebnisse hinsichtlich der Ökonomie, Nutzbarkeit und Zumutbarkeit

Um Anhaltspunkte hinsichtlich der Ökonomie der innerhalb vorliegender Arbeit konzipierten CBM zu erhalten, wurden die von den untersuchten Lehrkräften getroffenen Bewertungen der Aussagen 10) bis 13) des unter Abschnitt 6.3.7 beschriebenen Aussagenkatalogs analysiert.

Die Ergebnisse der Einschätzung der Aussage 10) „Der monatliche Einsatz von CBM in meiner Klasse ist realistisch.“ sind in Abbildung 57 grafisch dargestellt. Von den insgesamt 37 Umfrageteilnehmern bewerteten zehn (27 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 13 der Lehrpersonen (35.1 %) wählten die Kategorie „trifft eher zu“. Hingegen empfanden 14 Personen (37.8 %), dass die monatliche Durchführung von CBM in den Klassen einen nicht unwesentlichen Mehraufwand bedeutete und wählten die Antwort „trifft eher nicht zu“. Keine der Lehrkräfte wählte die Option „trifft nicht zu“. Insgesamt empfand die Mehrheit der Befragten (62.1 %) den monatlichen Einsatz der CBM als realistisch.

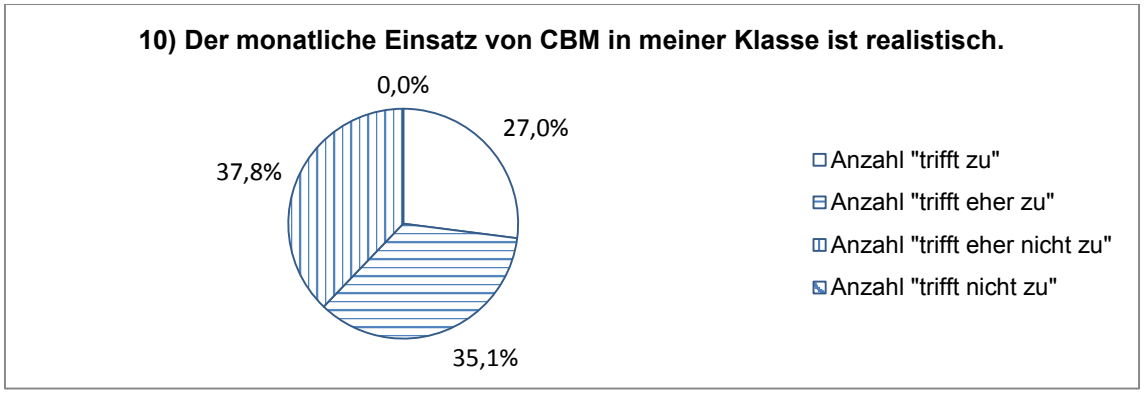


Abbildung 57: Ergebnisse der Beantwortung des Items 10) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Die Beantwortung der Aussage 11) stand in Abhängigkeit zur gewählten Antwortkategorie in Aussage 10). Teilnehmer, die den

monatlichen Einsatz von CBM als eher unrealistisch einschätzten, sollten Angaben zu einer, aus ihrer Sicht, realistischeren Frequentierung geben. Dies betraf 14 Personen. Die Antworten variierten in einem Intervall von mindestens 5 bis höchstens 12 Wochen zwischen den CBM-Testungen. Dabei ergab sich eine mittlere, als realistisch eingeschätzte Frequenz von etwa acht Wochen (M = 8.38 Wochen; SD = 2.29 Wochen). Während etwa 62.1 % der 37 Lehrpersonen (Abbildung 57) eine monatliche Durchführung der CBM im Unterricht für umsetzbar hielt, empfanden die restlichen 37.8 % im Durchschnitt eine Durchführung im zweimonatlichen Rhythmus als realistischer.

Weitere Anhaltspunkte zur Ökonomie der CBM sollten anhand der Aussage 12) „Der regelmäßige Einsatz von CBM bedeutet einen großen Mehraufwand für mich.“ erhoben werden. Die Ergebnisse der Einschätzung sind in Abbildung 58 visualisiert. Von den insgesamt 37 Umfrageteilnehmern bewerteten 12 (32.4 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 14 der Lehrpersonen (37.8 %) wählten die Kategorie „trifft eher zu“. Hingegen empfanden 11 Personen (29.7 %), dass die monatliche Durchführung von CBM in den Klassen eher keinen wesentlichen Mehraufwand bedeutete und wählten die Antwort „trifft eher nicht zu“. Keine der Lehrkräfte wählte die Option „trifft nicht zu“. Die Mehrheit der Befragten (70.2 %) empfand demnach den regelmäßigen Einsatz von CBM als einen hohen Mehraufwand.

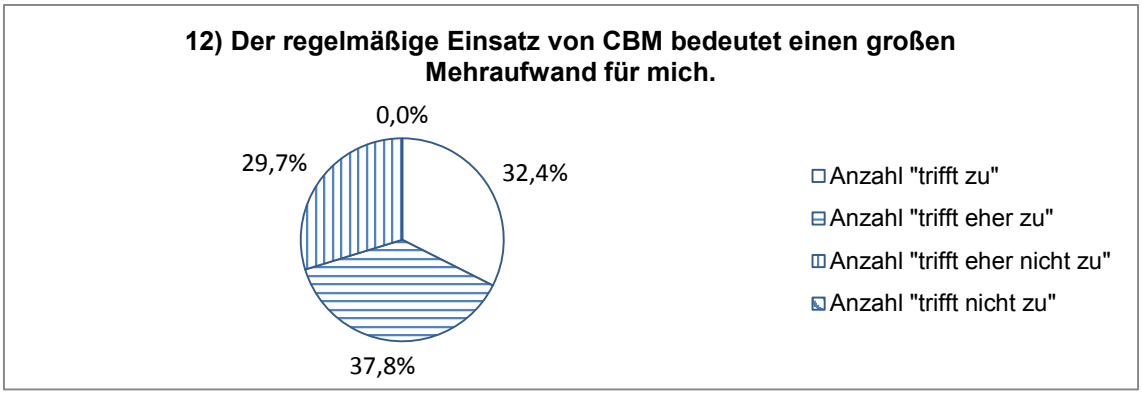


Abbildung 58: Ergebnisse der Beantwortung des Items 12) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Abhängig von der Einschätzung der Aussage 12) sollte die Aussage 13) bearbeitet werden. Teilnehmer, die den regelmäßigen Einsatz von CBM als einen hohen Mehraufwand einschätzten, sollten Angaben dazu treffen, ob dieser, aus ihrer Sicht, gerechtfertigt war und gern in Kauf genommen wurde. Dies betraf 26 Personen (70.2 %). Die Ergebnisse der Befragung sind in Abbildung 59 dargestellt. Die Antworten variierten über alle Kategorien. Dabei zeigte sich, dass von den insgesamt 26 Umfrageteilnehmern zwei (7.7 %) diese Angabe mit der Antwortkategorie „trifft zu“ bewerteten. Weitere 16 (61.5 %) der Lehrpersonen wählten die Kategorie „trifft eher zu“. Hingegen empfanden sechs Personen (23.1 %), dass der durch die regelmäßigen CBM-Testungen aufkommende Mehraufwand für sie eher unangemessen erschien und wählten die Antwort „trifft eher nicht zu“. Weitere zwei der Lehrkräfte (7.7 %) wählten die Option „trifft nicht zu“. Zusammenfassend bedeutet das, dass obwohl der regelmäßige Einsatz von CBM für die Mehrheit der Befragen (70.2 % von 37 Personen) einen großen Mehraufwand bedeutete, dieser von der Mehrheit dieser Personen (69.2 % von 26 Personen) als gerechtfertigt bzw. eher gerechtfertigt erlebt und gern in Kauf genommen wurde.

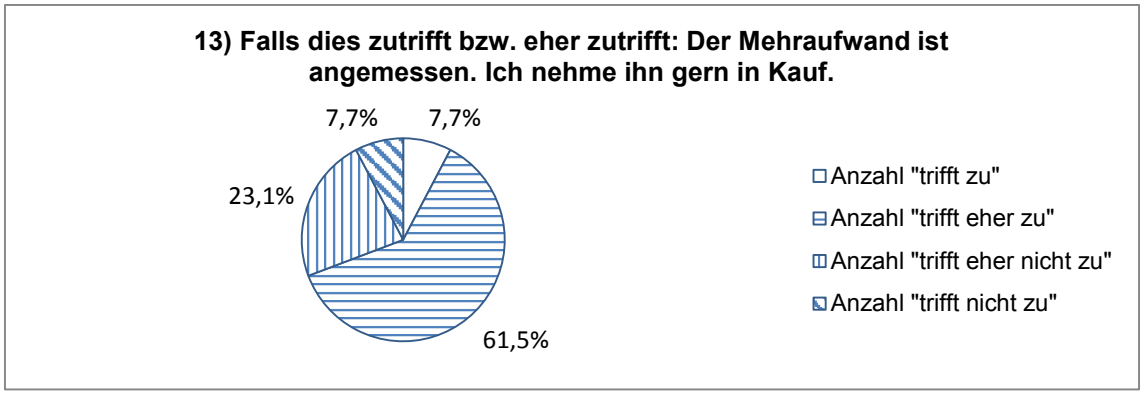


Abbildung 59: Ergebnisse der Beantwortung des Items 13) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 26)

Informationen zur Nutzbarkeit der innerhalb der vorliegenden Arbeit entwickelten CBM sollen anhand der Bewertungsergebnisse der Aussagen 4) bis 9) sowie 14) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz gefolgert werden.

Die Ergebnisse der Aussage 4) „Die Auswertung der Ergebnisse meiner Schüler stellt eine Hilfestellung für meine Unterrichtsgestaltung und für die Planung von Fördermaßnahmen dar.“ sind in Abbildung 60 visualisiert. Von den insgesamt 37 Umfrageteilnehmern bewerteten 16 (43.2 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 15 der Lehrpersonen (40.5 %) wählten die Kategorie „trifft eher zu“. Hingegen empfanden vier Personen (10.8 %), dass die Auswertung der CBM eher keine Hilfestellung zur Unterrichts- und Fördergestaltung darstellte und wählten die Antwort „trifft eher nicht zu“. Zwei der Lehrkräfte (5.4 %) sahen offenbar in den CBM-Ergebnissen der Schüler keine Hilfe zur Planung von Maßnahmen für Unterricht und Förderung und wählten die Option „trifft nicht zu“. Insgesamt lässt sich festhalten, dass die Mehrheit der Lehrpersonen (83.7 %) in den CBM-Daten der Schüler eine Hilfestellung zur Unterrichts- und Förderungsgestaltung sah.

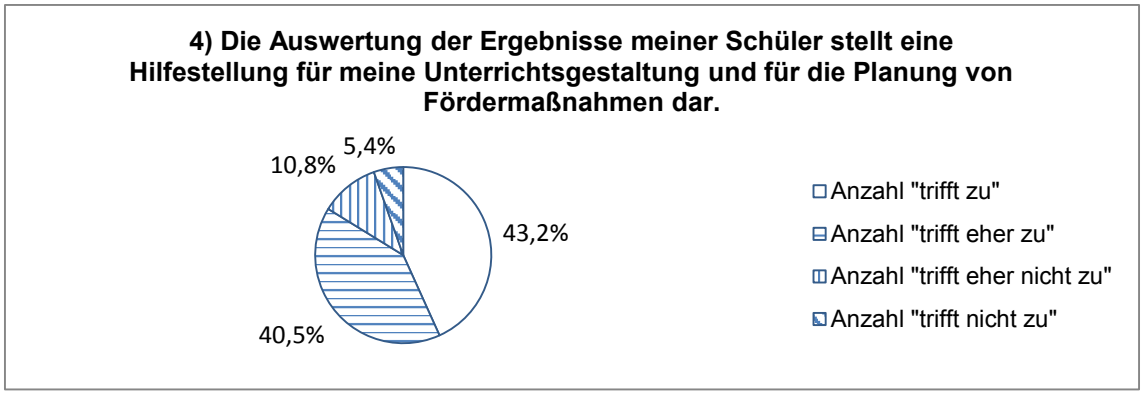


Abbildung 60: Ergebnisse der Beantwortung des Items 4) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Mittels Aussage 5) wurden die Lehrkräfte dahingehend befragt, inwieweit es ihnen gelang, anhand der CBM-Daten schwache Schüler zu identifizieren. Die Ergebnisse der Einschätzung sind in Abbildung 61 dargestellt. Von den insgesamt 37 bewerteten 12 Umfrageteilnehmer (32.4 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 16 der Lehrpersonen (43.2 %) wählten die Kategorie „trifft eher zu“. Acht Personen (21.6 %) hingegen empfanden, dass die monatlich erhobenen CBM-Daten nicht zur Identifikation schwacher Kinder beitrugen und wählten die Antwort „trifft eher nicht zu“. Eine der Lehrkräfte (2.7 %) sah in den CBM-Ergebnissen keine Hilfe zur Identifikation von schwachen Schülern und wählte die Option „trifft nicht zu“. Insgesamt gab die Mehrheit der Befragten (75.6 %) an, dass es ihnen gelang, förderbedürftige Kinder anhand der Schülerergebnisse auszumachen.

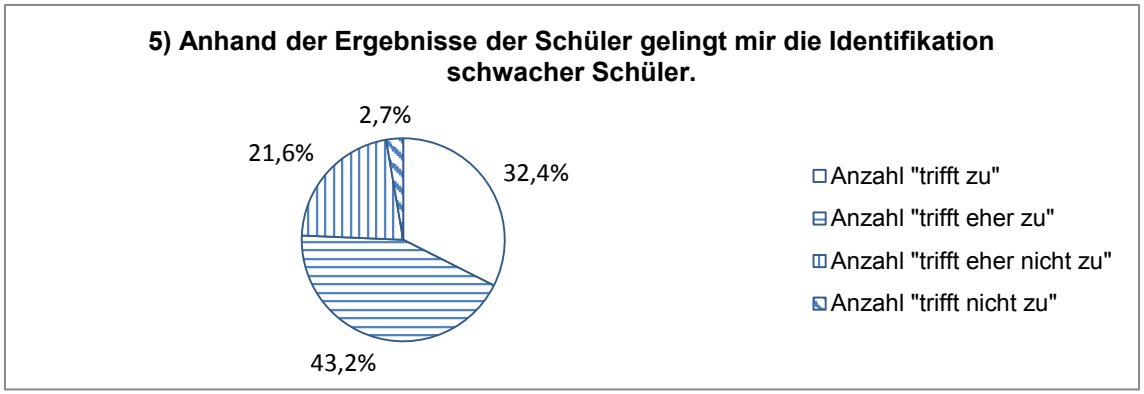


Abbildung 61: Ergebnisse der Beantwortung des Items 5) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Anhand der Aussagen 6) und 7) sollte ermittelt werden, inwieweit die Lehrkräfte die regelmäßig erhobenen CBM-Daten in der Kommunikation mit weiteren Bezugspersonen der Schüler nutzten. Aussage 6) bezog sich dabei auf die Kommunikation mit den Eltern. Die Ergebnisse sind in Abbildung 62 visualisiert. Von den insgesamt 37 Umfrageteilnehmern bewerteten 24 (64.9 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere neun der Lehrpersonen (24.3 %) wählten die Kategorie „trifft eher zu“. Drei der Personen (8.1 %) gaben an, die CBM-Daten in Gesprächen mit Eltern eher weniger genutzt zu haben und wählten die Antwort „trifft eher nicht zu“. Eine der Lehrkräfte (2.7 %) wählte die Option „trifft nicht zu“, sie nutzte die erhoben Schülerdaten im Austausch mit Eltern nicht. Es zeigte sich insgesamt, dass die Mehrheit der Befragten (89.2 %) die CBM-Daten der Schüler (in tabellarischer oder grafischer Form) in Elterngesprächen nutzten.

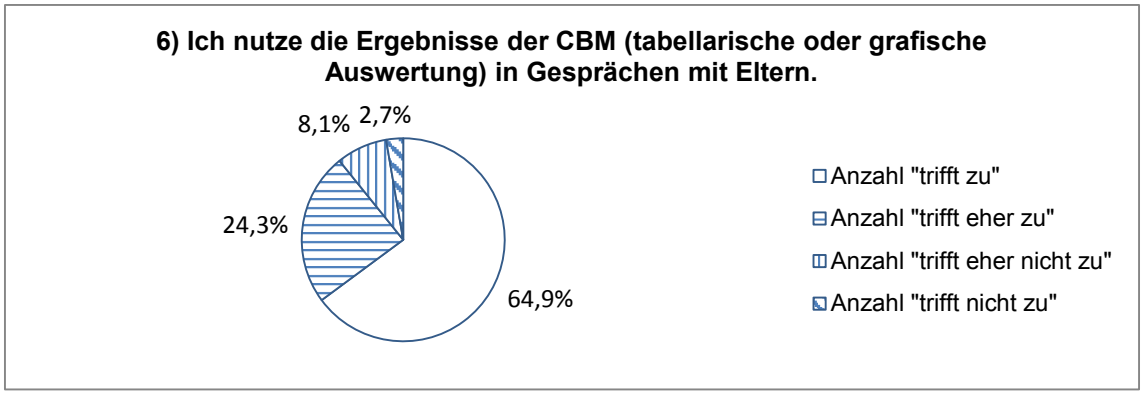


Abbildung 62: Ergebnisse der Beantwortung des Items 6) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Weitere Anhaltspunkte zum Nutzen der CBM-Daten im Austausch mit anderen Bezugspersonen der Schüler liefern die Ergebnisse der Aussage 4) „Ich nutze die Ergebnisse der CBM (tabellarische oder grafische Auswertung) in Gesprächen mit Kollegen.“. Die Ergebnisse sind in Abbildung 63 visualisiert. Von den insgesamt 37 Umfrageteilnehmern bewerteten acht (21.6 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 19 der Lehrpersonen (51.4 %) wählten die Kategorie „trifft eher zu“. Hingegen gaben sieben Personen (18.9 %) an, dass sie die CBM-Daten eher nicht im Gespräch mit Kollegen nutzten und wählten die Antwort „trifft eher nicht zu“. Drei Lehrkräfte (8.1 %) wählten die Option „trifft nicht zu“, nutzten die Daten gar nicht in der Kommunikation mit anderen Lehrpersonen. Insgesamt lässt sich festhalten, dass die Mehrheit der Befragten (73 %) die CBM-Daten der Schüler im Austausch mit Kollegen nutzten.

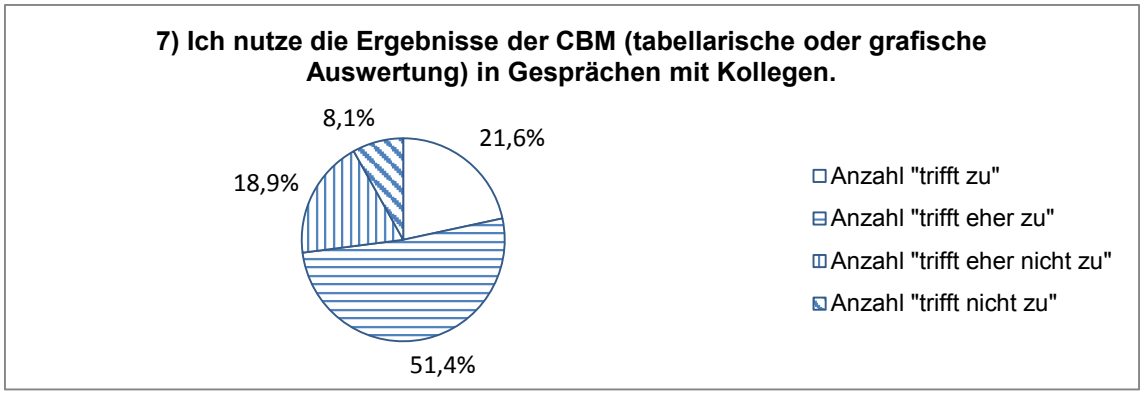


Abbildung 63: Ergebnisse der Beantwortung des Items 7) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Anhand der Aussage 8) „Eine zusätzliche Beratung/Weiterbildung zu Bedeutung sowie Durchführung und Auswertung von CBM würde mir bei der Arbeit helfen.“ sollte ermittelt werden, inwieweit sich die Lehrkräfte über Sinn und Durchführung regelmäßiger CBM-Testungen informiert fühlten. Die Ergebnisse der Einschätzung sind in Abbildung 64 dargestellt. Von den insgesamt 37 Umfrageteilnehmern bewerteten zwei (5.4 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere fünf der Lehrpersonen (13.5 %) wählten die Kategorie „trifft eher zu“. Hingegen empfanden 21 Person (56.8 %), dass sie eher keine weitere Beratung bzw. Weiterbildung zu Bedeutung und Durchführung der CBM benötigen und wählten die Antwort „trifft eher nicht zu“. Neun der Lehrkräfte (24.3 %) wählten die Option „trifft nicht zu“, sie schienen sich ausgiebig über Bedeutung und Durchführung der CBM informiert zu fühlen. Insgesamt gab die Mehrheit der Befragten (81.1 %) an, keine weitere Beratung oder Fortbildung zu benötigen.

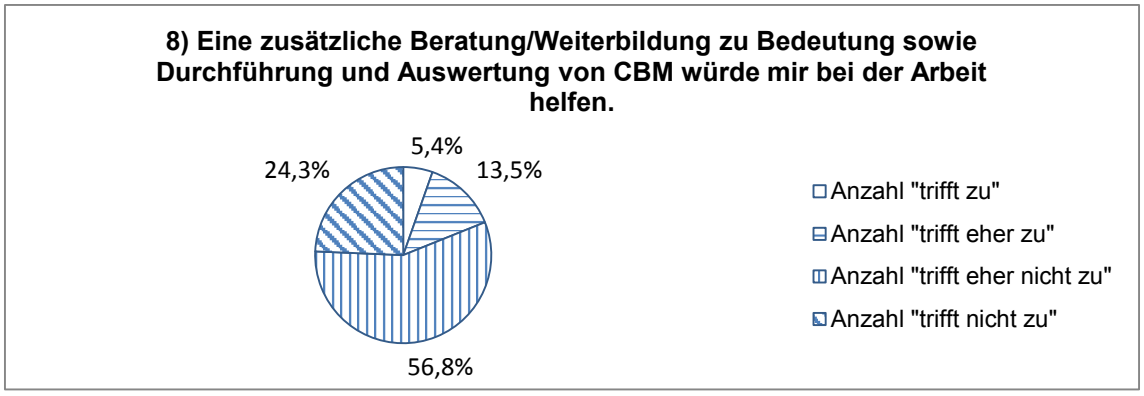


Abbildung 64: Ergebnisse der Beantwortung des Items 8) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Inwieweit die Lehrkräfte weitere spezifische Beratungs- oder Weiterbildungsangebote zur Förderung von schwachen Lernern als notwendig erachteten, sollte anhand der Aussage 9) „Ich brauche zusätzliche Beratung/Information wie ich identifizierte Risikokinder spezifisch fördern kann.“ ermittelt werden. Die Ergebnisse der Einschätzung sind in Abbildung 65 visualisiert. Von den insgesamt 37 Umfrageteilnehmern bewerteten sechs (16.2 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 12 der Lehrpersonen (32.4 %) wählten die Kategorie „trifft eher zu“. Hingegen empfanden insgesamt 19 Person (51.2 %), dass sie keine bzw. eher keine weitere Beratung bzw. Weiterbildung zur spezifischen Förderung schwacher Kinder benötigen und wählten die Antwort „trifft eher nicht zu“ (14 Personen, d. h. 37.8 %) bzw. „trifft nicht zu“ (5 Personen, d. h. 13.5 %).

Die Ergebnisse deuten auf eine recht ausgewogene Einschätzung der Aussage hin. Während etwa die Hälfte der Befragten (51.2 %) angab, keine weiteren Informationen hinsichtlich möglicher Fördermaßnahmen für schwache Lerner zu benötigen, zeigten die anderen Lehrkräfte (48.8 %) weiteren Bedarf in Form von Beratung und Fortbildung an. Es lässt sich an dieser Stelle vermuten, dass die Lehrkräfte der Region Rügen aufgrund der umfangreichen Fortbildungen, welche mit dem in Kapitel 3.1.4.4 erwähnten Beschulungskonzept des „Rügener Inklusionsmodells“ verbundenen

sind, eher zu der Gruppe zählten, die aus ihrer Sicht keine weiteren Informationen bezüglich spezifischer Fördermaßnahmen benötigten. Die Ergebnisse offenbaren jedoch ein anderes Bild, so entstammten 55.6 % der Gruppe mit Bedarf an weiteren Informationen der Region Rügen. Ebenso machten die Rügener Lehrkräfte knapp die Hälfte von der Gruppe der Personen ohne weiteren Bedarf an Weiterbildungen aus.

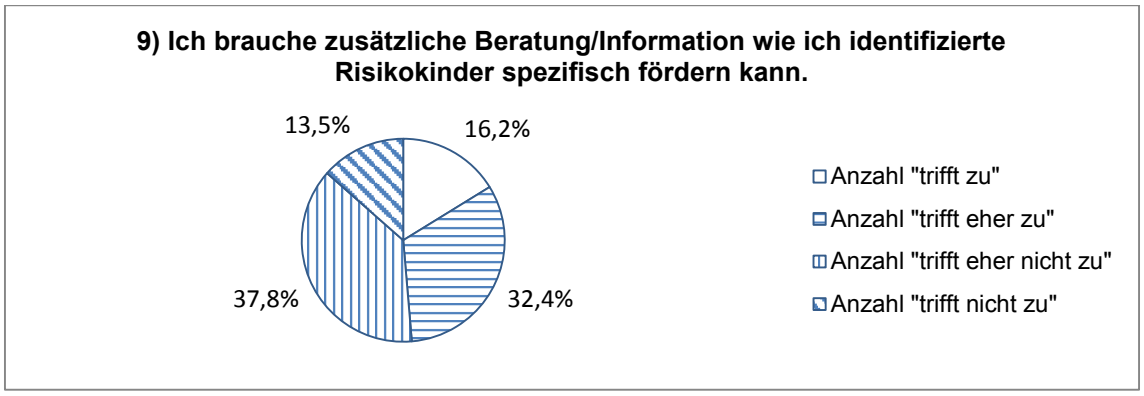


Abbildung 65: Ergebnisse der Beantwortung des Items 9) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N =37)

Mittels Aussage 14) wurden die Lehrkräfte dahingehend befragt, inwieweit sie es für sinnvoll hielten, die CBM-Daten der Schüler computergestützt auswerten zu lassen. Die Ergebnisse der Einschätzung sind in Abbildung 66 dargestellt. 18 (48.6 %) der insgesamt 37 Umfrageteilnehmer bewerteten diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 12 der Lehrpersonen (32.4 %) wählten die Kategorie „trifft eher zu“. Insgesamt schätzten sieben Personen (18.9 %) die computergestützte Auswertung der CBM-Daten als nicht nützlich bzw. eher nicht nützlich ein, sie wählten entsprechend die Antwortkategorien „trifft eher nicht zu“ (vier Personen, d. h. 10.8 %) bzw. „trifft nicht zu“ (drei Personen, d. h. 8.1 %). Insgesamt gab jedoch die Mehrheit der Befragten (81 %) an, dass sie die computergestützte Auswertung der regelmäßig erhobenen CBM-Daten als sinnvoll erachteten.

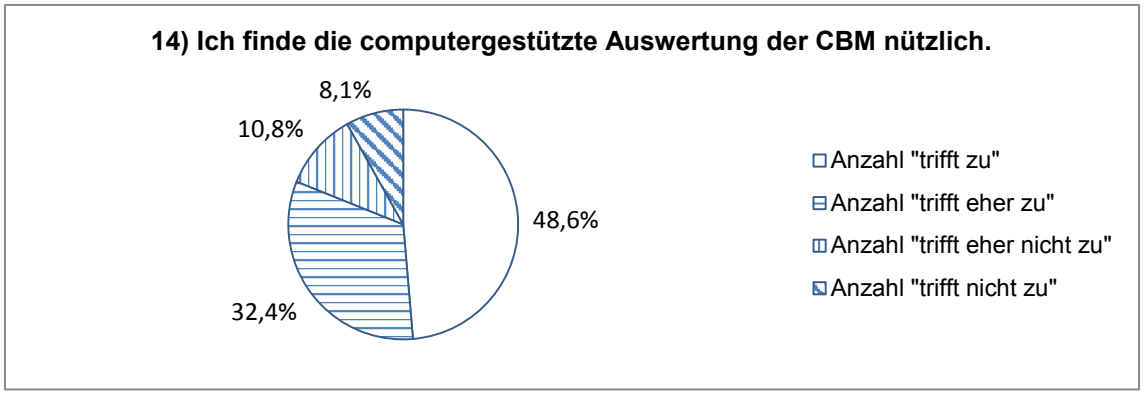


Abbildung 66: Ergebnisse der Beantwortung des Items 14) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

Hinweise bezüglich der Zumutbarkeit der innerhalb der vorliegenden Arbeit entwickelten CBM liefert die Bewertung der Aussage 2) durch die Lehrkräfte. Die Ergebnisse der Einschätzung der Aussage 2) „Das Bearbeiten der CBM hat den Kindern Spaß gemacht.“ sind in Abbildung 67 grafisch dargestellt. Von den insgesamt 37 Umfrageteilnehmern bewerteten 11 (29.7 %) diese Angabe mit der Antwortkategorie „trifft zu“. Weitere 20 der Lehrpersonen (54.1 %) wählten die Kategorie „trifft eher zu“. Hingegen hatten sechs Personen (16.2 %) den Eindruck, dass den Schülern das Bearbeiten der monatlichen CBM eher keinen Spaß gemacht hat und wählten die Antwort „trifft eher nicht zu“. Keine der Lehrkräfte wählte die Option „trifft nicht zu“. Insgesamt gab die Mehrheit der befragten Lehrkräfte (83.8 %) an, dass das Bearbeiten der CBM den Kindern Spaß gemacht bzw. eher Spaß gemacht hat.

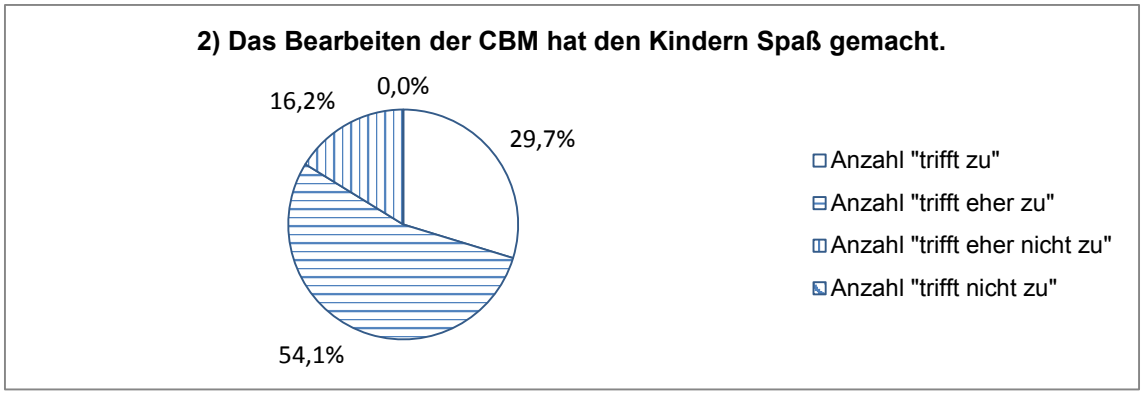


Abbildung 67: Ergebnisse der Beantwortung des Items 2) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37)

8 Zusammenfassung der Ergebnisse

Im folgenden Kapitel werden die Ergebnisse hinsichtlich der Reliabilität, Validität, Änderungssensibilität sowie zur Ökonomie, Nutzbarkeit und Zumutbarkeit der hier beschriebenen CBM zusammengefasst. Dabei werden die Ergebnisse jeweils hypothesenweise systematisiert dargestellt.

8.1 Zusammenfassung der Ergebnisse zur Reliabilität

Im folgenden Abschnitt werden die Ergebnisse hinsichtlich der Reliabilität der CBM zusammengefasst. Die Analysen bezogen sich auf die Splithalf-Reliabilität, Paralleltest-Reliabilität sowie die Retest-Reliabilität.

$H_{2\|1}$	Die Koeffizienten hinsichtlich der Splithalf-Reliabilität liegen im hohen Bereich ($r > .7$).

Für die CBM der Niveaustufe 1 wurden im Abschnitt 7.2.1 Reliabilitätskoeffizienten nach der Splithalf-Methode über die Teilung der Testbearbeitungszeit ermittelt. Die Reliabilitätswerte für die Versionen des CBM „Zahlenlesen im Zwanzigerraum“ streuten in einem Bereich von $r = .88$ bis $r = .95$ (Durchschnittswert $r = .91$) und sind damit insgesamt als hoch bis sehr hoch einzustufen. Geringere, aber dennoch durchweg hohe Werte ergaben sich für die Versionen des CBM „Zahlenreihe im Zwanzigerraum“ mit Koeffizienten in einem Intervall von $r = .77$ bis $r = .88$ (Durchschnittswert $r = .82$). Ein nicht ganz eindeutiges Bild ergab sich für die Versionen der CBM „Mengenvergleich im Zwanzigerraum“ sowie „Arithmetische Basiskompetenzen“, hier streuten die Koeffizienten in einem Bereich von gering bis hoch. Für eine Version des CBM „Mengenvergleich im Zwanzigerraum“ ergaben sich Werte im Bereich von $r = .44$ und $r = .76$ (Durchschnittswert: $r = .57$) und sind als mittel bis hoch einzustufen. Zwei Versionen des CBM „Mengenvergleich im

Zwanzigerraum“ konnten nicht ausgewertet werden, für sie liegen demnach keine Angaben zur Splithalf-Reliabilität vor. Die Reliabilitätskoeffizienten der Versionen des CBM „Arithmetische Basiskompetenzen“ streuen in einem Bereich von r = .35 und r = .89 (Durchschnittswert von r = .64). Ein Wert ist dabei als gering einzuschätzen, alle weiteren liegen im mittleren bis hohen Bereich.

$H_{2\|2}$	Die in Abständen von mehreren Schulwochen erhobenen Koeffizienten hinsichtlich der Retest-Reliabilität liegen zumindest im mittleren Bereich (r > .5).

Im Rahmen dieser Studie konnte für vereinzelte Versionen eines jeden CBM Aussagen zur Retest-Reliabilität getroffen werden. Dazu wurden diese Verfahren wiederholt eingesetzt. Für die CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ sowie „Mengenvergleich im Zwanzigerraum“ wurden jeweils die ersten Versionen zum ersten und zum 11 Wochen später festgelegten vierten Messzeitpunkt eingesetzt. Für das CBM „Arithmetische Basiskompetenzen“ lassen sich Aussagen zur Retest-Reliabilität zu den Versionen 1 bis 7 treffen. Die Abstände zwischen den Testungen variieren zwischen fünf und 25 Schulwochen. Die Versionen 1 und 2 der CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ sind jeweils zweimal mit einem zeitlichen Abstand von 17 (Version 1) bzw. 12 Wochen (Version 2) eingesetzt worden. Für das CBM „Mengenvergleich im Zwanzigerraum“ ergibt sich ein geringer Retest-Reliabilitätskoeffizient von r = .33, die CBM „Zahlenlesen im Zwanzigerraum“ und „Zahlenreihe im Zwanzigerraum“ weisen jeweils mittlere Korrelationen von r = .58 bzw. r = .53 auf. Alle ermittelten Werte für das CBM „Arithmetische Basiskompetenzen“ liegen in einem mittleren Bereich von r = .51 und r = .69, wobei sich keine zu verallgemeinernde Relation zwischen der zeitlichen Frequenz der wiederholten Testung sowie der Höhe der Korrelation erkennen lässt. Alle berechneten Retest-Reliabilitätswerte liegen für die CBM

„Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ im mittleren Bereich zwischen r = .53 und r = .60, bis auf für die Version 2 des CBM „Subtraktion im Zwanzigerraum“, für welches eine hohe Korrelation von r = .72 (zeitlicher Abstand 12) ermittelt wurde.

$H_{2\|3}$	Korrelationen der Ergebnisse zweier als parallel angenommener CBM, welche zeitgleich erhoben werden, liegen im hohen Bereich (r > .7).

Zur Überprüfung der Hypothese wurden in der Region Nordvorpommern zu jedem der etwa monatlichen Erhebungszeitpunkte jeweils zwei Versionen eines jeden CBM durchgeführt. Die Schülerergebnisse beider Versionen wurden korreliert. Die Ergebnisse der Analysen wurden in Abschnitt 7.2.3 berichtet. Für das CBM „Zahlenlesen im Zwanzigerraum“ streuen die Werte der ermittelten Paralleltestkoeffizienten in einem Intervall von r_{tt} = .86 und r_{tt} = .94 (Durchschnittswert r_{tt} = .90), welche als hoch bis sehr hoch angesehen werden können. Die sich in den Verteilungen je zweier zeitgleich eingesetzter Versionen ermittelten signifikanten Unterschiede weisen keine bzw. lediglich geringe Effekte auf (zwischen d = 0.15 und d = 0.41). Auch für das CBM „Zahlenreihe im Zwanzigerraum“ konnten durchweg hohe bis sehr hohe Reliabilitätskoeffizienten im Intervall von r_{tt} = .80 und r_{tt} = .90 (Durchschnittswert von r_{tt} = .85) ermittelt werden. Zwischen den Verteilungen der zu einem Messzeitpunkt eingesetzten Versionen ergeben sich jeweils signifikante Mittelwertunterschiede, welche keine bzw. eine lediglich geringe Effekte aufweisen (zwischen d = 0.17 und d = 0.48). Die Korrelationen zwischen jeweils zwei zeitgleich eingesetzten Versionen des CBM „Mengenvergleich im Zwanzigerraum“ streuen in einem als hoch einzuschätzenden Bereich von r_{tt} = .74 und r_{tt} = .76 (Durchschnittswert r_{tt} = .76). Auch bei diesem CBM zeigen sich signifikante Mittelwertunterschiede, welche bis auf den ersten Messzeitpunkt (d = 0.74, mittlerer Effekt)

keine oder nur geringe Effekte aufzeigen (zwischen d = 0.11 und d = 0.41). Für das CBM „Arithmetische Basiskompetenzen“ variieren die Werte der berechneten Paralleltestkoeffizienten in einem Bereich von r_{tt} = .65 und r_{tt} = .82 (Durchschnittswert r_{tt} = .75), welche als mittel bis hoch eingeschätzt werden können. Die sich in den Verteilungen je zweier zeitgleich eingesetzter Versionen ergebenden signifikanten Unterschiede weisen teilweise keine, geringe sowie mittlere Effekte auf (zwischen d = 0.11 und d = 0.58). Beim CBM „Zahlzerlegung im Zwanzigerraum“ ergeben sich mittlere bis hohe Paralleltest-Reliabilitätskoeffizienten zwischen r_{tt} = .67 und r_{tt} = .78 (Durchschnittswert r_{tt} = .73). In zwei Erhebungszeiträumen ergeben sich Mittelwertunterschiede zwischen den beiden jeweils eingesetzten Versionen des CBM zur Zahlzerlegung bei einer geringen bzw. mittleren Effektstärke (zwischen d = 0.21 und d = 0.53). Die Realibilitätskoeffizienten des CBM „Addition im Zwanzigerraum“ liegen in einem mittleren bis hohen Bereich zwischen r_{tt} = .66 und r_{tt} = .79 (Durchschnittswert r_{tt} = .73). Die Verteilungen jeweils zweier zu einem Messzeitpunkt eingesetzter Versionen des CBM zur Addition deuten durchgängig auf signifikante Mittelwertunterschiede hin, welche eine geringe bis mittlere Effektstärke aufweisen (zwischen d = 0.42 und d = 0.54). Für das CBM „Subtraktion im Zwanzigerraum“ konnten durchweg mittlere bis hohe Reliabilitätskoeffizienten im Intervall von r_{tt} = .67 und r_{tt} = .80 (Durchschnittswert r_{tt} = .77). ermittelt werden. Zwischen den Verteilungen der zu einem Messzeitpunkt eingesetzten Versionen ergeben sich jeweils signifikante Unterschiede, welche keine bzw. eine lediglich geringe Effekte aufwiesen (zwischen d = 0.10 und d = 0.22).

Darüber hinaus zeigt sich, dass, bis auf wenige Ausnahmen, die Mittelwerte der jeweils als zweites durchgeführten Version eines CBM immer höher als die der zuerst eingesetzten Version ausfallen. Beim CBM „Arithmetische Basiskompetenzen“ findet hierbei ein Wechsel

zur Schuljahresmitte statt. Während in der ersten Hälfte des Schuljahres die jeweils als zweite eingesetzte CBM-Version die höheren Mittelwerte aufweist, ist es in der zweiten Schuljahreshälfte die zuerst eingesetzte Version. Beim CBM „Zahlzerlegung im Zwanzigerraum“ ist ebenfalls ein solcher Wechsel der höheren Mittelwerte festzuhalten, jedoch nicht so systematisch wie beim CBM „Arithmetische Basiskompetenzen“.

8.2 Zusammenfassung der Ergebnisse zur Validität

Im folgenden Abschnitt werden die Ergebnisse hinsichtlich der Validität der CBM zusammengefasst. Dabei wird auf Aspekte der Konstrukt- sowie Kriteriumsvalidität eingegangen.

$H_{3\|1}$	Es bestehen zumindest mittlere Zusammenhänge ($r > .7$) zwischen den einzelnen CBM der Niveaustufen 1 und 2 (Konstruktvalidität).

Um die entwickelten CBM auf Zusammenhänge zu prüfen, wurden getrennt nach Niveaustufen Interkorrelationen zu jedem Messzeitpunkt berechnet. Für die Niveaustufe 1 wurden Korrelationen über die Messzeitpunkte 1 bis 4 (Tabelle 45 bis Tabelle 48) bestimmt. Dabei zeigte sich, dass alle CBM der Niveaustufe 1 geringe bis mittlere statistische Zusammenhänge im Bereich von $r = .38$ und $r = .56$ aufweisen. Für die CBM der Niveaustufe 2 erfolgten die Zusammenhangsanalysen über die Daten der Messzeitpunkte 5 bis 9 (Tabelle 49 bis Tabelle 53). Hier zeigte sich ein unterschiedliches Bild. Während für die erhobenen Daten des CBM „Arithmetische Basiskompetenzen“ mit den Daten der weiteren innerhalb der Niveaustufe 2 eingesetzten CBM geringe bis mittlere Korrelationskoeffizienten im Bereich von $r = .40$ und $r = .56$ ermittelt wurden, wiesen die CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“

untereinander durchweg mittlere bis hohe Korrelationen in einem Intervall von r = .58 bis r = .75 auf.

$H_{3\|2}$	Es bestehen mittlere Zusammenhänge (.5 < r ≤ .7) zwischen den einzelnen CBM der Niveaustufen 1 und den mathematischen Schülerleistungen gemessen mit dem Diagnoseprogramm Kalkulie Teil 1 (Fritz et al., 2007) zu Beginn der ersten Klasse (konvergente bzw. Übereinstimmungsvalidität).

Zur Einschätzung der konvergenten bzw. Übereinstimmungsvalidität wurden in der Region Rügen (Untersuchungsgruppe 1) zum ersten Messzeitpunkt in der 5. Schulwoche neben den CBM der Niveaustufe 1 das Diagnoseprogramm Kalkulie (Fritz et al., 2007) durchgeführt (Tabelle 54 bis Tabelle 58). Für die CBM „Zahlenlesen im Zwanzigerraum“ sowie „Mengenvergleich im Zwanzigerraum“ ergaben sich mittlere Korrelationen von r = .55 bzw. r = .51. Die berechnete Korrelation für das CBM „Zahlenreihe im Zwanzigerraum“ blieb mit einem Wert von r = .47 knapp unterhalb eines als mittel einzuschätzenden Koeffizienten aus. Noch geringer fiel die Korrelation zwischen Kalkulie (Fritz et al., 2007) und dem CBM „Arithmetische Basiskompetenzen“ aus (r = .41).

$H_{3\|3}$	Es bestehen mittlere Zusammenhänge (.5 < r ≤ .7) zwischen den einzelnen CBM der Niveaustufe 2 und den mathematischen Schülerleistungen gemessen mit dem DEMAT 1+ (Krajewski et al., 2002) zum Ende der ersten Klasse (konvergente bzw. Übereinstimmungsvalidität).

Zum neunten Messzeitpunkt in der 38. Schulwoche wurde in der Region Rügen (Untersuchungsgruppe 1) neben den CBM der Niveaustufe 2 zusätzlich das Diagnoseverfahren DEMAT 1+ (Krajewski et al., 2002) als Kriteriumsmaß zur Bestimmung der konvergenten bzw. Übereinstimmungsvalidität eingesetzt (Tabelle 58 bis Tabelle 61). Für die CBM „Zahlzerlegung im Zwanzigerraum“,

„Addition im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ ergaben sich mittlere Korrelationen zwischen r = .56 und r = .64. Die berechnete Korrelation für das CBM „Arithmetische Basiskompetenzen“ fällt mit von r = .45 gering aus.

$H_{3\|4}$	Es bestehen mittlere Zusammenhänge (.5 < r ≤ .7) zwischen den einzelnen CBM der Niveaustufen 1 und 2 und den mathematischen Schülerleistungen gemessen mit dem Inventar Rechenfische (Wagner & Hartke, 2006) zu verschiedenen Messzeitpunkten innerhalb der ersten Klasse (konvergente bzw. Übereinstimmungsvalidität).

Zum fünften Messzeitpunkt in der 21. Schulwoche (Tabelle 62) sowie zum achten Messzeitpunkt in der 34. Schulwoche (Tabelle 63) wurde in der Region Rügen (Untersuchungsgruppe 1) neben den CBM der Niveaustufe 2 zusätzlich das Inventar „Rechenfische“ (Wagner & Hartke, 2006) als Kriteriumsmaß zur Bestimmung der Übereinstimmungsvalidität eingesetzt. Die CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ zeigen über beide Messzeitpunkte hinweg mittlere bis hohe Korrelationskoeffizienten in einem Intervall von r = .55 bis r = .70. Für das CBM „Arithmetische Basiskompetenzen“ konnten eine mittlere Korrelation (r = .51) zum fünften sowie eine geringe Korrelation (r = .43) zum achten Messzeitpunkt errechnet werden. Um darüber hinaus weitere Hinweise zur Validität der CBM der Niveaustufe 1 zu erhalten, wurden auch die Korrelationen des Inventars „Rechenfische“ (Wagner & Hartke, 2006) mit den Daten der CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ sowie „Mengenvergleich im Zwanzigerraum“ zum vierten Messzeitpunkt (16. Schulwoche) berechnet. Bis auf den Koeffizienten für das Zahlenlesen (geringe Korrelation von r = .44) sind die Korrelationen hier als mittel einzuschätzen (r = .50 bzw. r = .54).

$H_{3\|5}$	Es bestehen lediglich geringe Zusammenhänge (r < .5) zwischen den einzelnen CBM der Niveaustufe 1 und den Voraussetzungen zum erfolgreichen Leseerwerb der Schüler gemessen mit dem MÜSC (Mannhaupt, 2005) zu Beginn der ersten Klasse (diskriminante Validität).

Zum ersten Messzeitpunkt in der 5. Schulwoche wurde in der Region Rügen (Untersuchungsgruppe 1) neben den CBM der Niveaustufe 1 auch das MÜSC (Mannhaupt, 2005) als Kriteriumsmaß zur Bestimmung der diskriminanten Validität eingesetzt (Tabelle 54 bis Tabelle 58). Für alle hierbei geprüften CBM ergaben sich geringe Korrelationen in einem Intervall von r = -.22 bis r = -.35. Der gegensinnige Zusammenhang lässt sich dadurch erklären, dass bei den CBM die Anzahl richtig gelöster Aufgaben, beim MÜSC (Mannhaupt, 2005) hingegen Risikopunkte erhoben werden.

$H_{3\|6}$	Es bestehen lediglich geringe Zusammenhänge (r < .5) zwischen den einzelnen CBM der Niveaustufen 1 und 2 und den Leseleistungen der Schüler gemessen mit der WLLP (Küspert & Schneider, 1998) zum Ende der ersten Klasse (diskriminante Validität).

Zum neunten Messzeitpunkt in der 38. Schulwoche wurde in der Region Rügen (Untersuchungsgruppe 1) neben den CBM der Niveaustufe 2 zusätzlich die Würzburger Leise Leseprobe (Küspert & Schneider, 1998) als Kriteriumsmaß zur Bestimmung der konvergenten Validität eingesetzt (Tabelle 58 bis Tabelle 61). Für alle CBM der Niveaustufe 2 wurden dabei geringe Korrelationen in einem Bereich von r = .30 und r = .45 ermittelt.

$H_{3\|7}$	Die über die Messzeitpunkte 1 bis 4 erhobenen Schülerergebnisse innerhalb der CBM der Niveaustufe 1 erlauben eine valide Prognose der Mathematikleistungen zum Ende der ersten Klasse erhoben mit dem DEMAT 1+ (Krajewski et al., 2002) (prognostische Validität).

Die Bestimmung der Prognosegüte der CBM erfolgte anhand von klassifikatorischen Analysen. Dabei wurde die Fähigkeit der Verfahren zur Identifikation von Kindern mit niedrigen Rechenleistungen zum Ende der ersten Klasse (Kriterium: Prozentrang kleiner bzw. gleich 15 im DEMAT 1+; Krajewski et al., 2002) ermittelt (Tabelle 64 bis Tabelle 67). Mit einem RATZ-Index zwischen RATZ = 0.40 bis 0.50 über die Messzeitpunkte 1 bis 4 liefert das CBM „Zahlenlesen im Zwanzigerraum“ nach Einteilung von Jansen et al. (1999) eine gute Klassifikationsleistung. Dennoch weist das CBM eine relativ hohe α-Fehlerquote (zwischen 26.0 % und 34.2 %) sowie β-Fehlerquote (zwischen 30.8 % und 40.4 %) auf. Sehr ähnliche Befunde ergeben sich für die CBM „Zahlenreihe im Zwanzigerraum“ (RATZ-Index: zwischen 0.38 und 0.66; α-Fehlerquote: zwischen 25.6 % und 32.1 %; β-Fehlerquote: zwischen 21.2 % und 40.0 %), „Mengenvergleich im Zwanzigerraum“ (RATZ-Index: zwischen 0.35 und 0.62; α-Fehlerquote: zwischen 15.8 % und 33.6 %; β-Fehlerquote: zwischen 26.9 % und 51.9 %) sowie „Arithmetische Basiskompetenzen“ zu den Messzeitpunkten eins bis vier (RATZ-Index: zwischen 0.41 und 0.54; α-Fehlerquote: zwischen 12.2 % und 44.1 %; β-Fehlerquote: zwischen 24.5 % und 46.2 %).

$H_{3\|8}$	Die über die Messzeitpunkte 5 bis 9 erhobenen Schülerergebnisse innerhalb der CBM der Niveaustufe 2 erlauben eine valide Prognose der Mathematikleistungen zum Ende der ersten Klasse erhoben mit dem DEMAT 1+ (Krajewski et al., 2002) (prognostische Validität).

Die Bestimmung der Prognosegüte der CBM der Niveaustufe 2 erfolgte ebenfalls anhand klassifikatorischer Analysen. Dabei wurde analog zur Vorgehensweise bezüglich der Niveaustufe 1 verfahren (Tabelle 67 bis Tabelle 70). Bis auf eine Ausnahme (CBM „Subtraktion im Zwanzigerraum“ zur 30. Schulwoche: RATZ = 0.65) weisen die CBM „Zahlerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ mit einem RATZ-Index jenseits der von Jansen et al. (1999) festgesetzten Grenze von RATZ = 0.66 eine sehr gute Klassifikationsleistung auf. Ebenso weisen die Verfahren eine geringe β-Fehlerquote auf (Zahlzerlegung: 4.3 % bis 19.0 %; Addition: 2.1 % bis 14.0 %; Subtraktion: 4.1 % bis 24.0 %). Die α-Fehlerquoten fallen jeweils recht hoch aus (Zahlzerlegung: 17.1 % bis 41.3 %; Addition: 23.5 % bis 47.3 %; Subtraktion: 23.8 % bis 54.9 %). Für das CBM „Arithmetische Basiskompetenzen“ ergibt sich innerhalb der zweiten Schuljahreshälfte kein eindeutiges Bild. In der 21. und 30. Schulwoche liegen die RATZ-Indices bei 0.50 bzw. 0.63, was auf eine gute Klassifikationsleistung hindeutet (Jansen et al., 1999), die α-Fehlerquoten sin relativ hoch (34.5 % bzw. 35.6 %), die β-Fehlerquoten ebenfalls (30.6 % bzw. 22.0 %). Zu den Messzeitpunkten in der 26. sowie 34. Schulwoche konnten RATZ-Indices von 0.77 bzw. 0.70 ermittelt werden, welche nach Jansen et al. (1999) eine sehr gute Klassifikationsleistung des Verfahrens anzeigen. Auch die geringe β-Fehlerquote (8.2 % bzw. 11.9 %) spricht für eine hohe prognostische Validität. Die α-Fehlerquote fällt jedoch auch zu diesen Messzeitpunkten recht hoch aus (60.9 % bzw. 56.1 %).

$H_{3\|9}$	Die erhobenen Schülerergebnisse aller CBM der Niveaustufe 1 zum Messzeitpunkt 1 geben kombiniert betrachtet eine valide Prognose der Mathematikleistungen zum Ende der ersten bzw. zweiten Klasse erhoben mit dem DEMAT 1+ (Krajewski et al., 2002) bzw. mit dem DEMAT 2+ (Krajewski et al., 2004) (inkrementelle Validität).

Im Rahmen der Bestimmung der inkrementellen Validität der CBM der Niveaustufe 1 wurde anhand eines Regressionsmodells der Zusammenhang der Ergebnisse aller zum ersten Messzeitpunkt in der fünften Schulwoche erhobenen CBM-Daten mit einem Risikostatus bezüglich der mathematischen Kompetenzen (Prozentrang kleiner bzw. gleich 15) zum Ende des ersten (DEMAT 1+; Krajewski et al., 2002) sowie des zweiten Schuljahres (DEMAT 2+; Krajewski et al., 2004) bestimmt (Tabelle 71 bzw. Tabelle 73). Insgesamt konnten 27 % der Varianz in der Risikogruppenzugehörigkeit zum Ende der ersten Klasse sowie 31.4 % der zum Ende der zweiten Klasse durch die CBM-Daten der Niveaustufe 1 aufgeklärt werden. Dabei stellte sich heraus, dass die Risikogruppenzugehörigkeit zum Ende der ersten Klasse lediglich durch die Ergebnisse der CBM „Zahlenlesen im Zwanzigerraum“, „Mengenvergleich im Zwanzigerraum“ sowie „Arithmetische Basiskompetenzen“ in signifikanter Weise vorhergesagt werden konnte. Die Daten des CBM „Zahlenreihe im Zwanzigerraum“ lieferten zu diesem Messzeitpunkt hierzu keinen signifikanten Beitrag. Zum Ende der zweiten Klasse konnte die Zugehörigkeit zur Risikogruppe durch die Ergebnisse der CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ sowie „Mengenvergleich im Zwanzigerraum“ in signifikanter Weise vorhergesagt werden. Die Daten des CBM „Arithmetische Basiskompetenzen“ erbrachten zu diesem Messzeitpunkt keinen signifikanten Beitrag.

$H_{3\mid 10}$	Die erhobenen Schülerergebnisse aller CBM der Niveaustufe 2 zum Messzeitpunkt 5 geben kombiniert betrachtet eine valide Prognose der Mathematikleistungen zum Ende der ersten bzw. zweiten Klasse erhoben mit dem DEMAT 1+ (Krajewski et al., 2002) bzw. mit dem DEMAT 2+ (Krajewski et al., 2004) (inkrementelle Validität).

Im Rahmen der Bestimmung der inkrementellen Validität der CBM der Niveaustufe 2 wurde anhand eines Regressionsmodells der Zusammenhang der Ergebnisse aller zum fünften Messzeitpunkt in der 21. Schulwoche erhobenen CBM-Daten mit einem Risikostatus bezüglich der mathematischen Kompetenzen (Prozentrang kleiner bzw. gleich 15) zum Ende des ersten (DEMAT 1+; Krajewski et al., 2002) sowie des zweiten Schuljahres (DEMAT 2+; Krajewski et al., 2004) bestimmt (Tabelle 72 bzw. Tabelle 74). Insgesamt konnten 33.3 % der Varianz in der Risikogruppenzugehörigkeit zum Ende der ersten Klasse sowie 52.8 % der zum Ende der zweiten Klasse durch die CBM-Daten der Niveaustufe 2 aufgeklärt werden. Dabei stellte sich heraus, dass die Risikogruppenzugehörigkeit zum Ende der ersten Klasse lediglich durch die Ergebnisse der CBM „Zahlzerlegung im Zwanzigerraum“ sowie „Addition im Zwanzigerraum“ in signifikanter Weise vorhergesagt werden konnte. Die Daten der CBM „Arithmetische Basiskompetenzen“ sowie „Subtraktion im Zwanzigerraum“ lieferten zu diesem Messzeitpunkt hierzu keinen signifikanten Beitrag. Zum Ende der zweiten Klasse konnte die Zugehörigkeit zur Risikogruppe durch die die Ergebnisse der CBM „Zahlzerlegung im Zwanzigerraum“ sowie „Addition im Zwanzigerraum“ in signifikanter Weise vorhergesagt werden. Die Ergebnisse der CBM „Arithmetische Basiskompetenzen“ sowie „Subtraktion im Zwanzigerraum“ erbrachten zu diesem Messzeitpunkt keinen signifikanten Beitrag.

8.3 Zusammenfassung der Ergebnisse zur Änderungssensibilität

Im folgenden Abschnitt werden die Ergebnisse hinsichtlich der Änderungssensibilität zusammengefasst.

$H_{4\|1}$	Die Korrelationen zwischen den CBM-Ergebnissen werden kleiner, je größer der Zeitraum zwischen den Erhebungszeitpunkten ist.

Bei der in Abschnitt 7.4 vorgenommenen Analysen der Korrelationen der CBM-Ergebnisse über die Messzeitpunkte hinweg, zeigt sich für die Verfahren „Zahlenlesen im Zwanzigerraum“ (Tabelle 75), „Mengenvergleich im Zwanzigerraum“ (Tabelle 83) sowie „Subtraktion im Zwanzigerraum“ (Tabelle 99), dass mit zunehmendem zeitlichen Abstand zwischen den Erhebungszeitpunkten ein Abfall der Korrelationskoeffizienten zu verzeichnen ist (Simplexstruktur). Für die CBM „Zahlenreihe im Zwanzigerraum“ (Tabelle 79), „Arithmetische Basiskompetenzen“ (Tabelle 87) und „Addition im Zwanzigerraum“ (Tabelle 95) gilt dies nur eingeschränkt, hier beinträchtigen einzelne Korrelationen die sonst vorliegende Simplexstruktur der Daten. Einzig für das CBM „Zahlzerlegung im Zwanzigerraum“ (Tabelle 91) sind durchweg steigende Korrelationen jeweils zum achten und zum neunten Messzeitpunkt zu beobachten.

$H_{4\|2}$	Die Splithalf-Reliabilität der CBM ist höher als die Retest- bzw. Paralleltest-Reliabilität.

Verglichen mit den ermittelten Rest-Reliabilitätskoeffizienten (Abschnitt 7.2.2.1 bis Abschnitt 7.2.2.7) ist festzuhalten, dass die berechneten Splithalf-Reliabilitätskoeffizienten der CBM der Niveaustufe 1 (Tabelle 33 bis Tabelle 36) bis auf wenige Ausnahmen (CBM „Arithmetische Basiskompetenzen“ Version 2 sowie 3) durchweg höher ausfallen. Bezogen auf die Paralleltest-Reliabilität (Tabelle 38 bis Tabelle 44) fielen die Koeffizienten zur Splithalf-Reliabilität für das CBM „Zahlenlesen im Zwanzigerraum“ ebenfalls

höher aus, beim CBM „Zahlenreihe im Zwanzigerraum“ waren die Werte annähernd gleich. Die Koeffizienten zur Splithalf-Reliabilität der Verfahren „Mengenvergleich im Zwanzigerraum“ sowie „Arithmetische Basiskompetenzen“ waren geringer als die ermittelten Paralleltest-Reliabilitätskoeffizienten. Für die CBM der Niveaustufe 2 wurden keine Analysen zur Splithalf-Reliabilität vorgenommen.

$H_{4\mid3}$	Die Varianzen der Schülerergebnisse in den CBM erhöhen sich von Messzeitpunkt zu Messzeitpunkt.

Während die CBM „Zahlenreihe im Zwanzigerraum“, „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ im Wesentlichen einen Anstieg in der Varianz aufweisen, zeigt sich bei den CBM „Zahlenlesen im Zwanzigerraum“ und „Arithmetische Basiskompetenzen“ keine Systematik in der Varianzentwicklung über die Messzeitpunkte. In den Daten des CBM „Mengenvergleich im Zwanzigerraum“ war ein Abfall der Varianz zu erkennen.

$H_{4\mid4}$	Zwischen den Mittelwerten der Schülerergebnisse in den CBM zu aufeinander folgenden Messzeitpunkten gibt es signifikante Unterschiede.

Zur Untersuchung der innerhalb der vorliegenden Arbeit konzipierten CBM hinsichtlich ihrer Änderungssensibilität wurden in Abschnitt 7.4 Mehrebenenanalysen (HLM) durchgeführt. Dabei wurden durchschnittliche Leistungsanstiege der Schüler für jedes CBM über die knapp monatlich frequentierten Messzeitpunkte errechnet. Für alle der untersuchten CBM ergaben sich Leistungsanstiege, welche sich als signifikant von Null verschieden erwiesen (Tabelle 77, Tabelle 81, Tabelle 85, Tabelle 89, Tabelle 93, Tabelle 97, Tabelle 101).

$H_{4\|5}$	In den einzelnen CBM zeigen die untersuchten Schüler einen durchschnittlichen monatlichen Leistungsanstieg von zumindest einem Punkt.

Für Kinder ohne festgestellte Risiken im mathematischen Bereich zu Beginn der ersten Klasse konnten für die CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“, „Mengenvergleich im Zwanzigerraum“ und „Addition im Zwanzigerraum“ durchschnittliche monatliche Leistungsanstiege von zumindest einem Punkt (wöchentliche Anstiege: ZL: $\gamma_{10} = 1.25$; ZR: $\gamma_{10} = 0.59$; MV: $\gamma_{10} = 0.59$; AD: $\gamma_{10} = 0.26$) festgehalten werden (Tabelle 77, Tabelle 81, Tabelle 85, Tabelle 97). Die durchschnittlichen monatlichen Leistungsanstiege für die CBM „Arithmetische Basiskompetenzen“, „Zahlzerlegung im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ lagen unter einem Punkt (wöchentliche Anstiege: AB: $\gamma_{10} = 0.15$; ZZ: $\gamma_{10} = 0.21$; SU: $\gamma_{10} = 0.24$; vgl. Tabelle 89, Tabelle 93 bzw. Tabelle 101).

Die monatlichen Durchschnittsanstiege der Kinder mit einem festgestellten Risiko im mathematischen Bereich zu Beginn der ersten Klasse fallen für die CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“, „Mengenvergleich im Zwanzigerraum“ und „Addition im Zwanzigerraum“ (Tabelle 77, Tabelle 81, Tabelle 85, Tabelle 97) ebenfalls größer als eins aus, für die CBM „Arithmetische Basiskompetenzen“, „Zahlzerlegung im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ liegen sie unter eins (vgl. Tabelle 89, Tabelle 93 bzw. Tabelle 101).

$H_{4\|6}$	Die durch die CBM erhobenen Daten weisen auf verschiedene Leistungsentwicklungen zwischen Kindergruppen mit unterschiedlich stark ausgeprägtem mathematischen Vorwissen hin: Kinder mit einem Risikostatus im mathematischen Vorwissen zu Beginn der ersten Klasse erzielen geringere Lernzuwächse als Kinder ohne einen solchen Risikostatus.

Während für die CBM „Zahlenreihe im Zwanzigerraum“, „Arithmetische Basiskompetenzen“, „Zahlzerlegung im Zwanzigerraum“ sowie „Addition im Zwanzigerraum“ kein Einfluss eines Risikostatus im mathematischen Bereich zu Beginn der ersten Klasse auf den monatlichen Lernzuwachs der Kinder festgestellt werden konnte (Tabelle 81, Tabelle 89, Tabelle 93, Tabelle 85Tabelle 97), erzielen Kinder mit einem Risiko in den CBM „Mengenvergleich im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ signifikant geringere monatliche Leistungsanstiege im Vergleich zu ihren Klassenkameraden (Tabelle 85 bzw. Tabelle 101). Im CBM „Zahlenlesen im Zwanzigerraum“ erzielen diese Kinder signifikant höhere Leistungsanstiege im Vergleich zu den anderen Kindern (γ_{11} = 0.31; Tabelle 77).

8.4 Zusammenfassung der Ergebnisse zur Ökonomie, Nutzbarkeit und Zumutbarkeit

Nachfolgend werden die Ergebnisse hinsichtlich der Ökonomie, Nutzbarkeit und Zumutbarkeit zusammengefasst. Dabei wird auf die Ergebnisse der Onlinebefragung der Lehrkräfte (UG 3) zurückgegriffen.

$H_{5\|1}$	Die Mehrheit der befragten Lehrkräfte schätzt die innerhalb dieser Forschungsarbeit konzipierten CBM als sinnvoll ein.

Abbildung 55 zeigt, dass die Mehrheit der Umfrageteilnehmern (86.4 %) die in dieser Arbeit beschriebenen CBM als sinnvoll einschätzen.

$H_{5\|2}$	Die Mehrheit der befragten Lehrkräfte schätzt den monatlichen Einsatz innerhalb dieser Forschungsarbeit konzipierter CBM als realistisch ein.

Die Mehrheit der befragten Lehrkräfte (62.1 %) an, dass sie einen monatlichen Einsatz der innerhalb dieser Forschungsarbeit konzipierten CBM als realistisch empfinden (Abbildung 57). Die restlichen 37.9 % der Befragten gaben an, dass das Intervall zwischen den Erhebungen auf durchschnittlich acht Wochen erhöht werden sollte.

$H_{5\|3}$	Die Mehrheit der befragten Lehrkräfte empfindet den regelmäßigen Einsatz innerhalb dieser Forschungsarbeit konzipierter CBM als erheblichen Mehraufwand.

Die Ergebnisse der Lehrerbefragung ergeben, dass die Mehrheit der befragten Lehrkräfte (70.2 %) den regelmäßigen Einsatz innerhalb dieser Forschungsarbeit konzipierter CBM als großen Mehraufwand empfindet (Abbildung 58).

$H_{5\|4}$	Die Mehrheit der befragten Lehrkräfte, die den regelmäßigen Einsatz innerhalb dieser Forschungsarbeit konzipierter CBM als großen Mehraufwand empfindet, schätzt diesen Mehraufwand als angemessen ein und nimmt ihn in Kauf.

Von der Mehrheit der Teilnehmer (69.2 %), die den regelmäßigen Einsatz von CBM als einen hohen Mehraufwand einschätzten, gab an, dass dieser aus ihrer Sicht gerechtfertigt sei und gern in Kauf genommen werde (Abbildung 59).

$H_{5\|5}$	Die Mehrheit der befragten Lehrkräfte nimmt an, dass es anhand der Schülerergebnisse, der innerhalb dieser Forschungsarbeit konzipierten CBM gelingt, Schüler mit schwachen Rechenleistungen zu identifizieren.

Die Mehrheit der Umfrageteilnehmer (75.6 %) gab an, dass es ihnen gelingt, anhand der CBM-Daten schwache Schüler zu identifizieren (Abbildung 61).

$H_{5\|6}$	Die Mehrheit der befragten Lehrkräfte sieht in der Auswertung der Schülerergebnisse in den innerhalb dieser Forschungsarbeit konzipierten CBM eine Hilfestellung für Unterrichtsgestaltung und für die Planung von Fördermaßnahmen.

Die Mehrheit der befragten Lehrkräfte (83.7 %) sieht in der Auswertung der CBM-Ergebnisse eine Hilfestellung für die Unterrichtsgestaltung und Förderplanung (Abbildung 60). Etwa die Hälfte der Befragten (51.2 %) gibt darüber hinaus an, keine weiteren Informationen hinsichtlich möglicher Fördermaßnahmen für schwache Lerner zu benötigen (Abbildung 65).

$H_{5\|7}$	Der Einsatz der CBM im Mathematikunterricht der Klassenstufe 1 ist den Kindern zumutbar.

Die Mehrheit der Befragten (83.3 %) gab an, dass das Bearbeiten der CBM den Kindern Spaß gemacht hat (Abbildung 67).

9 Diskussion der Ergebnisse

Im nachfolgenden Kapitel werden die hier dargestellten Ergebnisse der vorliegenden Studie differenziert für die untersuchten Gütekriterien (Abschnitte 9.1, 9.2, 9.3, 9.4 sowie 9.5) hinsichtlich der von L.S. Fuchs (2004) postulierten Forschungsebenen bei der Evaluation von CBM diskutiert. Neben Aspekten zur Objektivität, Reliabilität und Validität als Hauptgütekriterien werden ebenso die Ergebnisse bezüglich der Änderungssensibilität, der Anwendbarkeit und Zumutbarkeit sowie Ökonomie kommentiert, welche als Nebengütekriterien wesentliche Bedeutung beim praktischen Einsatz im schulischen Alltag besitzen. Dazu werden die jeweiligen Befunde methodenkritisch betrachtet und in aktuelle Forschungsstände aus den USA und Deutschland eingeordnet. In Kapitel 10 werden dann ein zusammenfassendes Fazit sowie ein Ausblick auf wünschenswerte weitere Forschungsbemühungen gegeben.

9.1 Diskussion der Ergebnisse hinsichtlich der Objektivität

In Abschnitt 6.2.1 wurden die ergriffenen Maßnahmen zur Reduktion des subjektiven Einflusses eines Testleiters bei der Durchführung, Auswertung und Interpretation der CBM beschrieben. In einem Manual wurden standardisierte Durchführungsbedingungen vorgegeben. Diese umfassten neben allgemeinen Aspekten auch spezifische Hinweise zur Durchführung (z. B. benötigtes Material wörtliche Instruktion) und Auswertung (z. B. Lösungsfolien) der einzelnen innerhalb dieser Studie erarbeiteten CBM. Zur Unterstützung bei der Interpretation der Schülerleistungsdaten wurde den teilnehmenden Lehrkräfte in Microsoft Excel programmierte Dateien zur Verfügung gestellt, welche eine automatische grafische Veranschaulichung der eingetragenen Werte für jeden Schüler und für jedes CBM, auch im Vergleich zum Klassendurchschnitt sowie zu

den gemittelten Schülerergebnissen der Untersuchungsgruppe 2 (Region Nordvorpommern) ermöglichten (vgl. Abbildung 34). Es wurden zudem klare Regeln zur Einschätzung der Schülerdaten auf Basis ausgewählter Prozentranggrenzen (PR < 10, PR ≤ 25, PR ≥ 75, PR > 90) definiert (vgl. Tabelle 15). Mit diesen vorgegebenen Regeln zur Testinterpretation ist davon auszugehen, dass auch die Interpretationsobjektivität weitgehend gesichert ist. In der vorliegenden Studie wurde kein Mittelwert der Vergleichsgruppe in den Verlaufsdiagrammen ausgegeben, was eine Einteilung in einen oberen und einen unteren Durchschnittsbereich erschwerte. Es wird angenommen, dass viele Lehrkräfte auf Grundlage der Grafiken das Ziel verfolgten, schwache Kinder zu Leistungen entsprechend eines Prozentrangs kleiner bzw. gleich 25 zu verhelfen. Ziel sollte es jedoch sein, Schüler an den Mittelwert bzw. in den oberen Durchschnittsbereich zu führen. Die Ausgabe des Mittelwertes der Vergleichsgruppe sollte zukünftig Berücksichtigung finden.

Der hier als hoch eingeschätzte Grad an Objektivität wird durch Angaben Forschungsergebnisse von Clarke und Shinn (2004) bestärkt. Für die CBM number identification, missing number und quantity discrimination schätzten die Autoren Werte zur Interrater-Reliabilität bei der Auswertung von CBM im Bereich von r = .98 und r = .99. berichtet. Ähnlich hohe Werte berichten auch Christ et al. (2008) in ihrer Literatursynthese. Angesichts der Einfachheit der Auswertung von CBM-Verfahren und vor dem Hintergrund der dargelegten Interrater-Reliabilitäten erscheint die Annahme einer hohen Objektivität der hier erarbeiteten CBM als gerechtfertigt ($H_{1|1}$, $H_{1|2}$ und $H_{1|3}$).

9.2 Diskussion der Ergebnisse hinsichtlich der Reliabilität

Zur Einschätzung der Reliabilität wurden verschiedene Analysen durchgeführt (vgl. Abschnitt 7.2). In einem ersten Schritt wurden auf Basis der Methode der Testhalbierung Splithalf-Reliabilitätskoeffizienten für die Versionen der einzelnen CBM bestimmt (Abschnitt 7.2.1). Zusätzlich konnten Angaben zur Retest- sowie zur Paralleltest-Reliabilität getroffen werden (Abschnitt 7.2.2 und 7.2.3), dies allerdings nur eingeschränkt, für einige Parallelversionen der vorliegenden CBM.

9.2.1 Splithalf-Reliabilität

Die Bestimmung der Splithalf-Reliabilität für die CBM der Niveaustufe 1 erfolgte über die Teilung der Tests nach der Hälfte der Bearbeitungszeit, d. h. die Ergebnisse der Schüler jeweils nach einer halben Minute wurden korreliert. Für die hier untersuchten CBM der Niveaustufe 1 ergaben sich mittlere bis hohe Durchschnittswerte für die Splithalf-Reliabilität („Zahlenlesen im Zwanzigerraum“: r = .91; „Zahlenreihe im Zwanzigerraum“: r = .82; „Mengenvergleich im Zwanzigerraum“: r = .57; „Arithmetische Basiskompetenzen“: r = .64) (vgl. Abschnitt 7.2.1). Bis auf die Werte für die CBM „Mengenvergleich im Zwanzigerraum“ und „Arithmetische Basiskompetenzen“ fallen die Reliabilitätskoeffizienten damit hypothesenkonform aus.

Ergebnisse hinsichtlich der Splithalf-Reliabilität zu vergleichbaren US-amerikanischen Verfahren der CBM der Niveaustufe 1 sind nicht bekannt, daher wird zur Einordnung der hier ermittelten Werte auf Angaben zur internen Konsistenz auf Grundlage von Cronbachs α zurückgegriffen. Während der Reliabilitätskoeffizient des CBM „Zahlenlesen im Zwanzigerraum“ sich in zuvor ermittelte Befunde

einordnet (number identification: α = .87 - .92), fallen die errechneten Werte der weiteren CBM der Niveaustufe 1 im Vergleich zu bisher veröffentlichten Ergebnissen eher gering aus (missing number: α = .95 - .96; quantity discrimination: α = .88 -.94) (Lee, Lembke, Moore, Ginsberg & Pappas, 2007). Für die CBM der Website easyCBM.com führen Anderson et al. (2010) Splithalf-Reliabilität in einem Bereich von r = .66 bis r = .82 an. Bei diesen Angaben muss jedoch angemerkt werden, dass die Ermittlung der Splithalf-Reliabilität oft nicht aufgrund einer Teilung der Bearbeitungszeit, sondern nach Teilung der Items eines Tests ohne Begrenzung der Zeit vorgenommen wurde. Ebenso wurde bei der Analyse von Cronbachs α der Faktor Zeit außer Acht gelassen. Durch dieses unterschiedliche Vorgehen bleibt offen, inwieweit die innerhalb der vorliegenden Arbeit ermittelten Befunde mit denen von Anderson et al. (2010) und Lee et al. (2007) vergleichbar sind. Obwohl die Teilung nach der Zeit zur Bestimmung der Splithalf-Reliabilität für Speedtests wie CBM die angepasste Vorgehensweise ist (Schermelleh-Engel & Werner, 2012), so ist nicht ganz klar, inwieweit ein durch die Halbierung der schon ursprünglich sehr knappen Bearbeitungszeit von einer Minute aufkommender Leistungsdruck des Schülers zu Verzerrungen der Daten führt. Dies gilt insbesondere für Schüler einer ersten Klasse. In der vorliegenden Untersuchung wird eine daraus resultierte Unterschätzung der Splithalf-Reliabilität vermutet.

Die Homogenität der CBM der Niveaustufe 2 wird durch die vorgenommene stratifizierte Itemanordnung bei der Konzeption auf Grundlage der im Vorfeld getätigten Itemanalyse unterstellt. Bereits publizierte Forschungsarbeiten erhärten diese Annahme. So wurden für die in ihrer Struktur vergleichbaren Rechen-CBM von L.S. Fuchs et al. (MBSP, 1998) Koeffizienten hinsichtlich der internen Konsistenz in einem Bereich von r = .93 bis r = .98 nachgewiesen (L.S. Fuchs et al., 1989a; L.S. Fuchs, D. Fuchs, Hamlett, Thompson, Roberts, Kupek et al., 1994; Phillips et al., 1993). Auch Strathmann und Klauer

(2012) geben für die LVD-M 2-4 hohe Werte von r = .79 bis r = .92 hinsichtlich der Splithalf-Reliabilität an.

Die Ergebnisse zur Splithalf-Reliabilität der innerhalb dieser Arbeit konzipierten CBM vermitteln einen ersten Eindruck zur Zuverlässigkeit der erhobenen CBM-Daten. Dabei ergeben sich im Wesentlichen zufriedenstellende Koeffizienten. Eine präzise Einschätzung der Itemschwierigkeiten und damit eine präzise Einschätzung der Homogenität von Verfahren zur Lernverlaufsdokumentation ist nach Wilbert und Linnemann (2011) nur auf Basis der Item-Response-Theorie möglich. Eine weitere Möglichkeit beschreiben Strathmann und Klauer (2012). Dabei wird eine gleiche Schwierigkeit von CBM durch die zufällige Auswahl der Items aus einem Pool homogener und kontentvalider Aufgaben erreicht. Das Ergebnis sind zufallsparallele Verfahren, welche nach dem Binomialmodell analysiert werden können (Klauer, 1972). Die vorliegende Studie sollte demnach zukünftig um weitere Analysen zur Homogenität ergänzt werden.

9.2.2 Retest-Reliabilität

Zur Bestimmung der Retest-Reliabilität wurden Daten einzelner Versionen der hier vorgestellten CBM wiederholt erhoben. Für die CBM der Niveaustufe 1 wurde jeweils die zuerst eingesetzte Version zum ersten Messzeitpunkt erneut zum 11 Wochen späteren vierten Messzeitpunkt durchgeführt. Weitere Versionen des CBM „Arithmetische Basiskompetenzen“ wurden in zeitlichen Abständen von zwischen 5 und 24 Schulwochen wiederholt eingesetzt. Zwischen den Testungen mit Versionen der CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ lagen 12 bzw. 17 Schulwochen. Für die CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ „Arithmetische Basiskompetenzen“, „Zahlzerlegung im

Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ ergeben sich hypothesenkonform als mittel einzustufende Retest-Reliabilitäten im Bereich r > .5 und r < .7 (vgl. Abschnitt 7.2.2). Eine Ausnahme stellt die Version 2 des CBM „Subtraktion im Zwanzigerraum“ dar, hier ergab sich ein hoher Wert von r = .72. Für das CBM „Mengenvergleich im Zwanzigerraum“ konnte lediglich ein geringer Wert von r = .33 ermittelt werden. Das Forschungsdesign ermöglichte lediglich für eine begrenzte Anzahl der Parallelversionen eines jeden CBM die Untersuchung der Retest-Reliabilität. Inwieweit diese dann auch übergreifend für die restlichen Versionen des jeweiligen CBM gelten, bleibt unbeantwortet. Die Analyse der Retest-Reliabilität sollte durch zukünftige Forschung für die weiteren Versionen ergänzt werden.

Ordnet man die ermittelten Koeffizienten der hier entwickelten CBM in bereits veröffentlichte Forschungsergebnisse in den USA ein, fällt auf, dass die hier erzielten Werte durchweg geringer ausfallen. Clarke und Shinn (2004) berichten von hohen Retest-Reliabilitäten nach einem zeitlichen Abstand von 13 bzw. 26 Wochen für die CBM number identification (r = .85 bzw. r = .76), missing number (r = .79 bzw. r = .81) sowie quantity discrimination (r = .85 bzw. r = .86). Für die CBM zum Rechnen von Additions- und Subtraktionsaufgaben werden in der Literatur hohe bis sehr hohe Retest-Reliabilitätskoeffizienten von r = .70 bzw. r = 72 bei einem Abstand von einer Woche (Tindal, Germann & Deno, 1983) und r = .80 bzw. r = .85 bei einem Abstand von zwei Wochen (Epstein, Polloway & Patton, 1989) angegeben.

Warum die Retest-Reliabilitäten der innerhalb der vorliegenden Arbeit konzipierten CBM geringer ausfallen, kann nicht abschließend geklärt werden. Generell muss in diesem Zusammenhang jedoch festgehalten werden, dass die Analyse der Retest-Reliabilität bei der Veränderungsmessung eher schwierig ist, da dabei von der

Instabilität des zu messenden Merkmals ausgegangen wird (Bühner, 2006). Da nicht davon auszugehen ist, dass die Schüler über die verschiedenen Messzeitpunkte Lernverläufe in gleichem Maße zeigen, so ist an dieser Stelle auch klar, dass die innerhalb dieser Studie ermittelten Korrelationen der Ergebnisse in den CBM über die Messzeitpunkte abnehmen müssen. Dies deutet jedoch nicht automatisch auf eine ungenügende Reliabilität der Verfahren hin. Klauer (2011) merkt in diesem Zusammenhang an:

> *„Der Schwierigkeitsgrad nimmt theoretisch also im Fall des Lernens kontinuierlich ab. Strathmann und Klauer (2008; 2010) haben in dieser Situation den Ausweg gewählt, immer nur zwei direkt aufeinanderfolgende Tests auf homogene Schwierigkeit zu testen. Dabei muss man allerdings unterstellen, dass der Lernzuwachs in dieser Zeit vernachlässigbar gering ist“* (S. 210).

Strathmann und Klauer (2012) konkretisieren dieses Vorgehen durch ihre Forderung, Testergebnisse von CBM zu vergleichen, die maximal zwei Wochen auseinander liegen. Aufgrund der vorliegenden Datenlage dieser Untersuchung können jedoch keine in dieser vorgeschlagen zeitlichen Frequenz erhobenen Ergebnisse analysiert werden. Um dieses Problem zu lösen, müssen zusätzlich die einzelnen Splithalf- Reliabilitätskoeffizienten der Verfahren über die Messzeitpunkte analysiert werden. Sind bzw. bleiben diese Werte entsprechend hoch, kann von reliablen Verfahren ausgegangen werden (vgl. dazu Klauer, 2011). Vor diesem Hintergrund sind die ermittelten Werte als zufriedenstellend einzustufen.

9.2.3 Paralleltest-Reliabilität

Die Koeffizienten hinsichtlich der Paralleltest-Reliabilität wurden durch Korrelation der Ergebnisse zweier zeitgleich eingesetzter Versionen der CBM ermittelt. Für alle in dieser Arbeit erarbeiteten

CBM wurden dabei hohe durchschnittliche Reliabilitätskoeffizienten im Bereich von $r > .7$ bis $r \leq .9$ berechnet (vgl. Abschnitt 7.2.3). In den meisten Fällen ergeben sich zwischen den Verteilungen je zweier zeitgleich eingesetzter Parallelversionen signifikante Mittelwertunterschiede, welche mehrheitlich keine bzw. lediglich geringe, teilweise jedoch auch als mittel einzustufende Effektstärken aufweisen. Für das CBM „Mengenvergleich im Zwanzigerraum“ weisen die Unterschiede zum ersten Messzeitpunkt mit $d = 0.74$ auf einen als mittel einzustufenden Effekt hin. Auffällig ist, dass in vielen Fällen die Mittelwerte der jeweils zweiten eingesetzten Version bei jeder Erhebung etwas höher ausfielen. Dies kann als Hinweis auf einen Übungseffekt gedeutet werden. Im Laufe des Schuljahres dreht sich dieses Verhältnis bei den Daten der CBM „Arithmetische Basiskompetenzen“ und „Zahlzerlegung im Zwanzigerraum“, dann fallen die Mittelwerte der jeweils zuerst eingesetzten CBM-Version höher aus als die der zeitgleich eingesetzten zweiten. Besonders hohe Mittelwertunterschiede sind in den meisten Fällen jeweils beim ersten Einsatz eines CBM zu beobachten. Es scheint, als müssen die Kinder beim ersten Bearbeiten eines CBM erst einmal mit den an sie gestellten Anforderungen vertraut werden.

Nicht für jede mögliche Kombination zweier Parallelversionen der CBM-Verfahren konnte in der vorliegenden Studie die Paralleltest-Reliabilität ermittelt werden. Da die ermittelten Durchschnittswerte zur Paralleltest-Reliabilität durchweg hoch ausfallen, kann jedoch angenommen werden, dass die gewählte Verfahrensweise bei der Konzeption der CBM dazu führt, parallele Verfahren zu generieren. Für die CBM der Niveaustufe 1 wurden dazu Items aus dem sehr abgegrenzten Raum der Zahlen bis 20 gewählt. Durch die Enge des Zahlenraums bleibt nur eine begrenzte Auswahl an Items, gleichzeitig sinkt damit die mögliche Variabilität der Schwierigkeiten der erstellten CBM. Für die hier vorgestellten Verfahren der Niveaustufe 2 wurde

eine stratifizierte Anordnung der Items, wie sie in Anschnitt 3.3.2.2 beschrieben ist, vorgenommen.

Es ist festzuhalten, dass die Befunde der vorliegenden Studie zur Parallelität der CBM sich in die in der Literatur beschriebene Forschungslage einreihen. So berichten Lembke, Foegen, Whittaker & Hampton (2008) von Koeffizienten hinsichtlich der Parallelität der CBM number identification in einem Bereich von r = .77 bis r = .89 (hier zum Vergleich Zahlenlesen: r_{tt} = .90). Für das CBM missing number beschreiben die Autoren Werte von r = .69 bis r = .79 (hier zum Vergleich Zahlenreihe: r_{tt} = .85), für das CBM quantity discrimination Werte von r = .70 bis r = .85 (hier zum Vergleich Mengenvergleich: r_{tt} = .76). Auch die ermittelten Kennwerte der Paralleltest-Reliabilität der CBM „Addition im Zwanzigerraum“ (r_{tt} = .73) sowie „Subtraktion im Zwanzigerraum“ (r_{tt} = .77) fallen ähnlich hoch aus wie die von L.S. Fuchs et al. (1998) beschriebenen Werte für das MBSP computation (zwischen r = .73 und r = .93).

Für die CBM „Arithmetische Basiskompetenzen“ und „Zahlzerlegung im Zwanzigerraum“ gibt es keine entsprechenden Ergebnisse, mit denen die hier erhaltenen Befunde verglichen werden könnten. Für das CBM numbers and operation (hier vergleichbar: „Arithmetische Basiskompetenzen“) führen Alonzo, Anderson und Tindal (2009) ausführliche Itemstatistiken an, die auf die Parallelität der Verfahren hinweisen. Das CBM „Arithmetische Basiskompetenzen“ ist strukturell an dieses Verfahren angelehnt. Lediglich ursprünglich sprachgebundene Items wurden systematisch durch neue, sprachfreie Aufgabenformate ersetzt. Die Grundstruktur des Verfahrens hat sich demnach nicht verändert. Vor dieser Tatsache und dem hohen Durchschnittswert für die Paralleltest-Reliabilität für das CBM „Arithmetische Basiskompetenzen“ (r = .75) kann auch für dieses Verfahren Parallelität angenommen werden.

Um mögliche Übungseffekte abschätzen zu können, wäre es hilfreich gewesen, in der vorliegenden Arbeit ein sog. Cross-over-design (Winer, 1971) einzusetzen. Dabei hätte man die Klassen jeweils in zwei Gruppen aufgeteilt, welche zu jedem Messzeitpunkt die beiden durchzuführenden CBM-Versionen jeweils in einer anderen Reihenfolge bearbeitet hätten (Gruppe 1: erst Version 1, dann 2; Gruppe 2: erst Version 2, dann 1).

Insgesamt betrachtet sind die sich ergebenden Werte hinsichtlich der Reliabilität der hier geprüften CBM überwiegend hypothesenkonform und als zufriedenstellend einzuschätzen. Für das CBM „Mengenvergleich im Zwanzigerraum“ ergeben sich nur mittlere Werte hinsichtlich der Splithalf-Reliabilität und geringe Werte hinsichtlich der Retest-Reliabilität bei hoher Parallelität. Eine Überarbeitung des Verfahrens scheint angezeigt, um höhere Reliabilitäten zu erreichen, eine stratifizierte Itemanordnung erscheint dabei eine sinnvolle Methode zu sein.

9.3 Diskussion der Ergebnisse hinsichtlich der Validität

9.3.1 Inhaltsvalidität

Abhängig von der Ausprägung ihrer mathematischen Basiskompetenzen können sich Schüler weiterführende mathematische Kompetenzen aneignen. So stellt das erfolgreiche Erlernen des Rechnens ein Produkt verschiedener mathematischer Einsichten dar. Modellvorstellungen, die die Entwicklung dieser Einsichten systematisieren, wurden in Kapitel 2 vorgestellt. Analysiert man die in der vorliegenden Studie erarbeiteten CBM vor diesem theoretischen Hintergrund, lassen sich die in den Verfahren abgefragten Kompetenzen in mathematischen

Kompetenzentwicklungsmodellen wiederfinden. Zentrale Meilensteine in der Entwicklung mathematischer Kompetenzen bilden:

- das Mengen- und Zahlenwissen (z. T. abgefragt in den CBM „Zahlenlesen im Zwanzigerraum“ und „Arithmetische Basiskompetenzen“),
- die integrierte Einsicht über Zusammenhänge von Mengen und Zahlen (z. T. abgefragt in den CBM „Mengenvergleich im Zwanzigerraum“ und „Arithmetische Basiskompetenzen“),
- die Einsicht in das System der Zahlreihe (z. T. abgefragt in den CBM „Zahlenreihe im Zwanzigerraum“ und „Arithmetische Basiskompetenzen“),
- die Einsicht in das Teile-Ganzes-Konzept (z. T. abgefragt in den CBM und „Arithmetische Basiskompetenzen“ und „Zahlzerlegung im Zwanzigerraum“),
- das Verständnis zum Enthaltensein von Zahlen in anderen Zahlen (z. T. abgefragt in den CBM und „Arithmetische Basiskompetenzen“ und „Zahlzerlegung im Zwanzigerraum“) sowie
- dem relationalen Verständnis der Zahlen zueinander (z. T. abgefragt in den CBM und „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“).

Darüber hinaus wurde in Abschnitt 2.9 zur Einschätzung der curricularen Validität der CBM eine Analyse der Curricula der deutschen Bundesländer vorgenommen. Die in den hier erarbeiteten CBM abgefragten mathematischen Aspekte sind indirekt bzw. teilweise explizit in den Curricula der deutschen Bundesländer wiederzufinden. Die curriculare Einordnung der CBM erfolgte in Abschnitt 4.1 (vgl. Tabelle 8). Die Verfahren erscheinen vor diesem Hintergrund als curricular valide.

Insgesamt muss jedoch festgehalten werden, dass durch das Auszählen der richtig gelösten Aufgaben bei der zeitlich begrenzten Bearbeitung eines CBM nicht direkt das Vorhandensein mathematischer Konzepte gemessen wird. Jedoch kommt der Höhe der Geschwindigkeit bei der Bearbeitung mathematischer Aufgaben (fluency) eine große Bedeutung zu, so gibt es einen Zusammenhang zwischen der Fähigkeit des schnellen Faktenabrufs mathematischer Inhalte mit der erfolgreichen Aneignung weiterführender mathematischer Kompetenzen (Garnett & Fleischner, 1983; Geary, 2004; Gersten et al., 2005; Jordan et al., 2003; VanDerHeyden & Burns, 2008). Diesbezüglich konstatiert Roick (2007):

> *„Eine ausgesprochen hohe Relevanz hat auch der Zeitdruck bei der Bearbeitung von Faktenaufgaben. Deutliche Leistungsunterschiede zwischen rechenschwachen und ungestörten Kindern zeigen sich erst dann, wenn die Bearbeitung der Faktenaufgaben unter Zeitdruck erfolgt, insbesondere wenn es sich dabei um komplexe Faktenaufgaben handelt (Jordan & Montani, 1997; Roick & Hasselhorn, 2005)“* (S. 59).

Es ist einleuchtend, dass Kinder mit gering ausgeprägtem mathematischen Wissen, welche lediglich über zählende und kaum elaborierte Strategien zum Lösen von Rechenaufgaben verfügen, mehr Zeit bei der Bearbeitung benötigen. Dies müsste sich demnach auch in den CBM-Daten wiederspiegeln. Jedoch weisen VanDerHeyden und Burns (2008) darauf hin, dass Kinder, die sich in einer Aneignungsphase einer neuen Rechenstrategie befinden, zunächst auch erst einmal langsamer Arbeiten, was sich ebenso in seinen CBM-Daten zur Addition bzw. Subtraktion zeigen müsste. Auf Grundlage von CBM-Daten könnte demnach ein Kind mit gering ausgeprägten mathematischen Kompetenzen, welches die Items zählend löst mit einem solchen gleichgestellt werden, welches zwar über weiter entwickelte mathematische Kompetenzen verfügt, diese jedoch noch nicht sicher anwenden kann und entsprechend

langsamer arbeitet. Ebenso kann ein Schüler, der sehr sorgsam alle bearbeiteten Items richtig löst die gleiche Punktzahl erhalten wie ein Schüler, der sehr schnell sehr viele Aufgaben bearbeitet hat, dabei aber viele Fehler gemacht hat. Um eine differenziertere Einschätzung der mathematischen Einsichten zu erhalten, muss folglich neben Aspekten der Geschwindigkeit auch die Präzision erfasst werden, mit welcher die Schüler die in den CBM gestellten Testaufgaben bearbeiten. Bei der Datenauswertung müssten demzufolge die Angaben zu richtig gelösten und den insgesamt bearbeiteten Aufgaben ins Verhältnis gesetzt werden, ähnlich wie in etablierten Konzentrationstests (z. B. Differentieller Leistungstest – KG; E.W. Kleber, G. Kleber & Hans, 1999). Pospeschill (2010) beschreibt einen Wert, der Informationen über den Anteil richtiger und falscher Lösungen unter Berücksichtigung der Speed-Aspekte liefert. Dieser Wert wird durch folgende Formel definiert (Pospeschill, 2010, S. 74):

$$\frac{Anzahl\ richtiger\ Items}{Anzahl\ richtiger\ Items + Anzahl\ falscher\ Items + Anzahl\ der\ ausgelassenen\ Items} \cdot 100$$

Eine Auswertung der erhobenen Schülerdaten auf Grundlage eines solchen Wertes war innerhalb der vorliegenden Studie nicht möglich. Die verfügbare Datenlage umfasste lediglich die von den Lehrern dokumentierten Anzahlen richtig gelöster Aufgaben. Angaben zur Fehleranzahl oder zur Menge von Aufgaben, welche die Schüler nicht bearbeitet haben, lagen nicht vor.

Dass anhand von CBM keine allgemeine mathematische Kompetenz gemessen werden kann, wird darüber hinaus deutlich, wenn man bedenkt, dass durch die hier vorliegenden CBM lediglich einige arithmetische Aspekte abgefragt werden. In der vorliegenden Studie wurde der Fokus auf CBM zur Identifikation von Zahlen, zur Seriation von Zahlen, zum Vergleich zweier Zahlen aufgrund ihrer Mächtigkeit, zur Zahlzerlegung sowie zum Rechnen beschränkt. Zukünftig bleibt

zu prüfen, inwieweit CBM konzipiert werden können, die weitere Aspekte abfragen wie z. B. die Zählkompetenzen (counting; Chard et al., 2005; Clarke & Shinn, 2004; VanDerHeyden et al., 2004; VanDerHeyden et al., 2001), die Fähigkeit zur Simultanerfassung von Mengen (subitizing; Chi & Klahr, 1975; Starkey & Cooper, 1995) oder auch die Fähigkeit zum Schätzen von Mengen (approximate number system; Mazzocco, Feigenson & Halberda, 2011). Eine Reihe von CBM zur Erfassung mathematischer Basiskompetenzen im Sinne des number sense wurde von Methe, Begeny und Leary (2011) konzipiert, implementiert und evaluiert. Wenn auch zunächst als vorläufig anzusehen, so kommen die Autoren zu vielversprechenden Ergebnissen bezüglich der wissenschaftlichen Güte verschiedener Verfahren. Zufriedenstellende Werte zur Reliabilität, Validität und Sensibilität ergaben sich für die CBM equal partitioning (eine vorgegebene Menge Punkte soll in gleiche Teile zerlegt werden bzw. es soll geprüft werden, ob eine Aufteilung in gleiche Teile korrekt durchgeführt wurde), ordinality (es soll der bspw. fünfte abgebildete Gegenstand benannt werden bzw. es soll die Stelle eines vorgegebenen Gegenstandes benannt werden), group by five (die Anzahl von Fünferbündeln soll bestimmt werden, anschließen die Gesamtanzahl) sowie verbal facts (einfache Kopfrechenaufgaben sollen gelöst werden) (Methe et al., 2011).

Festzuhalten ist zudem, dass weitere mathematische Inhalte wie bspw. die Geometrie durch die vorliegenden CBM nicht berücksichtigt werden. Auch die Ausprägung allgemeiner mathematischer Kompetenzen der Schüler, wie sie in den Bildungsstandards (Kultusministerkonferenz, 2004b) verordnet sind, können nicht durch die hier erarbeiteten CBM erfasst werden. Zur Diagnostik mathematischer Kompetenzen können die hier beschriebenen CBM somit nicht als hinreichend angesehen. Um einen differenzierteren Einblick in Ressourcen und Problemfelder eines Schülers zu erhalten, müssen durch CBM erhobene Schülerdaten durch weitere

diagnostische Maßnahmen flankiert werden. Dazu wurde in Abschnitt 3 eine Kombinationen diagnostischer Maßnahmen nach dem NCRTI (2010) bzw. nach Moser Opitz (2009; vgl. Abbildung 9) vorgeschlagenen.

Ausgehend von den geschilderten Überlegungen ist grundlegend festzuhalten, dass die in dieser Studie erarbeiteten CBM-Verfahren, sofern sie kombiniert über das erste Schuljahr eingesetzt werden, eine angemessene Repräsentation des theoretischen Merkmals erster arithmetischer Kompetenzen von Erstklässlern darstellen. Es handelt sich bei den CBM somit im Wesentlichen um curricular sowie kontentvalide Verfahren.

9.3.2 Konstruktvalidität

Durch Analysen von Multitrait-Multimethod-Matrizen konnte für alle innerhalb der vorliegenden Arbeit konzipierten CBM die Konstruktvalidität im Wesentlichen nachgewiesen werden. Im Zusammenhang mit Außenkriterien, welche die mathematischen Kompetenzen erfassten (Kalkulie; Fritz et al., 2007 bzw. DEMAT 1+; Krajewski et al., 2002), wurden mittlere bis fast mittlere Korrelationen errechnet (ausführlicher in Abschnitt 9.3.3), welche als empirische Evidenz für die konvergente Validität der CBM interpretiert werden können. Mit Außenkriterien, welche schriftsprachliche Aspekte fokussierten (MÜSC, WLLP bzw. CBM „Silbenlesen“) wurden erwartungskonform geringere Korrelationen ermittelt, ein Indiz für die diskriminante Validität der CBM. Für die CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“, „Mengenvergleich im Zwanzigerraum“ und „Arithmetische Basiskompetenzen“ ist einschränkend festzuhalten, dass wider der Erwartungen die Korrelationskoeffizienten zwischen den CBM und dem Münsteraner Screening ($r = -.35$, $r = -.30$, $r = -.35$ bzw. $r = -.22$) betragsmäßig höher ausfielen als die Korrelationskoeffizienten

zwischen den CBM und dem CBM „Silbenlesen“ (r = .29, r = .19, r = .17 bzw. r = .11). Beachtet werden muss an dieser Stelle, dass beim MÜSC Risikopunkte dokumentiert werden. Ein hoher Wert im MÜSC deutet demnach auf ein nur eingeschränkt ausgeprägtes schriftsprachliches Vorwissen hin. Der hier festgehaltene negative Zusammenhang muss demnach invers interpretiert werden. Diese Befunde decken sich mit den Ergebnissen von Krajewski (2003), die einen Zusammenhang zwischen phonologischen Fähigkeiten (als eine wesentliche Komponente im Schriftspracherwerb) und ersten mathematischen Kompetenzen festhalten konnte. So weist die Autorin darauf hin, dass der Grad der Ausprägung der phonologischen Bewusstheit im weiteren Sinn Auswirkungen auf basale Zahlenkompetenzen der Kinder hat. Dies wird durch ein weiteres Ergebnis der Analysen unterstützt, wonach mit r = -.48 ein mittlerer, entgegen gerichteter Zusammenhang zwischen dem Teil 1 des Diagnoseprogramms Kalkulie (Fritz et al., 2007) und dem Münsteraner Screening (Mannhaupt, 2006) belegt wurde. Die Annahme von gegebener Konstruktvalidität scheint vor diesem Hintergrund für die CBM der Niveaustufe 1 gerechtfertigt.

Die Interkorrelationen zwischen den CBM der Niveaustufe 1 fallen über die Messzeitpunkte hinweg eher gering aus (r = .31 bis r = .56). Es scheint, dass diese Verfahren erwartungsgemäß jeweils unterschiedliche Aspekte mathematischer Kompetenzen abfragen. Diese Befunde decken sich ebenso mit den Ergebnissen von Missall, Mercer, Martinez und Casebeer (2012). Sie berechneten Korrelationen für die CBM number identification und missing number in einem Bereich von r = .25 bis r = .50. Werte, die den Zusammenhang zwischen den CBM number identification und quantity discrimination beschreiben, fielen in einem Intervall von r = .16 bis r = .53 aus. Für die CBM missing number und quantity discrimination ergab sich Werte im Bereich von r = .22 bis r = .52 (Missall et al., 2012).

Die CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ weisen durchweg mittlere Korrelationen auf. Dies deutet auf einen höheren Zusammenhang zwischen den getesteten Kompetenzen hin. Dies lässt sich theoretisch hinreichend begründen. So erfassen diese CBM das Teile-Ganzes-Konzept und Aspekte des darauf aufbauenden relationalen Zahlbegriffs entsprechend der Entwicklungsstufen vier und fünf des Entwicklungsmodells mathematischer Kompetenzentwicklung von Fritz et al. (2007) (vgl. Abschnitt 2.5) bzw. der Stufe 3 des Entwicklungsmodells früher mathematischer Kompetenzen von Krajewski und Schneider (2006) (vgl. Abschnitt 2.4). Beide mathematische Einsichten stehen in einem engen inhaltlichen Zusammenhang zueinander und bilden die Grundlage für das Erlernen des sicheren und einsichtigen Rechnens. „Auf dieser Basis [des Teile-Ganzes-Konzepts] entsteht die Einsicht in die Umkehrbarkeit von Additions- und Subtraktionsaufgaben. Mehr noch, sämtliche Aufgaben des Typs a +/- b = ?, a +/- ? = c und ? +/- b = c werden für die Kinder gleich schwierig […]“ (Fritz et al., 2007, S. 13). Entsprechend sind die höheren Interkorrelationen der CBM der Niveaustufe 2 nachvollziehbar.

Das CBM „Arithmetische Basiskompetenzen“ wies durchweg geringe Korrelationen mit den anderen CBM-Verfahren auf. Dies kann ein Hinweis darauf sein, dass durch die Verfahren jeweils unterschiedliche Fähigkeiten abgefragt werden. Es liegt jedoch auch nahe, dass die geringen Korrelationen mit der unterschiedlichen Konzeption des CBM „Arithmetische Basiskompetenzen“ im Vergleich zu den anderen hier erarbeiteten Verfahren zusammenhängen. Das CBM umfasst mehrere Aufgabenformate, welche jeweils durch Ankreuzen der richtigen aus drei vorgegebenen Antwortalternativen (multiple choice) gelöst werden sollen. Dieses Aufgabenformat weicht stark vom Aufbau der weiteren hier beschrieben CBM ab. Zudem werden im CBM „Arithmetische

Basiskompetenzen“, entgegen der weiteren hier erarbeiteten CBM, verschiedene mathematische Kompetenzen aggregiert innerhalb eines CBM abgefragt. Der Wechsel der inhaltlichen Anforderungen, vor allem unter zeitlichem Druck, kann den Kindern besondere Schwierigkeiten bereitet haben.

9.3.3 Kriteriumsvalidität

Die Einschätzung des Grades der Kriteriumsvalidität erfolgte über die Analysen der Übereinstimmungs- der prognostischen sowie der inkrementellen Validität der CBM. Zur Bestimmung der Übereinstimmungsvalidität wurden Korrelationen der Ergebnisse der CBM-Daten mit verschiedenen Kriterien vorgenommen. Für die CBM der Niveaustufe 1 wurden Zusammenhänge mit den Ergebnissen des Diagnoseverfahrens Kalkulie (Fritz et al., 2007), für die Verfahren der Niveaustufe 2 Zusammenhänge mit dem DEMAT 1+ (Krajewski et al., 2002) und für die Verfahren beider Niveaustufen Zusammenhänge mit den Ergebnissen des Inventars Rechenfische (Wagner & Hartke, 2006) berechnet. Es ergaben sich geringe bis mittlere Korrelationen in einem Bereich zwischen $r = .41$ und $r = .70$. Diese Werte entsprechen etwa den bei Clarke und Shinn (2004) berichteten Koeffizienten (CBM number identification: $r = .63$ bis $r = .70$; CBM missing number: $r = .68$ bis $r = .74$; CBM quantity discrimination: $r = .71$ bis $r = .80$). Die Autoren nutzten dabei als Kriterien einen Untertest des Woodcock-Johnson (Woodcock & Johnson, 1989), bei dem verschiedene mathematische Rechenprobleme innerhalb einer Individual-Testung gelöst werden sollen (Klasse1: Addition und Subtraktion) sowie den Number Knowledge Test (NKT; Okamoto & Case, 1996), welcher neben Zählkompetenzen auch Kompetenzen zur Berechnung von einfachen und schweren Additions- und Subtraktionsaufgaben sowie zur zahlenmäßigen Benennung des Abstandes zwischen zwei Zahlen erfasst. In einer Studie von Chard et al. (2005) wurden ebenfalls mittlere Korrelationen in einem Intervall

von r = .45 bis r = .69 zu den CBM number identification, missing number sowie quantity discriminiation angegeben. Auch in dieser Studie diente der NKT (Okamoto & Case, 1996) als Kriterium.

Die im Rahmen der vorliegenden Studie konzipierten Rechen-CBM (CBM „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“) weisen mittlere Korrelationen mit den Kriteriumsmaßen auf. Zur Einordnung in bereits vorhandene Ergebnisse zur Kriteriumsvalidität der Rechen-CBM steht neben US-amerikanischen Veröffentlichungen auch eine deutsche Publikation (Strathmann & Klauer, 2012) zur Verfügung. Bezüglich CBM zum Rechnen für die Klassenstufe 2 bis 6 konnten L.S. Fuchs et al. (1998) für das MBSP computation mittlere bis sehr hohe Korrelationen im Bereich von r = .55 bis r = .93 mit zwei verschiedenen standardisierten Mathematiktests ermitteln. Ebenso reihen sich die innerhalb der vorliegenden Studie erhobenen Daten in die Ergebnisse der von Strathmann und Klauer (2012) entwickelten LVD-M 2-4 ein. Die Autoren berichten mittlere bis hohe Korrelationen mit Verfahren der DEMAT-Serie (DEMAT 2+, Krajewski et al., 2004; DEMAT 3+, Roick et al., 2004; DEMAT 4+, Gölitz et al., 2006) im Bereich von r = .53 bis r = .80.

Dass in der vorliegenden Studie keine als hoch einzuschätzenden Korrelationen gefunden werden konnten, begründet sich höchstwahrscheinlich durch die unterschiedlichen Diagnoseintentionen der gewählten Außenkriterien. Während das Diagnoseprogramm Kalkulie (Fritz et al., 2007) insbesondere das Vorhandensein zentraler mathematischer Einsichten und Konzepte (z. B. Teile-Ganzes-Konzept) im Sinne einer Erfassung des mathematischen Vorwissen fokussiert, dient der DEMAT 1+ (Krajewski et al., 2002) darüber hinaus eher als Maß für die Anwendbarkeit vorhandener mathematischer Kompetenzen bezüglich vergleichsweise komplexerer Aufgabenstellungen (z. B.

Lösen von Additionsaufgaben). Innerhalb beider Verfahren haben die Kinder jedoch genügend Zeit, um zumindest einen Großteil der Items zu lösen. Die CBM hingegen fokussieren durch die Einschränkung der Bearbeitungsdauer eher auf die Schnelligkeit bei der Bearbeitung der Aufgaben bzw. die Rechenflüssigkeit (fluency). Auch die Befunde von Foegen et al. (2007) zeigen, dass – entgegen der hohen Übereinstimmungsvaliditäten von Lese-CBM mit standardisierten Lesetests – CBM für den Bereich Mathematik eher mäßige Zusammenhänge mit Mathematiktests aufweisen. Ein vergleichbar guter robuster Indikator, wie es die Leseflüssigkeit für das erfolgreiche Leselernen ist, wurde im Bereich Mathematik mit den bisher entwickelten CBM noch nicht gefunden.

Zur Überprüfung der prognostischen Validität wurden die mathematischen Leistungen zum Ende der ersten Klasse der untersuchten Schüler analysiert. Dabei wurden die Kinder auf Grundlage der erhobenen CBM-Daten als Risiko- bzw. Nichtrisikokinder klassifiziert. Als Cutoff-Wert wurde eine Leistung entsprechend einem Prozentrang kleiner bzw. gleich 25 gewählt. Insgesamt ergeben sich für die überprüften CBM zu allen über das Schuljahr verteilten Messzeitpunkte ausschließlich gute bis sehr gute Ratz-Indices von 35% bis 95% (vgl. Abschnitt 7.3.3.2).

Anzumerken ist an dieser Stelle, dass die ermittelten Fehlerquoten (α- und β-Fehler) relativ hoch ausfallen. Es besteht jedoch die Vermutung, dass der Anteil der falsch positiv klassifizierten Kinder (α-Fehlerrate) durch studiendesignbedingte Faktoren überschätzt wird. Die Lehrkräfte der untersuchten Rügener Kinder (UG 1) erhielten regelmäßig anhand der CBM-Ergebnisse Rückmeldung über die Leistungsstände ihrer Schüler und erteilten auf dieser Grundlage jeweils spezifische Förderung entsprechend des Konzepts des RIM. Ausgehend davon, dass diese Fördermaßnahmen bei vielen Risikokindern Wirkung zeigte, wird der Anteil der falsch positiv

klassifizierten Kinder innerhalb dieser Studie zu hoch eingeschätzt. Die hier präsentierten Ergebnisse zur prognostischen Validität gelten demnach ausschließlich in Zusammenhang mit den im RIM verankerten Fördermaßnahmen. Eine Einschätzung der prognostischen Validität ohne weitere äußere Einflüsse ist auf Basis der vorliegenden Datenlage nicht möglich und sollte Gegenstand von Folgestudien sein.

Auch Clarke und Shinn (2004) konnten in ihrer Untersuchung die prognostische Güte der CBM number identification, missing number und quantity discrimination nachweisen. Kriterium dabei war der etwa drei und sechs Monate später eingesetzte Untertest zum Rechnen des Woodcock-Johnson (Woodcock & Johnson, 1989). Es ergaben sich knapp hohe bis hohe Korrelationskoeffizienten (Clarke & Shinn, 2004). Das CBM „Arithmetische Basiskompetenzen“ scheint in seiner Ursprungsform (numbers and oprations; Anderson et al., 2010) eine hinreichende prognostische Validität zu erreichen. Da es sich bei den dort berichteten Angaben jedoch um aggregierte Daten verschiedener Mathematik-CBM handelt, kann die prognostische Güte des CBM numbers and operations nicht genau bestimmt werden. Für Rechen-CBM (MBSP computation, L.S. Fuchs et al., 1998) berichten Jitendra, Sczesniak und Deatline-Buchman (2005) eine mittlere Korrelation von r = .59 mit einem standardisierten Testverfahren zur Erfassung der Rechenleistung.

Abgesehen von der Überschätzung der nach oben verzerrten α-Fehlerrate reihen sich die Erkenntnisse in die Ergebnisse von Blumenthal (2012) zur Vorhersagekraft von Testverfahren zum bereichsspezifischen Vorwissen und zur Intelligenz auf spätere schulische Minderleistungen ein. Auch dort zeigt sich, dass die untersuchten Verfahren sowohl einzeln als auch in kombinierter Betrachtung keine hinreichend validen langfristigen Prognosen von schulischen Minderleistungen ermöglichen können.

Die hier vorliegenden Analysen zur Bestimmung der prognostischen Validität werden durch verschiedene Aspekte limitiert. Zunächst muss kritisch angemerkt werden, dass Kinder, mit einem Ausgangsniveau nahe des Cutoff-Wertes sich nur unwesentlich zu verbessern brauchen, um als unauffällig eingestuft zu werden. Zudem unterliegen wiederholte Analysen innerhalb von Probandengruppen mit extremen Merkmalsausprägungen der Gefahr, durch einen Regressionseffekt verzerrt zu werden (Nachtigall & Suhl, 2002). Beide Aspekte führen wiederum zu einer Überschätzung der Rate der falsch-positiv klassifizierten Schüler. Weiterhin stellt sich die Frage, inwieweit die Festlegung des Cutoff-Wertes bei einem CBM-Ergebnis entsprechend eines Prozentrangs kleiner bzw. gleich 25 sinnvoll ist. Zwar wird in der vorliegenden Studie davon ausgegangen, dass das untere Leistungsquartil Unterstützung durch spezifische Fördermaßnahmen benötigt, fraglich ist jedoch, ob diese Risikogrenze wirklich in Hinblick auf die Vorhersagegüte der CBM sinnvoll gesetzt ist, oder ob durch eine andere Festlegung der prädiktive Gehalt optimiert werden kann. Mittels sog. Receiver-Operating-Characteristic-Analysen (ROC-Analysen), einer statistischen Methode, welche ursprünglich in der Sendetechnik Anwendung fand, können unter Berücksichtigung des Verhältnisses der Rate der Falsch-Positiven (α-Fehlerquote) und der Richtig-Negativen (Sensitivität) optimale Cutoff-Werte hinsichtlich einer möglichst geringen Fehlerquote bestimmt werden (Swets, Dawes & Monahan, 2000). Obwohl ROC-Analysen einen wesentlichen Teil dazu beitragen können, präzisere Klassifikationen zu ermöglichen, werden sie (auch bei der Testevaluation) noch recht selten durchgeführt (Swets et al., 2000). In zukünftigen Untersuchungen sollte versucht werden, die prädiktive Güte von CBM auf Basis der ROC-Analysen noch weiter zu erhöhen.

Mittels der Regressionsanalysen zur inkrementellen Validität der CBM zur Vorhersage der Mathematikleistungen zum Ende der ersten

bzw. zweiten Klasse stellte sich heraus, dass das CBM „Arithmetische Basiskompetenzen“ nicht signifikant zur Prognose der Leistungen zum Ende der zweiten Klasse und das CBM „Subtraktion im Zwanzigerraum“ nicht signifikant zur Prognose der Leistungen zum Ende der ersten und zweiten Klasse beitragen kann (vgl. Abschnitt 7.3.3.3). In diesem Kontext muss erwähnt werden, dass jeweils nur die Daten des ersten Einsatzes der CBM (in der fünften bzw. 21. Schulwoche) in die Regressionsmodelle aufgenommen wurden. Diese Ergebnisse decken sich ebenfalls mit den Erkenntnissen von Blumenthal (2012), wonach Prognosen mit zunehmender Länge des vorherzusagenden Zeitraums ungenauer ausfallen. Gleiches wurde in der vorliegenden Studie für diese beiden CBM ermittelt, jedoch auch für alle weiteren innerhalb dieser Untersuchung konzipierten Verfahren.

9.4 Diskussion der Ergebnisse hinsichtlich der Änderungssensibilität

Die in den Mehrebenenanalysen festgehaltenen signifikanten Verschiebungen der Mittelwerte über die Zeit (Tabelle 77, Tabelle 81, Tabelle 85, Tabelle 89, Tabelle 93, Tabelle 97 sowie Tabelle 101), können als Anzeichen dafür gesehen werden, dass mit den vorliegenden CBM monatliche Veränderungen in den Ausprägungen arithmetischer Kompetenzen im Mathematikunterricht der ersten Klasse hinweg erfasst werden können. Ebenso konnte auf Grundlage der Analysen gezeigt werden, dass die Schüler mit unterschiedlichen mathematischen Lernvoraussetzungen in die Schule kommen. Dies spiegelte sich in der jeweils hohen Varianz der intercepts der CBM wider. Die Analyse der Interkorrelationen über die Messzeitpunkte hinweg zeigt, dass die ermittelten Korrelationskoeffizienten für alle CBM mit Ausnahme des CBM „Zahlzerlegung im Zwanzigerraum“ einer sog. Simplexstruktur ähneln (vgl. Abschnitte 7.4.1, 7.4.2, 7.4.3, 7.4.4, 7.4.5, 7.4.6 sowie 7.4.7). D. h. mit zunehmendem zeitlichem

Abstand zwischen den Erhebungszeitpunkten konnte bei den CBM ein Abfall der Korrelationskoeffizienten verzeichnet werden. Humphreys (1960) weist darauf hin, dass derartige Simplexstrukturen bei Reifungs- bzw. Lernprozessen zu erwarten sind, d. h. bei Prozessen, in welchen ein Fortschritt beobachtet bzw. gemessen wird. Ursachen für eine solche Struktur sehen Lin und Humphreys (1977) entweder in systematischen Veränderungen im schulischen Bezugssystem oder aber in systematischen Veränderungen innerhalb des Schülers. Da die Wahrscheinlichkeit systematischer Veränderungen im Schulsystem bei der Größe der untersuchten Gruppe eher gering ausfällt, kann die Schlussfolgerung gezogen werden, dass durch die CBM tatsächliche Lernverläufe der Kinder erhoben wurden. Gestützt wird diese Annahme dadurch, dass sich zwischen verschiedenen Schülergruppen, welche aufgrund der Ausprägung ihres mathematischen Vorwissens zu Beginn der ersten Klasse gebildet wurden, signifikante Unterschiede der Lernverläufe für die CBM „Zahlenlesen im Zwanzigerraum“, „Mengenvergleich im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ zeigten. So erzielten Kinder mit einem Risikostatus im mathematischen Vorwissen zu Beginn der ersten Klasse geringere Lernzuwächse als Kinder ohne einen solchen Risikostatus.

Einzig der Nachweis des signifikanten Anstiegs der einzelnen CBM-Mittelwerte über die Zeit reicht jedoch nicht aus, um einen Beleg für die Änderungssensibilität zu erbringen. Ein an die hier entwickelten CBM angelegtes Kriterium war, dass durchschnittliche Lernzuwächse von zumindest einem Punkt von Messzeitpunkt zu Messzeitpunkt (etwa vier Schulwochen Abstand) ermittelt werden konnten. Dies entspricht einem wöchentlichen (bezogen auf Schulwochen) Durchschnittsanstieg von ca. 0.25 Punkten. Dies sollte für Schüler ohne als auch mit festgestelltem Risikostatus hinsichtlich ihres mathematischen Vorwissens gelten. Ist dies nicht der Fall, sind kurzfristige pädagogische Entscheidungen zur Notwendigkeit und

Gestaltung von Fördermaßnahmen in der Praxis nur schwer möglich. Die dafür erforderliche Datengrundlage für diese Entscheidungen wäre dann frühestens nach zwei Monaten gegeben. Für Risikokinder geht damit wertvolle Zeit zur Förderung unzureichend ausgeprägter Kompetenzen verloren (Clarke & Shinn, 2004). Für die CBM „Arithmetische Basiskompetenzen“, „Zahlzerlegung im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ fallen die ermittelten wöchentlichen Anstiege für Kinder mit Risiken hinsichtlich ihres mathematischen Vorwissens zu Beginn der ersten Klasse geringer als 0.25 Punkte aus, bei den restlichen CBM liegen die durchschnittlichen Anstiege zwischen 0.26 und 1.25 (Tabelle 77, Tabelle 81, Tabelle 85, Tabelle 89, Tabelle 93, Tabelle 97 sowie Tabelle 101). Da für diese CBM beim Einsatz im zweiten Halbjahr der Klassenstufe 1 jedoch eine hohe prognostische Validität nachgewiesen wurde, empfiehlt sich der Einsatz im Sinne eines Screenings bzw. für die längerfristige Dokumentation von Lernverläufen.

Ordnet man diese durchschnittlichen wöchentlichen Anstiege in bereits publizierte Studien aus den USA ein (vgl. Tabelle 103), ist festzustellen, dass es jeweils in Abhängigkeit vom betrachteten CBM deutliche Unterschiede zwischen den ermittelten und berichteten durchschnittlichen wöchentlichen Leistungsanstiegen gibt. Dabei fällt auf, dass ein Teil der US-amerikanischen CBM die Forderung nach einem durchschnittlichen monatlichen Anstieg von zumindest einem Punkt nicht erfüllen. Besonders auffällig sind die großen Unterschiede bei den CBM „Zahlenlesen im Zwanzigerraum“ (number identification), „Zahlenreihe im Zwanzigerraum“ (missing number) sowie „Mengenvergleich im Zwanzigerraum“ (quantity discrimination). Hier liegen die in der vorliegenden Arbeit ermittelten Koeffizienten deutlich über den in der Fachliteratur genannten Werten. Für die CBM „Arithmetische Basiskompetenzen“ (numbers and operations), „Addition im Zwanzigerraum“ (computation) sowie

„Subtraktion im Zwanzigerraum“ (computation) wurden in der vorliegenden Studie durchweg geringere, jedoch annähernd gleiche durchschnittliche Koeffizienten zum wöchentlichen Leistungsanstieg ermittelt. Für die CBM des Typs computation muss dabei angemerkt werden, dass es sich hier um Verfahren handelt, die Additions- und Subtraktionsaufgaben kombiniert abfragen. Da in den CBM der vorliegenden Studie Additions- und Subtraktionsaufgaben diskret voneinander erfasst werden, ist eine direkte Vergleichbarkeit der ermittelten Koeffizienten in diesem Fall nicht gegeben. Für das CBM „Zahlzerlegung im Zwanzigerraum“ gibt es derzeit in der Fachliteratur kein vergleichbares Verfahren, sodass eine Einordnung der Ergebnisse nicht möglich ist.

Tabelle 103: Einordnung ermittelter durchschnittlicher Anstiege je Woche in bereits veröffentlichte Referenzwerte (AIMSweb.com, 2012; Clarke & Shinn, 2004; easyCBM.com, 2012b; L.S. Fuchs & D. Fuchs, Hamlett, Walz & Germann, 1993; Lembke et al., 2008; Lembke & Foegen, 2009; Shapiro et al., 2005)

CBM	Durchschnittlicher wöchentlicher Leistungsanstieg	
	In vorliegender Arbeit ermittelte Anstiege	Angaben in Studien oder auf US-amerikanischen Servern
Zahlenlesen im Zwanzigerraum (number identification)	oR: 1.25 mR: 0.94	0.24 (Lembke et al., 2008) 0.25 (Lembke & Foegen, 2009) 0.47 (Clarke & Shinn, 2004) 0.64 (AIMSweb.com, 2012)
Zahlenreihe im Zwanzigerraum (missing number)	oR & mR: 0.59	0.11 (Lembke et al., 2008) 0.03 (Lembke & Foegen, 2009) 0.23 (Clarke & Shinn, 2004) 0.22 (AIMSweb.com, 2012)
Mengenvergleich im Zwanzigerraum (quantity discrimination)	oR: 0.67 mR: 0.38	0.12 (Lembke et al., 2008) 0.12 (Lembke & Foegen, 2009) 0.36 (Clarke & Shinn, 2004) 0.36 (AIMSweb.com, 2012)
Arithmetische Basiskompetenzen (numbers and operations)	oR & mR: 0.15	0.21 (uQ: 0.07; mQ: 0.17-0.22; oQ: 0.33)* (easyCBM.com, 2012b)
Zahlzerlegung im Zwanzigerraum	oR & mR: 0.21	-
Addition im Zwanzigerraum (computation)	oR & mR: 0.26	0.38 (Shapiro et al., 2005) 0.26 – 0.50 (L.S. Fuchs et al., 1993) 0.69 (AIMSweb.com, 2012)
Subtraktion im Zwanzigerraum (computation)	oR: 0.24 mR: 0.17	0.38 (Shapiro et al., 2005) 0.26 – 0.50 (L.S. Fuchs et al., 1993)

Erläuterungen: oR – gültig für Kinder ohne festgestelltes Risiko hinsichtlich des mathematischen Vorwissens zu Beginn der ersten Klasse; mR – gültig für Kinder mit festgestelltem Risiko hinsichtlich des mathematischen Vorwissens zu Beginn der ersten Klasse; uQ – unteres Quartil; mQ – mittlere Quartile; oQ – oberes Quartil; * - Angaben unter: http://www.rti4success.org/tools_charts/popups_progress/programPopup.php?url=easyCBMMath&title=easyCBM&area=Math

Grundlegend für die hier ermittelten durchschnittlichen Anstiege pro Woche ist die Annahme des linearen Verlaufs der Lernverlaufskurven. Fraglich ist derzeit jedoch, ob diese Annahme auch tatsächlich für alle hier dargestellten Verfahren gilt. Für den Bereich des Rechnens (Addition, Subtraktion, Multiplikation und

Division) in den Klassenstufen 3 und 4 konnten Shin et al. (2004) festhalten, dass ein quadratischer im Vergleich zu einem linearen Verlauf, die anhand von CBM erhobenen Schülerdaten am besten beschreibt. Auch Lembke et al. (2008) wiesen nach, dass die Ergebnisse der CBM missing number (äquivalent zu „Zahlenreihe im Zwanzigerraum“) sowie quantity discrimination (äquivalent zu „Mengenvergleich im Zwanzigerraum“) am besten durch Polynome dritten Grades beschrieben werden konnten. Hingegen zeigte sich für das CBM number identification (äquivalent zu „Zahlenlesen im Zwanzigerraum“), dass die regelmäßig erfassten Schülerleistungen eher einen linearen Anstieg aufzeigten. Wie aber sind solche überlinearen Verläufe zu erklären? Auf diese Frage bieten Lembke et al. (2008) zwei mögliche Antworten. So könnten überlineare Anstiege einerseits durch messfehlerbehaftete Verfahren zustande kommen. Dies würde darauf hinweisen, dass sich die untersuchten CBM nicht zur Dokumentation des Lernverlaufs eignen. Andererseits ist vorstellbar, dass der Lernverlauf der mathematischen Kompetenzen nicht kontinuierlich linear, sondern eher phasenhaft verläuft. Obwohl letztere Begründung für mathematisches Lernen plausibel erscheint, ist die tatsächliche Ursache derzeit noch nicht abschließend geklärt. Im Bereich des Lesens wurde die Annahme eines linearen Lernverlaufs durch verschiedene Forschungsbemühungen abgewiesen (Christ, Silberglitt, Yeo & Cormier, 2010; Nese, Biancarosa, Anderson, Lai, Alonzo & Tindal, 2012; Kamata, Nese, Patarapichayatham & Lai, 2012). Inwieweit diese Befunde auf die mathematische Domäne übertragbar sind, ist derzeit noch offen. Die in der Literatur dargestellten Ergebnisse geben lediglich erste Hinweise und müssen in Folgeuntersuchungen empirisch abgesichert werden. Ausgehend von einem nichtlinearen Lernverlauf der Schüler sind Anpassungen bestehender Bewertungsmaßstäbe von CBM notwendig. Die in Abschnitt 3.2.3 beschriebenen Bewertungsmöglichkeiten von CBM-Daten gehen von linearen Leistungsverläufen aus und werden diesem Umstand in der aktuellen

Form nicht gerecht. Die in Tabelle 103 dargestellten Ergebnisse zum durchschnittlichen Leistungsanstieg je Woche für die CBM können somit derzeit lediglich als grobe Orientierung dienen, da ihnen jeweils lineare Modelle zugrunde gelegt wurden (Lembke et al., 2008). Unklar ist ebenfalls, ob die innerhalb der vorliegenden Studie erhobenen CBM-Daten durch einen nichtlinearen Verlauf (quadratisch bzw. höhergradig) besser beschrieben werden können. Dies zu prüfen, sollte Gegenstand zukünftiger Untersuchungen sein.

Als einen weiteren limitierenden Aspekt geben Kamata et al. (2012) an, dass die auf Basis von Regressions- oder auch Mehrebenenanalysen bestimmten Angaben für Ausgangsniveaus und Lernzuwachs der CBM nur bedingt als gültige Vergleichswerte in Frage kommen. So basieren diese Daten auf gemittelten Werten einer umfassenderen Stichprobe und sind nicht ohne weiteres auf Individualfälle zu übertragen. Daher sollte zukünftig geprüft werden, ob sich gewisse typische Leistungsprofile bestimmter Schülergruppen in den CBM-Daten widerspiegeln. Sollten sich solche Profile von Kindern nachweisen lassen, ist weiterhin zu prüfen, welche Muster diese Kinder aufweisen und wie diese inhaltlich zu interpretieren sind. Auf Grundlage solcher Daten ließe sich die Identifikation und Förderung von Schülern präzisieren.

9.5 Diskussion der Ergebnisse hinsichtlich der Ökonomie, Nutzbarkeit und Zumutbarkeit

9.5.1 Ökonomie

Zusammenfassend lässt sich auf Grundlage der durchgeführten Lehrerbefragung festhalten, dass der monatliche Einsatz von CBM im Mathematikunterricht der Klassenstufe 1 als durchaus realisierbar eingeschätzt werden kann. Während die Testungen von den

befragten Personen generell als Mehraufwand wahrgenommen wurden (vgl. Abbildung 58), gab der Großteil dieser Lehrkräfte (69.2 % von 26 Personen, vgl. Abbildung 59) dennoch an, dass dieser als gerechtfertigt bzw. eher gerechtfertigt erlebt und gern in Kauf genommen wurde. Ähnliches berichteten auch die Lehrkräfte, die in der Untersuchung von Shapiro et al. (2005) teilnahmen: „Teachers reported that conducting the monitoring was feasible within their routine schedule and that using the probes helped to shape their instructional decisions for all students“ (S. 28).

Um den durch CBM entstehenden Mehraufwand zu reduzieren, könnte eine computerbasierte Durchführung konzipiert werden. Vorteil einer solchen Anwendung wäre, dass die Zeit zur Auswertung der Schülerergebnisse entfallen würde, dies würde automatisch durch den PC geschehen. Positiver Nebeneffekt dabei ist, dass die Durchführung, Auswertung als auch Interpretation dadurch noch weiter objektiviert werden könnte, da bspw. standardisierte Instruktionen und Bearbeitungszeiten sehr präzise eingehalten und detaillierte Schülerprofile sowie ggf. abgeleitete Förderhinweise ohne großen Aufwand generiert werden könnten. Hinzu kommt, dass durch den monatlichen bis wöchentlichen Einsatz von CBM in der Klasse ein sehr hoher Materialverschleiß und ein hohes Druckaufkommen auftreten. Auch diese Nachteile können durch eine computergestützte Anwendung minimiert werden. Die Bedeutung des Einsatzes computerbasierter Verfahren formativer Leistungsmessung auf Schülerleistungen konnte zudem in den USA in nationalen Tests großflächig belegt werden (Nunnery, Ross & McDonald, 2006; Yeh, 2006). Der Einsatz computerbasierter Maßnahmen zur formativen Leistungserhebung ergab laut Kingston und Nash (2011) eine durchschnittliche Effektstärke von d = 0.28 hinsichtlich der Steigerung der Schülerleistung.

Trotz der Vorteile, die eine internet- bzw. computerbasierte Anwendung der CBM mit sich bringt, ist zu beobachten, dass einige Lehrkräfte dahingehend eher etwas zurückhaltend reagieren. Innerhalb der Onlineumfrage wurden die Lehrkräfte (UG 3) um eine Einschätzung der computergestützten Auswertung der CBM-Daten gebeten. Trotz der genannten Vorteile zeigen die Ergebnisse in Abbildung 66, dass lediglich 30 der insgesamt 37 Umfrageteilnehmer (ca. 81 %) angaben, die computergestützte Auswertung sei nützlich bzw. eher nützlich. Worin dies begründet liegt, bleibt offen. Eine Vermutung ist, dass der Umgang mit dem Computer einigen Lehrkräften schwer fällt. Fraglich ist zudem, inwieweit Erstklässler schon sicher mit dem Medium Computer umgehen können. So kann es sein, dass in einer computerbasierten Testung anstelle der die mathematischen Kompetenzen die Fähigkeiten im Umgang mit dem Computer erfasst werden. Ein weiteres Problem könnte auch in der technischen Ausstattung der Schulen liegen. Einerseits müssen dazu eingerichtete Computer in der Schule, vorzugsweise in der Klasse, vorhanden sein, andererseits muss es eine verantwortliche Person vor Ort geben, die die Einrichtung und Wartung der technischen Systeme vornimmt.

9.5.2 Nutzbarkeit

Dass die CBM insgesamt so positiv aufgefasst werden, lässt sich durch den Nutzen erklären, den Lehrkräfte aus den Schülerdaten ziehen können. Dabei gab die Mehrheit der Befragten an, dass es ihnen gelingt, förderbedürftige Kinder anhand der Schülerergebnisse zu identifizieren (75.6 %, vgl. Abbildung 61) und in den CBM-Daten der Schüler eine Hilfestellung zur Unterrichts- und Förderungsgestaltung zu sehen (83.7 %, vgl. Abbildung 60). Darüber hinaus gab ein Großteil der Lehrpersonen an, dass die CBM-Daten der Schüler in tabellarischer oder grafischer Form ein sinnvolles Hilfsmittel im Rahmen von Elterngesprächen (89.2 %, vgl. Abbildung

62) oder in Gesprächen mit Kollegen (73 %, vgl. Abbildung 63) darstellen. Diese Ergebnisse reihen sich in die Befunde von Shapiro et al. (2005) ein.

Werden die CBM-Daten der Kinder, wie in Abschnitt 3.2.3 beschrieben, durch das Zusammenzählen der richtig gelösten Aufgaben ausgewertet, sollen anhand der Ergebnisse Hinweise auf Unterrichtseffektivität als auch mögliche Maßnahmen zur Förderung von Schülern gezogen werden können. Dennoch liefert diese Art der Datenauswertung nur begrenzte als auch undifferenzierte Informationen, wie schon in Abschnitt 9.3.1 beschrieben. So wird bei dieser Form der Auswertung nicht berücksichtigt, wie viele Items ein Schüler tatschlich bearbeitet hat, also wie viele Items richtig oder falsch gelöst wurden. Eine derartige Bewertung würde jedoch eine breitere Informationsgrundlage zur Schnelligkeit und Genauigkeit des Schülers bei der Bearbeitung der Aufgaben liefern. Gemäß Binder (1996), Burns et al. (2006) sowie VanDerHeyden und Burns (2008) kann auf Grundlage dieser Kombination an Informationen Auskunft darüber gegeben werden, an welcher Stelle im Lernprozess sich ein Kind befindet und von welchen Unterstützungsmaßnahmen es profitieren könnte (vgl. Abschnitt 3.3.2.1). Ausgehend vom Profil eines Schülers und von Studien zur Wirksamkeit von Maßnahmen zur mathematischen Förderung, wie sie bspw. Grünke (2006) beschreibt, wären folgende Aspekte bei der Förderplanung angezeigt:

- Kinder, die viele Fehler bei geringer Geschwindigkeit machen, sollten weiterhin am grundlegenden Verständnis des jeweiligen mathematischen Sachverhalts arbeiten. Dabei sollte kleinschrittig, eher direkt instruierend vorgegangen werden. Der Prozess sollte durch sinnvolle Veranschaulichungen und Verbalisierung des Vorgehens unterstützt werden und durch direktes korrektives Feedback seitens der Lehrkraft begleitet werden.

- Kinder, die wenig Fehler bei geringer Geschwindigkeit machen, sollten ihr vorhandenes Wissen durch Übungsphasen automatisieren. Die Übungseinheiten können eigenständig, mit einem Partner oder in der Gruppe stattfinden und sollten durch ein korrektives Feedback begleitet werden. Der Schwierigkeitsgrad der Aufgaben sollte sukzessive erhöht werden.
- Kinder, die wenig Fehler bei hoher Geschwindigkeit machen, haben die jeweiligen mathematischen Sachverhalte weitgehend verstanden. Ihnen sollte die Anwendung auf andere Sachverhalte ermöglicht bzw. mit anderen Darstellungsformen geübt werden (Generalisierung).

Insgesamt gaben 81.1 % der befragten Lehrkräfte der UG 3 an, keine weiteren Informationen zur Bedeutung von und zum Umgang mit CBM zu benötigten (vgl. Abbildung 64). In den Ausführungen von Gersten, Chard, Jayanthi, Baker, Morphy und Flojo (2009) ist jedoch ein gegenteiliges Bild zu finden. So berichten die Autoren von eher kleinen Effekten der regelmäßigen Lernverlaufsdokumentation zur Förderung von Kindern mit spezifischem Förderbedarf im Bereich der Mathematik durch Regelschullehrkräfte. Dies wird darauf zurückgeführt, dass diese weniger mit den Grundlagen der Diagnostik und dem Umgang mit erhobenen Daten vertraut sind (Gersten et al., 2009). Für den sinnvollen Einsatz diagnostischer Maßnahmen im Unterricht resümieren Hoge und Coladarci (1989), dass Lehrkräfte hinsichtlich der Bedeutung des Einsatzes standardisierter diagnostischer Maßnahmen als einen wesentlichen Aspekt des Unterrichts sensibilisiert werden müssen. Überdies sollten die Lehrer auch in der Umsetzung standardisierter diagnostischer Maßnahmen unterstützt werden. Dies betrifft nicht nur die Durchführung von Testverfahren, sondern ebenso die Auswertung und vor allem die Interpretation der erhobenen Schülerdaten. Informationen hinsichtlich des Einsatzes diagnostischer Maßnahmen stellen dabei jedoch nur

einen Teilaspekt dar. Andererseits erscheint es ebenso wichtig, den Lehrkräften sinnvolle Strategien zur Förderung von Kindern mit Schwierigkeiten beim Erlernen mathematischer Inhalte zu vermitteln. Dies zeigt sich auch in der vorliegenden Studie. So bekundeten knapp die Hälfte der befragten Lehrkräfte (51.3 %, vgl. Abbildung 65) starkes Interesse, zusätzlich hinsichtlich Möglichkeiten spezifischer Förderung von rechenschwachen Kindern informiert zu werden. Auch Gersten et al. (2009) führen an, dass die Unterstützung der Lehrkräfte zu weiteren Leistungssteigerungen führen kann und damit essentiell ist. Dies kann bspw. im Rahmen von Fortbildungen geleistet werden. Weitere Forschungsergebnisse deuten darauf hin, dass Regelschullehrkräfte bei der Förderplanung externe Unterstützung durch weiteres Fachpersonal, wie bspw. durch einen Sonderpädagogen bedürfen (D. Fuchs, L.S. Fuchs & Compton, 2012; Gersten et al., 2009). Denkbar ist zudem eine Erweiterung der CBM-Auswertungsrichtlinien um konkrete Interpretations- und Förderhinweise, wie sie zuvor beschrieben wurden. Ein solches Diagnose- und Förderkonzept ist nicht nur sehr praxisrelevant, sondern kann auch gut in einem wie in Abschnitt 9.5.1 beschriebenen computerbasierten System umgesetzt werden. Derzeit wird am ISER der Universität Rostock ein internetbasiertes System entwickelt, welches Lehrkräfte beim regelmäßigen Einsatz von CBM-Tests im schulischen Alltag unterstützen soll. Dazu werden die entwickelten CBM für die Bereiche Deutsch und Mathematik sowie weitere diagnostische Materialien für registrierte Nutzer als kostenfreie Kopiervorlage zur Verfügung gestellt. Die erhobenen Daten können die Lehrkräfte anschließend in diesem System speichern und organisieren sowie sich die Schülerergebnisse automatisch grafisch als auch tabellarisch auswerten lassen. Erste Erfahrungen mit dieser Arbeitsweise im Rahmen des Rügener Inklusionsmodells sprechen dafür, dass die Internetplattform die Lehrkräfte bei der regelmäßigen Lernverlaufsdokumentation ihrer Schüler unterstützt und sie zeitökonomischer und nutzbarer gestaltet.

9.5.3 Zumutbarkeit

Wie in Abschnitt 8.4 erwähnt, gab es zu Beginn der Studie einige kritische Stimmen, welche darauf hinwiesen, dass die zeitlich eng begrenzte Bearbeitung von mathematischen Anforderungen bei Erstklässlern zu hohen Druck ausübt und damit nicht zumutbar sein könnte. Diesbezügliche Informationen sind in der Fachliteratur zu CBM nicht zu finden und können auch in vorliegender Studie nicht gezeigt werden. Hinweise bezüglich der Zumutbarkeit der innerhalb der vorliegenden Arbeit entwickelten CBM sind in Abbildung 67 grafisch dargestellt. Von den insgesamt 37 Umfrageteilnehmern gaben 31 (84 %) an, dass das Bearbeiten der CBM den Kindern Spaß gemacht bzw. eher Spaß gemacht hat. Hingegen hatten sechs Personen (16 %) den Eindruck, dass den Schülern das Bearbeiten der monatlichen CBM eher keinen Spaß gemacht hat. Insgesamt ist demnach von einem zumutbaren Einsatz der innerhalb der Arbeit konzipierten CBM auszugehen. Dennoch müssen die Anwender der CBM dahingehend sensibilisiert werden, dass sie den Kindern jeweils vor dem Bearbeiten der CBM darauf hinweisen, dass die gestellten Aufgaben aufgrund ihrer Menge und Schwierigkeit innerhalb der vorgegebenen Bearbeitungszeit nicht gänzlich geschafft werden müssen. Dies könnte die Kinder weniger unter Druck setzen und damit den Grad der Zumutbarkeit erhöhen.

10 Fazit und Ausblick

Das nachfolgende abschließende Kapitel beginnt mit einer Zusammenschau der zentralen Befunde der vorliegenden Forschungsarbeit (Abschnitt 10.1). Im Anschluss daran werden die Ergebnisse hinsichtlich des Grades der internen sowie externen Validität diskutiert (Abschnitte 10.2 und 10.3). Abschließend werden die vorgestellten Ergebnisse vor dem Hintergrund der theoretischen Ausarbeitungen sowie der angeführten Forschungslage zu CBM in den USA und Deutschland eingeordnet (Abschnitt 10.4) und Implikationen für weiterführende Forschungsarbeiten sowie für Praktiker im schulischen und schulpolitischen Umfeld aufgezeigt (Abschnitt 10.5).

10.1 Kurzzusammenfassung zentraler Befunde

In den vorangegangenen Kapiteln wurde untersucht, inwieweit die in der vorliegenden Studie erarbeiteten CBM-Verfahren den wissenschaftlichen Hauptgütekriterien Objektivität, Reliabilität und Validität sowie darüber hinaus auch den Nebengütekriterien Ökonomie, Nutzbarkeit und Zumutbarkeit genügen und den Lernverlauf über die Zeit sensibel erfassen können. In einer etwa 30jährigen Forschungstradition ist man in den USA bereits zu umfassenden Erkenntnissen zur Entwicklung von CBM unter Berücksichtigung der oben genannten Aspekte gelangt. Während vor allem für den Bereich des Lesens bereits eine breite, empirisch belegte Datenbasis besteht (Colón & Kranzler, 2006; L.S. Fuchs, Deno & Mirkin, 1984; L.S. Fuchs et al., 1989b; Jones & Krause, 1988; Shapiro et al., 2006; Wayman et al., 2007; Wesson, 1991; Wiley & Deno, 2005; Yeo, 2010), fällt die Befundlage im Bereich der Mathematik deutlich geringer, jedoch ebenso vielversprechend aus (Christ et al., 2008; Foegen et al., 2007). So konnten in verschiedenen Literatur- und Metanalysen die positiven Effekte des regelmäßigen Einsatzes von CBM zur laufenden Unterrichtsevaluation sowie als

Feedback zur Förderung für Lehrer und Schüler (im Sinne des formative assessment) auf die Schulleistung der Kinder nachgewiesen werden (u. a. Black & Wiliam, 1998a, b; L.S. Fuchs & D. Fuchs, 1986; Kingston & Nash, 2011). Dies wurde ausführlich in den Abschnitten 3.1.3 bzw. 3.2.8.3 dargelegt.

Insgesamt betrachtet, kennzeichnen sich die in der vorliegenden Studie erarbeiteten und geprüften CBM durch zufriedenstellende Befunde hinsichtlich wissenschaftlicher Gütekriterien. Dennoch gibt es qualitative Unterschiede zwischen den Verfahren, wie aus Tabelle 104 hervorgeht. In der Tabelle wird auf Grundlage der in der vorliegenden Untersuchung ermittelten Befundlage für jedes hier vorgestellte CBM eingeschätzt, inwieweit es die Gütekriterien Objektivität, Reliabilität, Validität, Änderungssensibilität, Ökonomie, Nutzbarkeit sowie Zumutbarkeit erfüllt. Zur Ökonomie, Nutzbarkeit und Zumutbarkeit wurden in dieser Studie aggregierte Daten über alle hier entwickelten CBM erhoben, somit sind differenzierte Aussagen für einzelne Verfahren nicht möglich. Die Ergebnisse wurden zur Einschätzung der untersuchten Hauptgütekriterien zusammengefasst betrachtet und wie folgt kategorisiert:

- „++“ deutet an, dass ausreichend empirische Nachweise zur Annahme eines hohen Grades der Gültigkeit des jeweiligen Gütekriteriums gegeben sind.
- „+“ deutet an, dass einige empirische Nachweise zur Annahme eines hohen Grades der Gültigkeit des jeweiligen Gütekriteriums gegeben sind. Weitere zukünftige Analysen diesbezüglich werden empfohlen.
- „–“ deutet an, dass ungenügend empirische Nachweise zur Annahme eines hohen Grades der Gültigkeit des jeweiligen Gütekriteriums gegeben sind. Weitere zukünftige Analysen und ggf. Überarbeitungen des CBM diesbezüglich sind angezeigt.

- „k.A.“ deutet an, dass keine empirischen Angaben zur Einschätzung des Grades der Gültigkeit des jeweiligen Gütekriteriums vorliegen.

Am besten schneidet dabei das CBM „Addition im Zwanzigerraum“ ab, welches hinsichtlich aller geforderten Kriterien hinreichende Nachweise erzielt und sich damit empirisch bestens bewährt hat. Die CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“, „Arithmetische Basiskompetenzen“, „Zahlzerlegung im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ sind, wenn auch mit etwas geringerer empirischer Bewährung (siehe Kapitel 7.2, 7.3 sowie 7.4), für den praktischen Einsatz in der Schule geeignet. Für die CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ sowie „Subtraktion im Zwanzigerraum“ konnten, aufgrund unzureichender Datenlage, keine Angaben zur Splithalf-Reliabilität gemacht werden. Die in Abschnitt 3.2.8.1 beschriebene Verfahrensweise zur Konzeption der Verfahren spricht jedoch für eine hohe interne Konsistenz der Verfahren. Des Weiteren wurde für das CBM „Mengenvergleich im Zwanzigerraum“ lediglich ein geringer Koeffizient zur Retest-Reliabilität berechnet, in Folgestudien sollte die Retest-Reliabilität des Verfahrens erneut untersucht werden. Sollten sich erneut geringe Werte ergeben, ist eine Überarbeitung des Verfahrens angezeigt.

Tabelle 104: Überblick über zentrale Befunde der innerhalb der Studie erarbeiteten CBM-Verfahren

Gütekriterium		ZL	ZR	MV	AB	ZZ	AD	SUB
Objektivität		++	++	++	++	++	++	++
Reliabilität	Splithalf-Reliabilität	++	++	+	+	k.A.	k.A.	k.A.
	Retest-Reliabilität	++	++	-	++	++	++	++
	Paralleltest-Reliabilität	++	++	++	++	++	++	++
Validität	Augenschein- und Inhaltsvalidität	++	++	++	++	++	++	++
	Übereinstimmungsvalidität	+	++	++	+	++	++	++
	Prognostische Validität	+	+	+	+	++	++	+
	Inkrementelle Validität	++	++	++	+	++	++	+
Änderungssensibilität		++	++	++	+	+	++	++
Ökonomie					+			
Nutzbarkeit					+			
Zumutbarkeit					++			

Erläuterungen: ZL – CBM „Zahlenlesen im Zwanzigerraum“; ZR – CBM „Zahlenreihe im Zwanzigerraum“; MV – CBM „Mengenvergleich im Zwanzigerraum“; AB – CBM „Arithmetische Basiskompetenzen“; ZZ – CBM „Zahlzerlegung im Zwanzigerraum“; AD – CBM „Addition im Zwanzigerraum“; SUB – CBM „Subtraktion im Zwanzigerraum“; ++ – ausreichende Nachweise gegeben; + – einige Nachweise gegeben; – – ungenügend Nachweise gegeben; k.A. – keine Angabe möglich

10.2 Aspekte zur Einschätzung der internen Validität der Befunde

Trotz der insgesamt positiven Befunde zur Güte der hier geprüften CBM, müssen dennoch einige, die praktische Verwendung der Verfahren einschränkende Gesichtspunkte der internen und externen Validität der Ergebnisse skizziert werden. Zunächst werden dazu die Ergebnisse limitierende Faktoren beleuchtet. Dies betrifft vor allem Aspekte des Forschungsdesigns, der hier vorgestellten CBM-Verfahren und statistische Verfahrensweisen. Im Anschluss wird die Frage nach der Generalisierbarkeit der in der vorliegenden Arbeit ermittelten Befunde diskutiert. Für die weitere Forschung sind aus diesen Überlegungen methodische Schlüsse zu ziehen.

Die vorliegende Untersuchung ist als ein Teil in eine umfassendere Hauptstudie eingegliedert, welche grundlegend eine andere Forschungsintention besitzt (Voß et al., 2012). Aufgrund dessen ist

das verwendete Untersuchungsdesign nicht optimal auf die Zielstellungen der vorliegenden Arbeit ausgerichtet, sondern an eine vorhandene Datenlage und die daraus resultierenden Restriktionen gebunden. Einige der Analysen konnten daher nur für vereinzelte Parallelversionen eines jeden CBM durchgeführt werden. Es ist noch nicht abschließend geklärt, inwieweit die ermittelten Ergebnisse auf weitere Parallelversionen des jeweiligen CBM übertragen werden können. Diesbezüglich müssten weitere Analysen durchgeführt werden.

Innerhalb der Studie konnten zudem keine Kontrollmaßnahmen zur korrekten Durchführung der CBM implementiert werden, so dass auftretende systematische Fehler bei der Durchführung nicht auszuschließen sind, die zu einer Verzerrung der hier präsentierten Ergebnisse geführt haben könnten. Bspw. hatten die Lehrkräfte der UG 2 (Nordvorpommern) den Auftrag je zwei, zur Bestimmung der Paralleltest-Reliabilität, eingesetzten, Versionen jedes CBM an jeweils zwei aufeinander folgenden Tagen durchzuführen. Ob diese Anweisungen eingehalten wurden, konnte im Rahmen dieser Studie nicht abgesichert werden.

Wie meist in longitudinal angelegten Feldstudien, war auch in der vorliegenden Untersuchung eine hohe Variabilität des Stichprobenumfangs über die Messzeitpunkte hinweg, vor allem in der Region Nordvorpommern (UG 2), zu verzeichnen. Nicht in jedem Fall liegen Angaben zur Begründung des Wegfalls der Daten vor. Der Datenausfall in Nordvorpommern folgt zum Teil einer Systematik, so haben teilweise ganze Klassen oder gar Schulen im Laufe der Zeit den regelmäßigen CBM-Einsatz abgebrochen. Dies ist wahrscheinlich durch den empfundenen Mehraufwand durch die regelmäßigen Testungen zu begründen. Ist das der Fall, würde dies bedeuten, dass zum Ende des Schuljahres verstärkt Klassen von besonders engagierten Lehrkräften, die sich der Bedeutung der CBM

verstärkt bewusst sind und den resultierenden Mehraufwand in Kauf nehmen, an den Testungen teilgenommen haben. Denkbar ist bspw. auch, dass es Lehrkräfte gab, die aufgrund eines hohen Anteils an förderbedürftigen Kindern in ihren Klassen nicht weiter an der Untersuchung teilnahmen, da ihnen dies neben der Förderarbeit im Unterricht als zu aufwendig erschien. Diese rein spekulativen Annahmen konnten jedoch auf Grundlage der erhobenen Informationen nicht geklärt werden. Abbrüche des CBM-Einsatzes können zu einer Verzerrung der hier ermittelten Ergebnisse geführt haben und ihre Gültigkeit einschränken. Auf der Insel Rügen (UG 1) hingegen gehörten die regelmäßigen Lernstandserhebungen zum Treatment des RIM und wurden in allen untersuchten Klassen durchweg über die Zeit der Klassenstufe 1 durchgeführt. Hier sind nur einzelne Datenausfälle zu beobachten (z. B. krankheitsbedingt oder durch Wegzüge).

Für die Analysen innerhalb der vorliegenden Studie wurden fehlende Datenwerte im Datensatz (Missings) fallweise ausgeschlossen. In der methodologischen Forschung werden hingegen Schätzmethoden zum Umgang mit Missings, wie bspw. der Maximum-Likelihood-Methode oder multiplen Imputationen, angewendet, welche gegenüber traditionellen Forschungsmethoden wie dem Fallausschluss überlegen sind (Enders, 2010; Baraldi & Enders, 2010). Grundvoraussetzung bei der Schätzung fehlender Werte ist jedoch, dass die Ursache für das Fehlen der Werte keiner Systematik folgt (Enders, 2010). Wie zuvor beschrieben ist dies für die UG 2 (Nordvorpommern) nicht, für die UG 1 (Rügen) nur bedingt gegeben. Die hier ermittelten Ergebnisse müssen demnach vor diesem Hintergrund interpretiert werden.

Bei der Interpretation der Ergebnisse der vorliegenden Studie muss darüber hinaus beachtet werden, dass die Reliabilität, ebenso wie die Validität, maßgeblich durch die Wahl der Testitems bestimmt wird, so

können Kennwerte der Gütekriterien durch statistisch gesehen „schlechte“ Items (z. B. Items mit nur geringer Trennschärfe bzw. zu einfache oder zu schwere Items) nach unten verzerrt werden. Da die CBM-Daten innerhalb der vorliegenden Studie aggregiert, d. h. als Gesamtwert für jedes Verfahren erhoben wurden, sind keine differenzierten Aussagen über einzelne Testitems möglich. Entsprechend können auch keine Aussagen zu einer solchen Verzerrung der Reliabilitäts- bzw. Validitätskoeffizienten getroffen werden.

Weiterhin muss angemerkt werden, dass die hier dargestellten CBM mit Ausnahme des CBM „Zahlenlesen im Zwanzigerraum“ als Gruppenverfahren konzipiert wurden, um den Grad der Ökonomie bezüglich der Zeit und damit auch der Nutzbarkeit zu erhöhen. In den USA werden vor allem die hier beschriebenen CBM der Niveaustufe 1 in Einzelsitzungen angewendet. Diese grundlegenden Unterschiede in der Durchführung können der Grund für die festgehaltenen Abweichungen der innerhalb der vorliegenden Studie ermittelten und den in der US-amerikanischen Literatur beschriebenen Reliabilitäts- bzw. Validitätskoeffizienten sein (vgl. Abschnitte 9.2 und 9.3).

Ein weiterer Aspekt, welcher zu Verzerrungen der Daten führen könnte, ist das sog. Phänomen des *Regressionseffekts*, welches besagt, dass es bei wiederholter Messung eines Merkmales zu einer Annäherung der individuellen Messwerte zum Gruppenmittelwert kommt. Vor allem Probanden mit extremer Merkmalsausprägung zur ersten Messung zeigen bei der zweiten Erhebung des Merkmals weniger extreme Ausprägungen. Gemessene Unterschiede können somit durch die Veränderung des Personenmerkmals oder durch messfehlerbedingte Schwankungen zustande kommen. Speziell bei Veränderungsmessungen wird dieser Effekt bedeutsam, da nicht eindeutig ist, was ursächlich für eine festgehaltene Veränderung des

gemessenen Merkmals ist. Für eine detaillierte Beschreibung des Regressionseffekts sei auf Rost (2007) verwiesen.

Die Betrachtung der Verteilungen der CBM „Zahlzerlegung im Zwanzigerraum" (Abbildung 49 bzw. Abbildung 50), „Addition im Zwanzigerraum" (Abbildung 51 bzw. Abbildung 52) sowie „Subtraktion im Zwanzigerraum" (Abbildung 53 bzw. Abbildung 54) führt zu der Annahme, dass diese recht spitzgipflig ausfallen. Dies ist ein Anzeichen dafür, dass die in den Verfahren gestellten Aufgaben für die untersuchten Kinder recht schwer waren und sie nicht viele Items in einem Bearbeitungszeitraum von einer Minute bearbeiten konnten. Um noch mehr Varianz und damit zeitgleich differenziertere Aussagen gerade auch im unteren Leistungsbereich für diese Verfahren zu ermöglichen, sollten die vorgenommenen Analysen zukünftig mit einer erhöhten Bearbeitungszeit von z. B. anderthalb oder zwei Minuten wiederholt werden. Evtl. erweisen sich die hier dargestellten CBM dann auch noch sensibler auf Veränderungen der Schülerleistungen über die Zeit. Ebenfalls ist zu vermuten, dass die ermittelten Reliabilitäts- und Validitätskoeffizienten durch ein solches Vorgehen höher ausfallen könnten.

Die hier vorgenommen Analysen basieren auf der Klassischen Testtheorie (KTT), wonach das Ergebnis einer Messung die Summe der wahren Merkmalsausprägung und einem zufälligen Fehler aufgrund des Messinstruments darstellt. Dabei sind die wahren Werte und die Messfehler unkorreliert (Moosbrugger, 2012). Mit den Axiomen der KTT ergeben sich einige Restriktionen für die Evaluation von diagnostischen Testverfahren, welche vor allem auch im Zusammenhang mit der Messung von Veränderungen relevant sind. So ist eine Erfassung von Veränderungen des wahren Merkmalsanteils einer Person über mehrere Messzeitpunkte nur bedingt mit den Axiomen der Klassischen Testtheorie (KTT) vereinbar, da hierbei der wahre Wert als konstant angenommen wird

und Veränderungen als Messfehler interpretiert werden (Lienert & Raatz, 1998). An vielen Stellen wurde folglich der ergänzende Einsatz von Analysemaßnahmen auf Basis der Item-Response-Theorie (IRT, synonym Probabilistische Testtheorie) vorgeschlagen (u. a. Lienert & Raatz, 1998; Moosbrugger, 2012; Wilbert & Linnemann, 2011). Da die CBM-Daten innerhalb der vorliegenden Studie nicht itemweise erhoben wurden, sind Analysen im Rahmen der IRT an dieser Stelle nicht möglich, sollten jedoch in Folgestudien angestrebt werden.

Die Analysen der hier vorliegenden Studie bezogen sich ausschließlich auf die im Jahr 2004 von Lynn S. Fuchs postulierten Forschungsstufen. Diese haben sich seither in der US-amerikanischen Forschung zum Thema CBM etabliert (Foegen et al., 2007). Mit dem wachsenden Forschungsinteresse an CBM im deutschsprachigen Raum, geht eine Debatte einher, die methodische Aspekte bei der Konzeption und Evaluation von den Verfahren fokussiert. Hierzu weisen Klauer (2011) sowie Wilbert und Linnemann (2011) auf verschiedene methodologische Aspekte hin. Die Anregungen von Klauer beziehen sich auf Aspekte der Reliabilität und Validität von CBM für den mathematischen Bereich sowie die Messung von Leistungsveränderungen der Verfahren und wurden bereits in den Abschnitten 3.3.3 und 9.2.1 skizziert. Wilbert und Linnemann (2011) stellen darüber hinaus programmatische Kriterien auf, welche jedes Testverfahren mit dem Ziel der Lernverlaufsdiagnostik erfüllen sollte. Dabei werden in fünf Schritten Analysen auf Basis der KTT mit denen der IRT verbunden (Wilbert & Linnemann, 2011):

1. Analyse der Items nach der KTT hinsichtlich Trennschärfen, Schwierigkeiten, interner Konsistenz, Homogenität und Retest-Reliabilität (evtl. unter Verwendung zuvor bestimmter Personen-Parameter nach der IRT)

2. Prüfen auf Eindimensionalität entweder durch Nachweis der Passung eines Rasch-Modells oder durch eine konfirmatorische Faktorenanalyse
3. präzise Bestimmung der Itemschwierigkeit auf Basis einer Raschmodellierung
4. Ermittlung der Testfairness durch Berechnung von Differential Item Functioning-Werten

Untersuchungen in der von Wilbert und Linnemann (2011) beschriebenen Form sind auf der Datengrundlage der vorliegenden Studie nicht möglich. Insgesamt erscheinen die hier vorgenommenen Analysen vor dem Hintergrund der von L.S. Fuchs (2004) beschriebenen Forschungsstufen als hinreichend, um eine Einschätzung zur Güte und Anwendbarkeit der im Rahmen dieser Studie erarbeiteten CBM vornehmen zu können. Weitere Untersuchungen in die von Wilbert und Linnemann (2011) beschriebene Richtung sind im Rahmen zukünftiger Studien wünschenswert.

10.3 Aspekte zur Einschätzung der externen Validität der Befunde

Nachfolgend wird die Frage nach der Allgemeingültigkeit bzw. der Generalisierbarkeit der hier geschilderten Befunde diskutiert. In dem Zusammenhang muss analysiert werden, inwieweit die hier erzielten Ergebnisse allgemeine Gültigkeit besitzen, wie bspw. in anderen Regionen und Settings.

Bei der vorliegenden Studie handelt es sich um ein Quasi-Experiment im Sinne einer Feldstudie. Innerhalb der Untersuchung wurden die regelmäßigen Erhebungen von CBM-Daten im etwa monatlichen Abstand durch die entsprechenden Lehrkräfte im schulischen Alltag realisiert. Es gab keine externen Unterstützungssysteme, wie bspw.

eine Anleitung der Datenerhebung durch studentische Hilfskräfte. Beurteilungen zur Ökonomie und Nutzbarkeit der Daten stellen demnach eine realistische Einschätzung für den Einsatz der CBM im schulischen Alltag dar.

Die Generalisierbarkeit der Befunde hängt zudem wesentlich davon ab, inwieweit die hier untersuchten Probandengruppen bzw. deren erhobenen Leistungsdaten repräsentativ sind. Bei der Repräsentativität von Daten wird zwischen zwei Formen unterschieden: der spezifischen und der globalen Repräsentativität (Goldhammer & Hartig, 2012). Globale Repräsentativität gilt als nachgewiesen, wenn eine auf Basis einer echten Zufallsstichprobe aus einer definierten Population gezogene Stichprobe in ihrer Zusammensetzung der Zielpopulation entspricht. Spezifische Repräsentativität liegt vor, wenn die untersuchte Stichprobe zumindest hinsichtlich des zu untersuchenden Merkmals mit der Population (hier: Schüler erster Grundschulklassen in Deutschland) vergleichbar ist (Goldhammer & Hartig, 2012). Aufgrund des Vorgehens bei der Stichprobenziehung kann in der vorliegenden Studie nicht von einer globalen Repräsentativität der ermittelten Daten ausgegangen werden. Grundlage für die beschriebenen Untersuchungen waren verschiedene Schülerstichproben aus Mecklenburg-Vorpommern. Die UG 1 sowie die UG 2 umfassten jeweils nahezu alle Schüler der Regionen Rügen und Nordvorpommern, in beiden Fällen lag jedoch keine Totalerhebung vor. Dennoch ist davon auszugehen, dass die Häufigkeitsverteilungen der erhobenen Ergebnisse für Mecklenburg-Vorpommern repräsentativ sind. Die in Abschnitt 1.1 dargestellten Ergebnisse von PISA deuten jedoch auf Leistungsunterschiede zwischen den Schülern der verschiedenen deutschen Bundesländer hin, ein Rückschluss auf die Erstklässler aus ganz Deutschland erscheint an dieser Stelle demnach nicht uneingeschränkt zulässig zu sein. Um zu prüfen, inwieweit eine Verallgemeinerung der in dieser

Studie ermittelten Befunde in der UG 1 auf andere Bundesländer möglich ist, können die Leistungen der untersuchten Rügener Kinder zu Beginn der ersten Klasse im Vergleich zu den angegebenen Bundesnormen der eingesetzten Verfahren betrachtet werden. Ein derartiger Vergleich wurde in der Stichprobenbeschreibung in Abschnitt 6.1 vorgenommen. Bis auf geringe Schwankungen erzielten die Rügener Erstklässler im Mittel vergleichbare Leistungen mit der Bundesnorm. Zur Generalisierbarkeit der Ergebnisse auf Datengrundlage der UG 2 kann keine genaue Auskunft gegeben werden, zumal diese durch die zuvor beschriebene Variabilität der Schülergruppe Nordvorpommerns eingeschränkt wird. Um die Repräsentativität weitergehend umfassend abzusichern, wäre eine Erhebung von CBM-Daten auf bundesweiter Ebene erstrebenswert. Vor allem die hier errechneten wöchentlichen Durchschnittsanstiege (vgl. Abschnitt 9.4) sollten auf bundesweiter Ebene auf Gültigkeit überprüft werden. Die Anstiege dienen, neben der Einschätzung der Schülerleistung auf intraindividueller Ebene über die Zeit, als interindividuelle Vergleichswerte zu den verschiedenen Messzeitpunkten, auf deren Grundlage eine Einschätzung des Lernverlaufs im sozialen Vergleich möglich ist. Diesen Werten kommt eine besonders hohe praktische Relevanz zu, da auf ihrer Grundlage Förderziele festgelegt werden können und Lernverläufe von Kindern präzise eingeordnet und eingeschätzt werden können (vgl. Abschnitt 3.2.3).

10.4 Abschließende Einordnung der Ergebnisse

Zentrales Anliegen der vorliegenden Arbeit lag in der Erarbeitung und Evaluation von CBM-Verfahren für den mathematischen Erstunterricht, um einerseits eine frühe Identifikation von Risikokindern zu ermöglichen und andererseits Hinweise auf die Effektivität der durch die Lehrkraft dargebotenen Unterrichts- und Fördermaßnahmen zu erhalten (formative Evaluation). Wie wichtig

die frühzeitige Erfassung von Entwicklungsrisiken im Bereich der Mathematik und der zeitnahe Einsatz von spezifischen Fördermaßnahmen ist, wurde in Abschnitt 2.11 auf Grundlage theoretischer Ausarbeitungen zur mathematischen Kompetenzentwicklung ab dem frühen Kindesalter zusammenfassend dargelegt. Eine zentrale Aufgabe der wissenschaftlichen Forschung stellt an dieser Stelle die Konzeption und Evaluation diagnostischer Verfahren für den Einsatz im schulischen Alltag dar, um eine objektive, reliable als auch valide Erfassung der Schülerleistungen zu ermöglichen. Der regelmäßige Einsatz von CBM-Verfahren im mathematischen Erstunterricht soll die Dokumentation der Leistungsentwicklung der Kinder ermöglichen. Damit gliedert sich die Studie in die aktuelle Präventionsforschung ein. Im Rahmen der Präventionsdebatte und -forschung in den USA hat sich in einer nunmehr 30jährigen Forschungstradition der Response-to-Instruction-Ansatz (vgl. 3.1.4) als ein bedeutsames präventiv ausgerichtetes Konzept zur gemeinsamen Beschulung in der Regelschule etabliert. Ein Kernelement des RtI-Ansatz, neben nach Intensität auf mehreren Stufen systematisierten Präventionsmaßnahmen und Maßnahmen der evidenzbasierten sowie datengeleiteten Praxis, stellen CBM dar (National Center on Response to Intervention, 2010). In jüngerer Vergangenheit gewinnt der RtI-Ansatz als Rahmenkonzept für eine präventive und integrative Beschulung ebenso im deutschsprachigen Raum an Bedeutung (Diehl & Hartke, 2007; Hagen & Hillenbrand, 2012; Hartke & Diehl, 2013; Hartmann, 2008; Hartmann & Müller, 2009; Huber & Grosche, 2012; Diehl et al., 2012; Mahlau et al., 2011; Parmar, 2008; Walter, 2008a). Vor diesem Hintergrund ist es dahingehend notwendig, objektive, reliable und valide curriculumbasierte Messverfahren für den deutschsprachigen Raum zu entwickeln. Dieses Ziel wurde mit der vorliegenden Arbeit verfolgt, indem CBM für den Erstmathematikunterricht erarbeitet und evaluiert wurden (vgl. Kapitel 4). Diese Verfahren stellen Adaptionen bereits existierender

Verfahren aus den USA dar. In der vorliegenden Arbeit wurde geprüft, inwieweit sich diese Adaptionen zur Dokumentation von Lernverläufen im Bereich des Erstunterrichts Mathematik auch in ersten Grundschulklassen innerhalb Deutschlands einsetzen lassen. Zentraler Aspekt dabei war es, zu analysieren, ob die entwickelten CBM-Verfahren dazu geeignet sind, den Lernverlauf der Schüler anhand monatlicher Datenerfassungen abzubilden. Davon ausgehend sollte geprüft werden, ob es mit Hilfe der Verfahren gelingt, ungenügende Lernzuwächse von Kindern möglichst früh zu identifizieren und Prognosen für den weiteren Entwicklungsverlauf zu geben. Grundlage für die hier beschriebenen Analysen stellten die von L.S. Fuchs (2004) postulierten Voraussetzungen zur Gültigkeit von CBM dar (vgl. Abschnitte 7.2, 7.3, 7.4 sowie 7.5). Diese betreffen neben klassischen Gütekriterien der Objektivität, Reliabilität und Validität ebenso Aspekte der Veränderungssensibilität, sowie Nutzbarkeit. In Tabelle 104 wurde ein Überblick über die zentralen Befunde dieser Studie gegeben. Zusammenfassend ist zu konstatieren, dass bezüglich der Kriterien der Objektivität, der Reliabilität und der Validität sowie der Änderungssensibilität, Ökonomie, Nutzbarkeit und Zumutbarkeit für die im Rahmen dieser Arbeit entwickelten sieben CBM zufriedenstellende Ergebnisse erzielt wurden. Die Befunde reihen sich im Wesentlichen in die Befunde der US-amerikanischen Forschungslage der vergangenen 30 Jahre ein. So erscheint anhand der hier vorgestellten CBM, der Verlauf einzelner mathematischer Kompetenzen (Zahlenwissen, Menge-Zahl-Korrespondenz, Teile-Ganzes-Konzept und relationales Zahlenverständnis) abbildbar zu sein. Ebenso erscheint eine valide Identifikation von Risikokindern möglich zu sein. Dies zeigt sich anhand der hohen prognostischen Güte der Verfahren. Auf Grundlage der wiederholt erhobenen CBM-Daten können Lehrkräfte demnach objektiv Kinder erkennen, deren Entwicklung mathematischer Kompetenzen und Einsichten gefährdet ist. Durch frühzeitig einsetzende und spezifische Förderangebote kann einer

derartigen Gefahr entgegengewirkt werden. Je öfter diese Lernstandserhebungen mithilfe von CBM durchgeführt werden, umso engmaschiger wird das Informationsmetz hinsichtlich der Entwicklung eines Schülers. In der vorliegenden Studie hat sich ein zeitlicher Abstand von etwa vier Schulwochen zwischen den Testungen als sinnvoll erwiesen. Für die Festsetzung einer solchen Frequenz bedarf es jedoch auch CBM-Verfahren, die geeignet sind, den Lernverlauf in diesem zeitlichen Rahmen sensibel zu erfassen. Dieser Anforderung werden fünf der in der vorliegenden Studie erarbeiteten CBM gerecht. Die CBM „Arithmetische Basiskompetenzen“ und „Zahlzerlegung im Zwanzigerraum“ erfüllen diese Anforderung nur eingeschränkt. Der erfolgreiche Einsatz von CBM hängt neben der Qualität der Verfahren auch maßgeblich davon ab, wie sie eingesetzt werden und wie die erhobenen Ergebnisse interpretiert und in adäquate Förderung umgesetzt werden. In diesem Zusammenhang spielt die Lehrkraft eine zentrale Rolle. Damit CBM als wesentliches Element erfolgreicher Lernentwicklung regelmäßig und gewinnbringend von Lehrkräften in der Praxis genutzt werden, muss eine hohe Akzeptanz gegenüber den Verfahren bestehen. Eine weitere wichtige Voraussetzung ist das Erkennen des mit dem CBM-Einsatz verbundenen Nutzens. Auf Basis einer Lehrerbefragung konnte innerhalb der vorliegenden Studie abgesichert werden, dass sich die entwickelten CBM aus Perspektive der befragten Lehrkräfte, wenn auch mit einem gewissen Mehraufwand, gewinnbringend in die Unterrichtspraxis integrieren lassen. So bestätigte die Mehrheit der befragten Lehrkräfte, anhand der CBM-Daten Risikokinder frühzeitig zu erkennen und Hinweise für die Förderung zu erhalten. Trotz der Erkenntnisse zum Nutzen und zur allgemeinen Akzeptanz der Verfahren, deutet die Befragung der Lehrkräfte jedoch auch daraufhin, dass ein Teil der Befragten weitere Unterstützung bei der Förderung rechenschwacher Kinder benötigt.

Generell ist festzuhalten, dass die vorliegenden sieben Adaptionen US-amerikanischer CBM-Verfahren als ein Werkzeug zur unterrichtsbegleitenden Lernverlaufsdokumentation im Erstmathematikunterricht auch im deutschen Schulsystem einsetzbar sind. Insgesamt muss jedoch einschränkend festgehalten werden, dass CBM lediglich einen Einblick in einen Teilausschnitt der mathematischen Einsichten und Kompetenzen eines Schülers geben und damit kein umfassendes Bild liefern (Christ et al., 2008). Im Gegensatz zu den Forschungsergebnissen zum Erwerb des Lesens ließen sich im Bereich der Mathematik bisher keine robusten Indikatoren nachweisen, die umfassend Aufschluss über die mathematischen Kompetenzen eines Kindes geben. Isoliert für den Bereich der Arithmetik betrachtet, scheint die Rechenflüssigkeit, verstanden als das Zusammenspiel der Geschwindigkeit und Genauigkeit beim Lösen von Rechenaufgaben, jedoch ein zentraler Hinweisgeber für die Leistungsentwicklung zu sein. Ebenso spielen basale mathematische Kompetenzen eine tragende Rolle bei der Entwicklung des Rechnens. Dies wurde im Kapitel 2 ausführlich dargestellt. Besonders relevant sind dabei sichere Zählkompetenzen, das Verständnis für die Korrespondenz von Mengen und Zahlen (Kardinalität), die Einsicht in das Teile-Ganzes-Konzept sowie das Verständnis für die Relationen von Zahlen. Auf Grundlage dieser Einsichten und Kompetenzen, verbunden mit einem breiten Netz an automatisiertem arithmetischem Basiswissen, ist ein Kind fähig, Rechenaufgaben schnell und genau zu lösen. Die im Rahmen der vorliegenden Studie erarbeiteten CBM zusammen betrachtet erfassen die Ausprägung zentraler arithmetischer Einsichten auf Basis der Geschwindigkeit bei der Anwendung und geben ein Maß für die Flüssigkeit des Rechnens. Will man ein umfassenderes Bild der mathematischen Kompetenzen eines Schülers erstellen, müssen die erhobenen CBM-Daten durch weitere diagnostische Maßnahmen flankiert werden, z. B. Verfahren, welche neben arithmetischen auch weitere mathematische Aspekte, wie geometrische oder

stochastische Kompetenzen abfragen (siehe dazu NCRTI [2010] bzw. Moser Opitz [2009] in Kapitel 3).

10.5 Ausblick

Um die zukünftige Entwicklung von CBM als einen Aspekt der Prävention von Schwierigkeiten beim Erwerb der mathematischen Kompetenzen (primäre Prävention) bzw. als einen Teil der Unterstützung von Kindern mit festgestellten Schwierigkeiten (sekundäre und tertiäre Prävention) zu etablieren und weiterzuentwickeln, wäre eine langfristige und kooperative Zusammenarbeit zwischen Schulen, Universitäten und Bildungsbehörden erstrebenswert. Dabei kommen der Schulpraxis, der Bildungspolitik und der Forschung verschiede Aufgaben zu. Eine der Aufgaben von Schule besteht u. a. darin, sich Innovationen, wie bspw. dem regelmäßigen Einsatz von CBM im Unterricht, nicht zu verschließen und sich den damit einhergehenden Herausforderungen zu stellen. So führt auch die Kultusministerkonferenz (2004a) das Innovieren als einen Kompetenzbereich auf, der von einer Lehrkraft aufgrund der besonderen Verantwortung, die mit dem Berufsbild einhergeht, abverlangt wird. Konkret werden eine ständige Weiterentwicklung sowie das Mitwirken bei der Organisation und Umsetzung schulischer Entwicklungsprozesse erwartet. Eine derzeit hochaktuelle Innovation im Bereich der schulischen Bildung stellt bspw. die UN-Behindertenrechtskonvention dar. Durch die Verabschiedung der Behindertenrechtskonvention im Jahr 2006 und ihre Ratifizierung in Deutschland im Jahr 2009 verpflichten sich die Unterzeichnerstaaten, behinderten Menschen eine gleichberechtigte Teilhabe am gesellschaftlichen Leben, so bspw. auch im schulischen Umfeld, zu ermöglichen. Schule sieht sich damit der Herausforderung gegenüber, ein inklusionsförderliches Umfeld zu schaffen, das eine erfolgreiche gemeinsame Beschulung an der Regelschule gestattet. In diesem Kontext wurde in Mecklenburg-Vorpommern im Jahr 2012

eine Expertenkommission initiiert, welche den Auftrag hatte, grundlegende Fragen zur Gestaltung und schrittweisen Umsetzung schulischer Inklusion in Mecklenburg-Vorpommern zu klären (Expertenkommission "Inklusive Bildung in M-V bis zum Jahr 2020", 2012). Ein wesentliches Postulat des Gremiums stellt die Sicherstellung von qualitativ hochwertigem Unterricht dar. Dabei spielen, neben einer inklusionsförderlichen schulischen Gesamtatmosphäre, vor allem Aspekte der schulischen Prävention sowie einer inklusionsförderlichen Didaktik und Methodik eine zentrale Rolle, um allen Kindern hinreichende schulische Leistungen zu ermöglichen und schulischen Minderleistungen und Schulversagen vorzubeugen. Unter Rückbezug auf Hartke (2005) führt die Expertenkommission "Inklusive Bildung in M-V bis zum Jahr 2020" (2012) neben Ansatzpunkten, wie u. a. der Förderung schulischer Vorläuferfähigkeiten, dem lückenschließenden Lernen sowie der verstärkten Binnendifferenzierung „regelmäßige Lernstandserfassungen (Lernfortschrittsdokumentationen) zur Steigerung der Passung von Lernvoraussetzungen des Kindes, Unterricht und Förderung“ (S. 72) als wesentlich für eine gelingende schulische Prävention an. Somit reiht sich der regelmäßige Einsatz von CBM in Elemente inklusionsförderlicher Maßnahmen ein und nimmt einen zentralen Stellenwert bei der Gestaltung einer inklusiven Schule in Mecklenburg-Vorpommern ein. Damit CBM zu einer gelingenden schulischen Prävention und Inklusion beitragen, müssen jedoch einige Anforderungen erfüllt sein. Hinsichtlich des regelmäßigen Einsatzes von CBM im Unterricht muss Lehrkräften bewusst sein, dass das bloße Erheben von Schülerdaten allein nicht ausreichend ist, um eine Steigerung der Schülerleistungen zu erzielen. Viel wesentlicher ist die Frage, wie diese Informationen verwertet und in adäquate spezifische Förderung überführt werden können. Der in Abschnitt 3.1.4 beschriebene RtI-Ansatz als dynamisches Fördersystem erscheint ein vielversprechender Ansatz zur praktischen Umsetzung von Maßnahmen zur

Lernverlaufsdokumentation und darauf ausgerichteten Förderung. Die Befunde zur Effektivität dieses Konzepts bestätigen dies.

Bezogen auf den Erstunterricht im Bereich Mathematik in der Grundschule sollten Lehrkräfte wesentliche Bedingungsgefüge und Determinanten für die Entwicklung mathematischer Kompetenzen kennen. Ausgehend von diesem Wissen müssen Konsequenzen für Unterricht und Förderung gezogen werden. Dies wird auch in den Empfehlungen der Kultusministerkonferenz (2004a) deutlich:

> „*Das Ziel ist die Entwicklung eines gesicherten Verständnisses* [Herv. d. Verf.] *mathematischer Inhalte. Die allgemeinen mathematischen Kompetenzen verdeutlichen, dass die Art und Weise der Auseinandersetzung mit mathematischen Fragen ein wesentlicher Teil der Entwicklung mathematischer Grundbildung ist. Deren Entwicklung hängt nicht nur davon ab, welche* [Herv. d. Verf.] *Inhalte unterrichtet wurden, sondern in mindestens gleichem Maße davon, wie* [Herv. d. Verf.] *sie unterrichtet wurden, d. h. in welchem Maße den Kindern Gelegenheit gegeben wurde, selbst Probleme zu lösen, über Mathematik zu kommunizieren usw.*“ (S. 6).

Letzter Teil des Zitats verweist auf die Qualität des Unterrichts und damit auf die wesentliche Rolle und Verantwortung einer Lehrperson in der Vermittlung zentraler Kenntnisse und Kompetenzen. Hinzu kommt, dass sich guter Unterricht an den Lernvoraussetzungen der Schüler anpassen muss. Die Lehrkraft steht also in dem Spannungsfeld zwischen der hohen Heterogenität in den Lernausgangslagen und der Entscheidung über den Einsatz etablierter, sich als wirksam erwiesener Methoden (vgl. dazu auch Hattie, Beywl & Zierer, 2013) mit dem Ziel, dass möglichst alle Schüler von ihrem Unterricht profitieren. Eine Möglichkeit zur formativen Evaluation des Unterrichts und der Förderung stellen in regelmäßigem Abstand erhobene CBM-Daten dar (vgl. Abschnitt 3.2.6). Treten Lernentwicklungen nicht in dem avisierten Ausmaß ein,

ist dies ein Indikator dafür, dass der Unterricht bzw. die Förderung nicht ausreichend wirkt. D. h. bei der Anwendung von CBM darf der Grundgedanke der formativen Evaluation der eingesetzten pädagogischen Maßnahmen und Methoden nicht aus den Augen verloren werden.

Eine zentrale Aufgabe der Bildungspolitik besteht, neben der Qualitätsentwicklung und Standardsicherung, darin, Schule bei der Umsetzung ihrer Aufgaben zu unterstützen. Vor allem schulische Innovationen bedürfen besonderer Unterstützung. Vor dem Hintergrund der zuvor beschriebenen Inklusionsdebatte ergibt sich für die Bildungspolitik ein großer Handlungsdruck. Da nicht davon ausgegangen werden kann, dass Regelschullehrkräfte über ausreichend spezifisches sonderpädagogisches Fachwissen zur gelingenden Umsetzung einer präventiven und inklusiven Beschulung verfügen, müssen von Seiten der Politik Unterstützungssysteme bereitgestellt werden. Dabei stellt eine Möglichkeit die Aus-, Fort- und Weiterbildung des pädagogischen Personals dar. Dabei erscheint es essenziell, dass u. a. Aspekte einer qualitativ hochwertigen Unterrichts- und Fördergestaltung als auch Aspekte der pädagogischen Diagnostik Berücksichtigung finden. In diesem Rahmen sollten auch CBM als ein Teil einer objektiven, reliablen, validen als auch förderrelevanten Diagnostik thematisiert werden. Dies fordert auch die Expertenkommission "Inklusive Bildung in M-V bis zum Jahr 2020" (2012): „Ein hohes Maß an entwicklungspsychologischen und diagnostischen Kenntnissen und Fertigkeiten ist auch für Lehrkräfte allgemeiner Schulen wünschenswert“ (S. 83). Jedoch kann nicht erwartet werden, dass sich Regelschullehrkräfte in Aus- bzw. Weiterbildung umfassende Kenntnisse und Kompetenzen entsprechend eines ausgebildeten Sonderpädagogen aneignen und übergreifende Experten jeglicher (sonder-)pädagogischer Gesichtspunkte werden. In diesem Zusammenhang konstatieren D. Fuchs et al. (2012):

> “We believe it is naive to expect—and very bad policy to demand—that generalists will be cross-trained to teach skillfully to an academically diverse class of 28 children (primary prevention); to implement with fidelity a validated standard protocol to three to six students, some with behavior problems, while collecting and reviewing data on their progress (secondary prevention); and to use ‘experimental teaching’ with the most difficult-to teach children (tertiary prevention)” (S. 270).

Über die Aus-, Fort- und Weiterbildungsmaßnahmen hinaus sollten demnach auch zusätzliche externe Ressourcen, wie bspw. Stunden für Sonderpädagogen, Schulpsychologen oder externe Unterrichtsberater, an den Schulen bereitgestellt werden.

Genauso wünschenswert wie die Aus- und Fortbildung sowie der Einsatz von verschiedenen Spezialisten in der Schule wären weitere Hilfssysteme in Form von Materialdatenbanken, welche Übersichten über sinnvolle verfügbare Testverfahren, evidenzbasierte Förderkonzepte und -maßnahmen u. ä. zugänglich machen. Beispiele aus den USA wurden in Abschnitt 3.1.4.3 bereits genannt.

Zu den Aufgaben der Universität gehört es, die weitere Bildungsforschung zu konzipieren. Ein Hauptziel der Bildungsforschung sollte es sein, unterrichtsrelevante und -konforme Konzepte und Maßnahmen zur effektiven Leistungssteigerung aller Schüler zu entwickeln, zu evaluieren und der Praxis zugänglich zu machen. Dies gilt laut Hoge und Coladarci (1989) insbesondere für Testverfahren zur Objektivierung von Lehrereinschätzungen. Mit dem Ansatz des curriculumbasierten Messens ist in den USA in einer nunmehr über 30 Jahre andauernden Forschungstradition ein Konzept entstanden, was in der Schule ebenso tragfähig wie auch effektiv ist. Die vorliegende Arbeit liefert grundlegende Befunde für die CBM-Entwicklung, -erforschung und -etablierung in Deutschland. Grundsätzlich muss jedoch angemerkt werden, dass dieser Prozess im deutschsprachigen Raum noch „in den Kinderschuhen“ steckt.

Neben den in den vorangegangenen Abschnitten benannten anstehenden Schritten zur notwendigen, sich anschließenden Forschung der hier vorliegenden CBM, sind noch eine Vielzahl an allgemeinen Fragen offen. So bleibt es zu klären,

1. inwieweit sich weitere Verfahren entwickeln lassen, welche sich eignen, den Lernverlauf mathematischer Kompetenzen der Kinder differenziert zu dokumentieren. Erste Ansätze zur Erfassung weiterer mathematischer Basiskompetenzen wurden bspw. durch Methe et al. (2011) in den USA beschrieben (vgl. Abschnitt 9.3.1).
2. wie CBM für den Einsatz in höheren Klassenstufen konzipiert sein müssen. Auch hier gibt es erste Vorarbeiten, welche u. a. bei Foegen et al. (2007) aufgeführt sind.
3. inwieweit CBM für mathematische Aspekte über den Bereich der Arithmetik hinaus konzipiert werden können. Neben Aspekten der Arithmetik spielen bspw. noch die Geometrie, die Stochastik, das Sachrechnen und der Umgang mit Längen und Größen eine bedeutende Rolle im Mathematikunterricht in der Grundschule. So fragt bspw. das Verfahren des Typs concepts and applications vom CBM Monitoring basic skills progress (Fuchs et al., 1999) auch solche mathematischen Aspekte ab.
4. wie CBM bezüglich anderer Lernbereiche konzipiert sein müssen. Für den Bereich des Lesens gibt es hierzu in den USA (Colón & Kranzler, 2006; Foegen et al., 2007; Wayman et al., 2007) als auch in Deutschland (Strathmann & Klauer, 2012; Walter, 2010a) verschiedene Vorarbeiten. Jedoch erscheint es sinnvoll, auch CBM bspw. für den Englisch- oder anderweitigen Sprachunterricht zu konzipieren und zu evaluieren.
5. inwieweit sich die Vorhersagekraft von CBM durch das kombinierte Betrachten der Ergebnisse verschiedener Verfahren optimieren lässt (D. Fuchs et al., 2012).

6. inwieweit durch den Einsatz von CBM positive Effekte auf die Schülerleistungen, vor allem auch im Rahmen des deutschen Schulsystems, erreicht werden. Dies sollte im Rahmen von methodologisch kontrollierter, randomisierter Studien erfolgen.
7. welche Faktoren den Lernverlauf beeinflussen. Bspw. Christ et al. (2008) sehen hier ein wichtiges Forschungsfeld. Insbesondere sollten dabei Unterrichtsvariablen, wie z. B. der Grad und die Art der Differenzierung im Unterricht, jedoch auch Schülervariablen, wie bspw. die aktive Teilnahme am Unterricht, die Motivation des Schülers oder die verfügbare Zeit für Hausaufgaben bzw. Lernen berücksichtigt werden.
8. inwieweit die auf Durchschnittswerten basierenden erhobenen Angaben über zu erwartende Leistungsniveaus und -anstiege in CBM auch für einzelne Schüler anwendbar, also individuell übertragbar sind. In diesem Zusammenhang schlagen Kamata et al. (2012) vor, anhand von latenten Klassenanalysen zu prüfen, ob sich bestimmte typische Schülerprofile aus den CBM-Daten der Kinder identifizieren lassen. Anhand derartiger Erkenntnisse können differenzierte Hinweise hinsichtlich sinnvoller Fördermaßnahmen gewonnen werden. So könnten für typische Schülerprofile spezifische Förderprogramme erarbeitet werden. Zudem können sie weiteren Aufschluss darüber geben, wie sich Kinder mit gewissen Leistungsprofilen über die Zeit entwickeln. Somit könnten derzeitige Annahmen und Modellvorstellungen zur Entwicklung mathematischer Kompetenzen erweitert werden.

Die dargelegten Forschungsdesiderate zeigen, dass die CBM-Forschung im Allgemeinen, vor allem aber auch in Deutschland noch nicht abgeschlossen ist. Während zum einen die Neukonzeption von CBM-Verfahren eine zentrale Forschungsaufgabe darstellt, ergeben sich daraus zum anderen weiterführende Fragen, welche auf Grundlage von derartigen Lernverlaufsdaten Beantwortung finden

könnten. Insgesamt wird die Auffassung von Strathmann und Klauer (2010) geteilt, dass: „die Lernverlaufsdiagnostik ein weites und methodologisch umfangreiches wie herausforderndes Forschungsfeld zu öffnen vermag, das die Aufmerksamkeit von Forschern verdient“ (S. 121).

Abstract

Situationen und Probleme, in denen mathematische Kenntnisse und Kompetenzen erforderlich sind, begegnen dem Menschen heutzutage alltäglich, sei es im persönlichen oder beruflichen Alltag. Umso bedenklicher ist es, dass internationale Vergleichsstudien für fast ein Viertel der Schulkinder deutliche Schwierigkeiten im mathematischen Bereich belegen. Übertragen auf eine durchschnittliche Schulklasse zeigen demnach fünf bis sechs Schüler im Verlauf Ihrer Schullaufbahn Probleme im Fach Mathematik. Ein Risiko für ein erfolgreiches Erlernen des Rechnens in der Schule stellen frühe, nicht identifizierte Wissenslücken dar. Ohne entsprechende spezifische Förderung münden diese häufig in mathematischen Minderleistungen (Aunola, Leskinen, Lerkkanen & Nurmi, 2004; Helmke & Weinert, 1997; Krajewski, 2003; Mazzocco &Thompson, 2005). Ein wesentliches Element schulischer Prävention ist daher die frühe sowie valide Identifikation von Kindern mit Risiken in der Entwicklung mathematischer Kompetenzen (Hartke & Diehl, 2013). Zwar gibt es zu diesem Zweck diagnostische Verfahren, welche eine frühzeitige Prognose zukünftiger schulischer Minderleistungen ermöglichen sollen, aktuelle Forschungsergebnisse deuten jedoch darauf hin, dass dieses Ziel noch nicht in ausreichendem Maß erfüllt wird (Blumenthal, 2012). Vielmehr zeigen die Ergebnisse auf, dass langfristige Prognosen schulischer Leistungen aufgrund der hohen Komplexität der Bedingungs- sowie von Lehr-Lerngefügen nicht möglich und eher kurzfristige Vorhersagen auf Grundlage innerhalb kurzer Zeitintervalle wiederholter Leistungserhebungen sinnvoll sind. Der in der Literatur beschriebene Ansatz des formative assessment (Black & Wiliam, 1998b) stellt eine Reihe von Methoden zur Erfüllung dieses Zwecks bereit, von welchen vor allem die in den USA entwickelten und erforschten curriculum-based measurements (CBM; Deno, 1985) aufgrund ihrer engen Bindung an wissenschaftliche Gütekriterien in den Fokus der Forschung geraten sind. In einer nun etwa 30jährigen

Forschungstradition ist man in den USA bereits zu umfassenden Erkenntnissen zur Entwicklung, zur Evaluation und zum Einsatz von CBM in der schulischen Praxis gekommen. So liegen neben verschiedenen Wirksamkeitsnachweisen auf die schulischen Leistungen der Kinder (u. a. Black & Wiliam, 1998a, b; L.S. Fuchs & D. Fuchs, 1986; Kingston & Nash, 2011) auch programmatische Forschungskriterien zur Einschätzung der Güte von CBM vor (L.S. Fuchs, 2004), welche sich seither in der US-amerikanischen CBM-Forschung etabliert haben (Foegen, Jiban & Deno, 2007).

Das zentrale Ziel der vorliegenden Studie lag darin, zu prüfen, inwieweit Adaptionen US-amerikanischer CBM zur Prävention von Rechenschwächen auch im deutschen Schulsystem angewendet werden können. Besonderer Fokus liegt dabei auf einer möglichst frühen Identifikation von Kindern mit Risiken zur Ausprägung einer Rechenschwäche und der formativen Unterrichtsevaluation. Da sich die vorliegende Teilstudie im Kontext des umfassenden Forschungsprojekts „Rügener Inklusionsmodell“ unter der Leitung von Prof. Dr. Bodo Harke realisieren ließ, konnte auf Leistungsdaten von Erstklässlern des Jahrganges 2010/11 der Insel Rügen zurückgegriffen werden (N = 472). Ergänzend wurden die Daten verschiedener Schülerstichproben aus den Regionen Nordvorpommern (N = 700) und Rostock (N = 174) herangezogen.
Theoretische Basis der entwickelten CBM in der vorliegenden Studie bilden Erkenntnisse:

- zur Entwicklung mathematischer Kompetenzen im frühen Kindes- bis Schulalter,
- zur Konzeption und Evaluation von CBM im Allgemeinen sowie
- zur Konzeption und Evaluation speziell im Bereich der Erstmathematik der Schuleingangsstufe.

Auf Grundlage dieser umfassenden Vorarbeiten wurden sieben CBM mit jeweils mehreren Parallelversionen erarbeitet. Diese Verfahren erfassen Aspekte zum Mengen- und Zahlenwissen, zur Einsicht über Zusammenhänge von Mengen und Zahlen, zur Einsicht in das System der Zahlreihe, zur Einsicht in das Teile-Ganzes-Konzept, zum Verständnis zum Enthaltensein von Zahlen in anderen Zahlen sowie zum relationalen Verständnis der Zahlen zueinander. Diese Einsichten und Kompetenzen gelten als zentrale mathematische Meilensteine und lassen sich aktuellen Annahmen und Modellen der Entwicklungspsychologie sowie curricularen Vorgaben für den Mathematikunterricht in der ersten Klasse in Deutschland zuordnen (Fritz, Ricken & Gerlach, 2007; Krajewski & Schneider, 2006).

Die Erforschung der Verfahren erfolgte auf Grundlage der von L.S. Fuchs (2004) beschrieben Forschungskriterien. Diese sind auf drei Stufen systematisiert. Während Analysen der Stufe 1 sich auf die Hauptgütekriterien Objektivität, Reliabilität und Validität von CBM beziehen, werden auf Stufe 2 Untersuchungen zur Änderungsensibilität und auf Stufe 3 Analysen zur Nutzbarkeit und Effektivität der Verfahren durchgeführt.

Im Rahmen der Forschungsstufe 1 konnte ein hohes Maß an Objektivität der hier vorliegenden Verfahren festgehalten werden. Darüber hinaus konnten zufriedenstellende Kennwerte hinsichtlich der Splithalf-Reliabilität (Koeffizienten zwischen $r = .57$ und $r = .91$), der Retest-Reliabilität (Koeffizienten zwischen $r = .33$ und $r = .72$, mit zeitlichen Abständen zwischen 5 bis 25 Schulwochen) und der Paralleltest-Reliabilität (Koeffizienten zwischen $r = .65$ und $r = .94$) ermittelt werden. Die Einschätzung der Validität der CBM erfolgte auf Grundlage von Analysen zur Konstruktvalidität sowie zur Kriterumsvalidität. Korrelationen mit konvergenten Außenkriterien (Kalkulie, Fritz et al., 2007; Rechenfische, Wagner & Hartke, 2006; DEMAT 1+, Krajewski, Küspert, Schneider & Visé, 2002) liegen

zwischen r = .41 und r = .70. Korrelationen mit diskriminanten Außenkriterien (MÜSC, Mannhaupt, 2005; CBM Silbenlesen, Diehl, 2010a, b; WLLP, Küspert & Schneider, 1998) liegen zwischen r = .11 und r = .45. Zudem wurden RATZ-Indices zwischen 35 % und 95 % ermittelt, die auf eine gute bis sehr gute klassifikatorische Güte hindeuten. Insgesamt kennzeichnen sich die in der vorliegenden Studie erarbeiteten CBM durch zufriedenstellende Befunde hinsichtlich der wissenschaftlichen Hauptgütekriterien. Dennoch gibt es qualitative Unterschiede zwischen den Verfahren, die in Folgestudien weiterer Analysen bedürfen.

Innerhalb der Forschungsstufe 2 konnte auf Grundlage von Mehrebenenanalysen die Änderungssensibilität der sieben CBM-Verfahren nachgewiesen werden. Über die Messzeitpunkte ergab sich für jedes erforschte CBM ein signifikanter Leistungszuwachs der untersuchten Schüler. Es wurden durchschnittliche monatliche Anstiege für verschiedene Kindergruppen berechnet. Während vier der sieben CBM eine monatliche Lernverlaufsdokumentation von Kindern mit und ohne einem festgestellten Risiko im mathematischen Bereich zu Beginn der ersten Klasse ermöglichen, scheinen drei der Verfahren lediglich für eine längerfristige Überwachung des Lernverlaufs bzw. als Screeningverfahren geeignet.

Innerhalb einer Lehrerbefragung konnten im Rahmen der Forschungsstufe 3 Anhaltspunkte zur Nutzbarkeit, Ökonomie sowie Zumutbarkeit ermittelt werden. Dabei zeigte sich, dass der monatliche Einsatz von CBM im schulischen Alltag zwar als gewisser Mehraufwand erlebt wird, welcher sich jedoch durch den gewonnenen Nutzen rechtfertigt. So gibt ein Großteil der Lehrkräfte an, dass es ihnen auf Grundlage der CBM-Daten gelingt, rechenschwache Kinder zu identifizieren und Hinweise für Unterricht- und Förderung abzuleiten. Empirische Untersuchungen zur

Wirksamkeit der CBM stehen noch aus und sollten in weiteren Studien Berücksichtigung finden.

Unter Bezugnahme auf die hier ermittelten Befunde kann zusammenfassend konstatiert werden, dass es sich bei den hier untersuchten CBM um geeignete Verfahren zur Dokumentation kurz- bzw. mittelfristiger Lernverläufe im Erstunterricht Mathematik handelt, welche sich durch eine zufriedenstellende bis hohe wissenschaftliche Güte auszeichnen. Die Verfahren sind in besonderem Maße geeignet Kinder mit Risiken zur Ausprägung einer Rechenschwäche zu identifizieren und können als ein Element schulischer Prävention auch im deutschen Schulsystem eingesetzt werden. Trotz der positiven Befunde der Arbeit ist eine umfassende Forschungsarbeit im Bereich der CBM-Entwicklung im deutschsprachigen Raum weiterhin angezeigt.

Literatur

Aaron, P.G. (1997). The Impending Demise of the Discrepancy Formula. *Review of Educational Research, 67* (4), 461–502.

Achtenhagen, F. & Baumert, J. (1999). Alternative Evaluationsverfahren. *Unterrichtswissenschaft, 27*(2), 98–101.

Aebli, H. (1962). *Über die geistige Entwicklung des Kindes*. Stuttgart: Klett-Cotta.

Aebli, H. (1976). *Grundformen des Lehrens. Eine Allgemeine Didaktik auf kognitionspsychologischer Grundlage* (vollst. überarbeitete, stark erweiterte Neuaufl.). Stuttgart: Klett.

AIMSweb.com (2012). AIMSweb. Zugriff am 27.04.2012. Verfügbar unter http http://www.aimsweb.com.

Alarcón, M., DeFries, J.C., Light, J.G. & Pennington, B.F. (1997). A Twin Study of Mathematics Disability. *Journal of Learning Disabilities, 30* (6), 617–623.

Allinder, R.M. & BeckBest, M.A. (1995). Differential effects of two approaches to supporting teachers' use of curriculum-based measurement. *School Psychology Review, 24* (2), 287–298.

Allinder, R.M., Bolling, R.M., Oats, R.G. & Gagnon, W.A. (2000). Effects of teacher self-monitoring implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial and Special Education, 21,* 219–226.

Allinder, R.M., Fuchs, L.S., Fuchs, D. & Hamlett, C.L. (1992). Effects of Summer Break on Math and Spelling Performance as a Function of Grade Level. *The Elementary School Journal, 92* (4), 451–460.

Alonzo, J. & Tindal, G. (2009). *The development of K-8 progress monitoring measures in mathematics for use with the 2 % and general education populations: Kindergarten (Technical Report No. 0921).* University of Oregon, Behavioral Research and Teaching. Zugriff am 27.04.2012. Verfügbar unter: http://www.brtprojects.org/publications/dl/90.

Alonzo, J., Anderson, D. & Tindal, G. (2009). *IRT Analysis of General Outcome Measures in Grades 1-8. (Technical Report No. 0916).* Eugene, OR: Behavioral Research and Teaching, University of Oregon. Zugriff am 11.12.2012. Verfügbar unter http://www.brtprojects.org/publications/dl/88.

Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (4. Auflage). Heidelberg: Springer.

American Psychiatric Association. (2005). *Diagnostic and statistical manual of mental disorders. DSM-IV-TR* (4. Aufl.). Arlington, VA: American Psychiatric Association.

Amrein A.L. & Berliner, D.C. (2002). High-Stakes Testing & Student Learning. *Education Policy Analysis Archives, 10*(18). Zugriff am 27.04.2012. Verfügbar unter http://epaa.asu.edu/ojs/article/view/297.

Anderson, D., Lai, C.F., Nese, J.F.T., Park, B.J., Sáez, L., Jamgochian, E.M. et al. (2010). *Technical Adequacy of the easyCBM Primary-Level Mathematics Measures (Grades K-2), 2009-2010 Version (Technical Report No. 1006).* University of Oregon, Behavioral Research and Teaching. Zugriff am

27.04.2012. Verfügbar unter http://www.brtprojects.org/publications/dl/1016.

Antell, S. & Keating, D. (1983). Perception of numerical invariance in neonates. *Child Development, 54,* 695–701.

Ashcraft, M. H. & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning & Memory, 4* (5), 527–538.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44* (1-2), 75–106.

Assessment Reform Group. (2002). Assessment for Learning: 10 principles. Zugriff am 27.04.2012. Verfügbar unter http://www.assess-ment-reform-group.org.

Atkinson, R. & Shiffrin, R. (1968). Human Memory: A proposed system and its control processes. In J. T. Spence & K. W. Spence (Hrsg.), *The psychology of learning and motivation. Advances in research and theory: Volume 2* (The psychology of learning and motivation, Bd. 2, S. 89–195). New York: Academic Press.

Augst, G. (1989). *Schriftwortschatz: Untersuchungen und Wortlisten zum orthographischen Lexikon bei Schülern und Erwachsenen*. Frankfurt am Main: Peter Lang.

Aunola, K., Leskinen, E., Lerkkanen, M.-K. & Nurmi, J.-E. (2004). Developmental Dynamics of Math Performance From Preschool to Grade 2. *Journal of Educational Psychology, 96* (4), 699–713.

Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G. H. Bower (Hrsg.), *The psychology of learning and motivation. Advances in*

research and theory: Volume 8 (The psychology of learning and motivation, Bd. 8, 47–90). New York, London: Academic Press.

Baddeley, A.D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A.D. (1990). *Human memory. Theory and practice*. London: Erlbaum.

Baddeley, A.D. (2000). The episodic buffer: a new component of working memory? *Trends Cognitive Science, 4* (11), 417–423.

Baker, J.M. & Zigmond, N. (1990). Are regular education classes equipped to accommodate students with learning disabilities? *Exceptional Children, 56,* 515–526.

Bamonto Graney, S. & Shinn, M.R. (2005). Effects of Reading Curriculum-Based Measurement (R-CBM) Teacher Feedback in General Education Classrooms. *School Psychology Review, 34* (2), 184–201.

Bangert-Drowns, R.L., Kulik, J.A., and Morgan, M.T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Baraldi, A.N. & Enders, C.K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48* (1), 5–37.

Baroody, A. (1992). The development of preschoolers' counting skills and principles. In J. Bideaud, C. Meljac & J. Fischer (Hrsg.), *Pathways to number. Children's developing numerical abilities* (S. 99–126). Hillsdale, NJ: Erlbaum.

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *ZfE, 9* (4), 469–520.

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D. & Smith, T. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TMSS)*. Chestnut Hill MA: TIMSS International Study Center Boston College.

Becker, M., Lüdtke, O., Trautwein, U. & Baumert, J. (2006). Leistungszuwachs in Mathematik. Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? *Zeitschrift für Pädagogische Psychologie*, *20*, 233–242.

Behörde für Schule und Berufsbildung Hamburg. (2011). *Bildungsplan Grundschule. Mathematik.* Zugriff am 01.09.2012. Verfügbar unter http://www.hamburg.de/contentblob/2481796/data/mathematik-gs.pdf.

Bell, B. & Cowie, B. (2001). *Formative assessment and science education* (Bd. 12). Dordrecht, Boston: Kluwer Academic.

Bennett, R. E. (2009). *A critical look at the meaning and basis of formative assessment*. Princeton, NJ: ETS.

Berch, D.B. (2005). Making Sense of Number Sense. Implications for Children With Mathematical Disabilities. *Journal of Learning Disabilities,38*(4), 333-9.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *Behavior Analyst, 19,* 163–197.

Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., Wiesemes, R. & Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review, 1*, 61-67.

Black, P. & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5* (1), 7–74.

Black, P. & Wiliam, D. (1998b). Inside the black box. Raising standards through classroom assessment. *Phi Delta Kappan, 80* (2), 139–148.

Black, P. & Wiliam, D. (2003). 'In Praise of Educational Research': formative assessment. *British Educational Research Journal, 29*(5), 623-37.

Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational assessment, evaluation and accountability, 21* (1), 5–31.

Black, P.J., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for learning. Putting it into practice*. Maidenhead: Open University Press.

Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. et al. (2004). Mathematische Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47-92). Münster: Waxmann.

Blumenthal, Y. (2012). *Kognitive Prädiktoren schulischer Minderleistungen. Zur Möglichkeit valider Schulleistungsprognosen im frühen Grundschulalter*. Saarbrücken: SVH.

Bobis, J. (1991). The effect of instruction on the development of computation estimation strategies. *Mathematics Education Research Journal, 3*, 7-29.

Bortz, J. & Döring, N. (2006*). Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.

Bos., W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich.* Münster: Waxmann.

Boston, Carol (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation, 8*(9).

Brandtstädter, J. (1982a). Methodologische Grundfragen psychologischer Prävention. In J. Brandtstädter & A. von Eye (Hrsg.), *Psychologische Prävention. Grundlagen, Programme, Methoden* (S. 37-79). Stuttgart: Hans Huber.

Brandtstädter, J. (1982b). Prävention von Lern- und Entwicklungsproblemen im schulischen Bereich. In J. Brandtstädter & A. von Eye (Hrsg.), *Psychologische Prävention. Grundlagen, Programme, Methoden* (S. 275-302). Stuttgart: Hans Huber.

Broadfoot, P. & Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education, 11*(1), 7–27.

Broadfoot, P. (1998). Records of achievement and the learning society: A tale of two discourses. *Assessment in Education, 5*(3), 447–477.

Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrerinnens. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie Praxisgebiete Pädagogische Psychologie, Bd. 3, 177–214). Göttingen: Hogrefe Verlag.

Bruner, J. (1951). Personality dynamics and the process of perceiving. In R.R. Blake & G.V. Ramsey (Hrsg.), *Perception an approach to personality* (S. 121–147). New York: Ronald Press.

Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models. Applications and data analysis methods* (Advanced quantitative techniques in the social sciences, Bd. 1). Newbury Park: Sage.

Bühl, A. (2006). *SPSS 14.0. Einführung in die moderne Datenanalyse*. München: Pearson.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (ps Methoden/Diagnostik, 1. Aufl.). München: Pearson Studium.

Bull, R. & Johnston, R.S. (1997). Children's Arithmetical Difficulties: Contributions from Processing Speed, Item Identification, and Short-Term Memory. *Journal of Experimental Child Psychology, 65* (1), 1–24.

Bull, R., Johnston, R.S. & Roy, J. (1999). Exploring the Roles of the Visual–Spatial Sketch Pad and Central Executive in Children's Arithmetical Skills. *Views From Cognition and Developmental Neuropsychology, 15* (3), 421–442.

Burns, M.K. & Symington, T. (2002). A meta-analysis of pre-referral intervention teams: Student and systemic outcomes. *Journal of School Psychology, 40,* 437–447.

Burns, M.K. (2010). Response-to-Intervention Research: Is the Sum of the Parts as Great as the Whole? *Perspectives on Language and Literacy, 36* (2), 13–15.

Burns, M.K., Appleton, J.J. & Stehouwer, J.D. (2005). Meta-Analytic Review of Responsiveness-To- Intervention Research: Examining Field-Based and Research-Implemented Models. *Journal of Psychoeducational Assessment, 23* (4), 381–394.

Burns, M.K., VanDerHeyden, A.M. & Jiban, C. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review, 35*, 401-418.

Butterworth, B. & Reigosa, V. (2007). Information processing deficits in dyscalculia. In D. B. Berch & M. M. M. Mazzocco (Hrsg.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (S. 65–81). Baltimore, Md: Brookes.

Büttner, G. (2003). Gedächtnisentwicklung im Kindes- und Jugendalter. *Sprache Stimme Gehör, 27,* 24–30.

Campbell, D.T. & Fiske, D.W. (1959). Convergent and Discriminant Validation by the Multitrait Multimethod Matrix. *Psychological Bulletin, 56*(2),81-105.

Campbell, J.I.D. (1992). *The Nature and Origin of Mathematical Skills* (Advances in psychology, Bd. 91, 1. Aufl.). Amsterdamm: Elsevier textbooks.

Caplan, G. (1964). *Principles of preventive psychiatry*. New York: Basic Books Inc.

Carpenter, T. & Moser, J. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15,* 179–202.

Case, R. & Sowder, J. (1990). The development of computational estimation: A neo-Piagetian analysis. *Cognition and Instruction, 7*, 79-104.

Cattell, R.B., Weiß, R.H. & Osterland, J. (1997). Grundintelligenztest Skala 1. CFT 1 (5. Aufl.). Göttingen: Hogrefe Verlag.

Chard, D.J., Clarke, B., Baker, S., Otterstedt, J., Braun, D. & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30,* 3–14.

Chi, M. & Klahr, D. (1975). Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology, 19,* 434–439.

Christ, T.J. & Vining, O. (2006). Curriculum-Based Measurement Procedures to Develop Multiple-Skill Mathematics Computation Probes: Evaluation of Random and Stratified Stimulus-Set Arrangements. *School Psychology Review, 35* (3), 387–400.

Christ, T.J., Scullin, S., Tolbize, A. & Jiban, C.L. (2008). Implications of Recent Research. Curriculum-Based Measurement of Math Computation. *Assessment for Effective Intervention, 33* (4), 198–205.

Christ, T.J., Silberglitt, B., Yeo, S. & Cormier, D. (2010). Curriculum-Based Measurement of Oral Reading: An Evaluation of Growth Rates and Seasonal Effects among Students Served in General and Special Education. *School Psychology Review, 39* (3), 447–462.

Clarke, B. & Shinn, M.R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33* (2), 234–248.

Clearfield, M.W. & Mix, K.S. (2001). Amount Versus Number: Infants' Use of Area and Contour Length to Discriminate Small Sets. *Journal of Cognition and Development, 2* (3), 243–260.

Coalition for Evidence-Based Policy. (2002). *Bringing Evidence-Driven Progress To Education: A Recommended Strategy for the U.S. Department of Education.* Zugriff am 14.03.2012. Verfügbar unter http://evidencebasedpolicy.org/docs/Evid-based_educ_strategy_for_ED.pdf.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B. & Perlwitz, M., (1991). Assessment of a problem-centred second-grade mathematics project. *Journal for Research in Mathematics Education, 22*, 3-29.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology, 78,* 141–146.

Colón, E.P. & Kranzler, J.H. (2006). Effect of Instructions on Curriculum-Based Measurement of Reading. *Journal of Psychoeducational Assessment, 24* (4), 318–328.

Council Of Chief State School Officers (CCSSO) (2008a). *Attributes Of Effective Formative Assessment.* Zugriff am 03.11.2011. Verfügbar unter http://www.ccsso.org/Documents/2008/Attributes_of_Effective_2008.pdf.

Council Of Chief State School Officers (CCSSO) (2008b). *Formative Assessment: Examples of Practice.* Zugriff am 12.03.2012. Verfügbar unter http://www.ccsso.org/Documents/2008/Formative_Assessment_Examples_2008.pdf.

Crooks, T.J. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of*

Cumming, J.J. & Elkins, J. (1999). Lack of automaticity in the basic addition facts as a characteristic of arithmetic learning problems and instructional needs. *Mathematical Cognition, 5,* 149–180.

Das Ministerium für Bildung, J. u. S. L. B., Senatsverwaltung für Bildung, J. u. S. B., Senator für Bildung und Wissenschaft Bremen & Des Ministerium für Bildung, W. u. K. L. M. V. (2004). *Rahmenplan Grundschule Mathematik.* Zugriff am 01.09.2012.

Verfügbar unter http://www.bildungsserver-mv.de/download/rahmenplaene/rp-mathe-gs.pdf.

Das Ministerium für Bildung, W. F. u. K. L. S.-H. (2005). *Lehrplan Grundschule. Mathematik.* Zugriff am 01.09.2012. Verfügbar unter http://www.uni-flensburg.de/fileadmin/ms2/inst/mathematik/personen/lorenzen/Vorlesungen/Problemloesen/Mathematik__PS_-1.pdf.

Dehaene, S. & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsychologia, 29,* 1045–1074.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1–42.

Deno, S.L. & Fuchs, L.S. (1987). Developing curriculum-based measurement systems for databased special education problem solving. *Focus on Exceptional Children, 19* (8), 1–15.

Deno, S.L. & Mirkin, P.K. (1977). *Data-based Program Modification: A Manual*. Reston, VA: Data-based Program Modification: A Manual.

Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232.

Deno, S.L. (2003a). Curriculum-Based Measures: Development and Perspectives. *Assessment for Effective Intervention, 28* (3-4), 3–12.

Deno, S.L. (2003b). Developments in Curriculum-based Measurement. *Journal of Special Education, 37* (3), 184–192.

Deno, S.L., Mirkin, P.K. & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49* (1), 36–45.

Deutsches Cochrane Zentrum. (2011). *Von der Evidenz zur Empfehlung (Klassifikationssysteme).* Zugriff am 14.03.2012. Verfügbar unter http://www.cochrane.de/de/evidenz-empfehlung.

Diehl, K. & Hartke, B. (2007). Curriculumnahe Lernfortschrittsmessungen. *Sonderpädagogik, 37* (4), 195–211.

Diehl, K. & Hartke, B. (2012). *Inventar zur Erfassung der Lesekompetenz im 1. Schuljahr. IEL-1.* Göttingen: Hogrefe.

Diehl, K. (2010a). *CBM L-N1 Curriculumbasierte Messung Lesen Niveaustufe 1*. Universität Rostock, unveröffentlicht.

Diehl, K. (2010b). *CBM L-N2 Curriculumbasierte Messung Lesen Niveaustufe 2*. Universität Rostock, unveröffentlicht.

Diehl, K., Hartke, B. & Knopp, E. (2009). Curriculum-Based Measurement & Leerlingonderwijsvolgsystemm - Konzepte zur theoriegeleiteten Lernfortschrittsmessung im Anfangsunterricht Deutsch und Mathematik? *Zeitschrift für Heilpädagogik, 60* (4), 122–130.

Diehl, K., Mahlau, K., Voß, S. & Hartke, B. (2010). Das Rügener-Integrations-Modell (RIM). *Gemeinsam leben, 19* (3), 162-167.

Diehl, K., Mahlau, K., Voß, S. & Hartke, B. (2012). *Das Rügener Inklusionsmodell (RIM). Konzeption einer präventiven und inklusiven Grundschule nach dem Response to Intervention-Ansatz (RTI)*. Rostock: Universität Rostock.

Dilling, H., Mombour, W. & Schmidt, M.H. (2011). *Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F); klinisch-diagnostische Leitlinien* (8., überarb. Aufl.). Bern: Huber.

Ditton, H. (1998). Mehrebenenanalyse. Grundlagen und Anwendungen des Hierarchisch Linearen Modells. Weinheim: Juventa.

Dochy, F. (2001). A new assessment era: Different needs, new challenges. *Research Dialogue in Learning and Instruction, 2*, 11–20.

Donovan, S. & Cross, C. T. (2002). *Minority students in special and gifted education*. Washington, D.C: National Academy Press.

Drouin, M. (2010). Group-Based Formative Summative Assessment Relates to Improved Student Performance and Satisfaction. *Teaching of Psych., 37* (2), 114–118.

easyCBM.com (2012a). easyCBM. Zugriff am 27.04.2012. Verfügbar unter http://www.easycbm.com.

easyCBM.com (2012b). *Technical Reports.* Zugriff am 27.04.2012. Verfügbar unter http://www.brtprojects.org/publications/technical-reports.

Educational Research, 58 (4), 438–481.

Elawar, M.C., and Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behaviour a little rather than a lot. *Journal of Educational Psychology, 77*(2), 162-173.

Enders, C.K. (2010). *Applied missing data analysis* (Methodology in the social sciences). New York: Guilford Press.

Engle, R.W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23.

Epstein, M.H., Polloway, E.A. & Patton, J.R. (1989). Academic achievement probes: Reliability of measures for students with mild mental retardation. *Special Services in the Schools, 5* (1/2), 23–31.

Expertenkommission "Inklusive Bildung in M-V bis zum Jahr 2020". (2012). Zur Entwicklung eines inklusiven Bildungssystems in Mecklenburg- Vorpommern bis zum Jahr 2020. Bericht mit Empfehlungen der Expertenkommission "Inklusive Bildung in M-V bis zum Jahr 2020". Zugriff am 24.07.2013. Verfügbar unter http://service.mvnet.de/_php/download.php?datei_id=73004.

Farlow, L.J. & Snell, M.E. (1989). Teacher use of student performance data to make instructional decisions: Practices in programs for students with moderate to profound disabilities. *Journal of the Association for Persons with Severe Handicaps, 14,* 13–22.

Feigenson, L., Carey, S. & Hauser, M. (2002). The Representations Underlying Infants' Choice of More: Object Files Versus Analog Magnitudes. *Psychological Science, 13* (2), 150–156.

Feigenson, L., Carey, S. & Spelke, E. (2002). Infants' Discrimination of Number vs. Continuous Extent. *Cognitive Psychology, 44,* 33–66.

Fingerle, M. & Ellinger, S. (2008). Sonderpädagogische Förderprogramme im Vergleich. In M. Fingerle & S. Ellinger (Hrsg.),

Sonderpädagogische Förderprogramme im Vergleich. Orientierungshilfen für die Praxis (Heil- und Sonderpädagogik, S. 7–10). Stuttgart: Kohlhammer.

Fischer, F. (1990). A part-part-whole curriculum for teaching number to kindergarten. *Journal for Research in Mathematics Education, 21*, 207-215.

Fletcher, J.M., Coulter, W.A., Reschly, D.J. & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia*, 54 (2), 304-331.

Fletcher, J.M., Francis, D.J., Shaywitz, S.E., Lyon, G.R., Foorman, B.R., Stuebing, K.K. et al. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabil Res Pract, 13,* 186–203.

Fletcher, J.M., Lyon, G.R., Barnes, M., Stuebing, K.K., Francis, D.J., Olson, R.K. et al. (2002). Classification of learning disabilities: An evidence-based evaluation. In R. Bradley, L. C. Danielson & D. P. Hallahan (Hrsg.), *Identification of learning disabilities. Research to practice* (The LEA series on special education and disability, S. 185–250). Mahwah, NJ: Erlbaum.

Fletcher, J.M., Shaywitz, S.E., Shankweiler, D., Katz, L., Liberman, I., Fowler, A. et al. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86,* 6–23.

Foegen, A. & Deno, S.L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *The Journal of Special Education, 35,* 4–16.

Foegen, A., Jiban, C. & Deno, S.L. (2007). Progress Monitoring Measures in Mathematics: A Review of the Literature. *The Journal of Special Education, 41* (2), 121–139.

Fontana, D. & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology, 64* (3), 407–417.

Foorman, B.R., Francis, D.J., Fletcher, J.M. & Lynn, A. (1996). Relation of phonological and orthographic processing to early reading: Comparing two approaches to regression-based, reading-level-match designs. *Journal of Educational Psychology, 88,* 639–652.

Frederiksen, J. R., & White, B. (1996). *Reflective assessment: assessing students' research within an inquiry-based middle-school science curriculum. Technical Report CM-96-02*. UC Berkeley, School of Education.

Frey, A., Asseburg, R., Carstensen, H.C., Ehmke, T. & Blum, W. (2007) Mathematische Kompetenz.In PISA-Konsortium Deutschland (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie (S. 249-276)*. Münster: Waxmann.

Frey, B.B. & Schmitt, V.L. (2007). Coming to Terms With Classroom Assessment. *Journal of Advanced Academics, 18*(3), 402-423.

Fritz, A. & Ricken, G. (2008). *Rechenschwäche*. München: Reinhardt.

Fritz, A. & Ricken, G. (2009). Grundlagen des Förderkonzeptes »Kalkulie«. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Handbuch*

Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie (2. Aufl., S. 274–395). Weinheim: Beltz.

Fritz, A., Ricken, G. & Gerlach, M. (2007). *Diagnose- und Trainingsprogramm für rechenschwache Kinder. Handreichung zur Durchführung der Diagnose* (Kalkulie: Diagnose- und Trainingsprogramm für rechenschwache Kinder…, 1. Aufl., 1 Dr.). Berlin: Cornelsen.

Fuchs, D. & Deshler, D.D. (2007). What We Need to Know About Responsiveness To Intervention (and Shouldn't Be Afraid to Ask). *Learning Disabil Res Pract, 22* (2), 129–136.

Fuchs, D. & Fuchs, L.S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41* (1), 93–99.

Fuchs, D., Fuchs, L.S. & Compton, D.L. (2012). Smart RTI: A Next-Generation Approach to Multilevel Prevention. *Exceptional Children, 78*(3), 263–279.

Fuchs, D., Mock, D., Morgan, P.L. & Young, C.L. (2003). Responsiveness-to-Intervention: Definitions, Evidence, and Implications for the Learning Disabilities Construct. *Learning Disabil Res Pract, 18* (3), 157–171.

Fuchs, L. S. & Fuchs, D. (2007). *Progress Monitoring in the Context of Responsiveness-to-Intervention.* Zugriff am 12.03.2012. Verfügbar unter http://www.studentprogress.org/summer_institute/2007/RTI/RTIManual_2007.pdf.

Fuchs, L. S., Fuchs, D. & Hamlett, C. L. (1989b). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10,* 43–52.

Fuchs, L.S. & Deno, S.L. (1994). Must Instructionally Useful Performance Assessment Be Based in the Curriculum? *Exceptional Children, 6* (1), 15–24.

Fuchs, L.S. & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53,* 199–208.

Fuchs, L.S. & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21* (1), 45–58.

Fuchs, L.S. & Fuchs, D. (2005). *Using Curriculum-Based Measurement for Progress Monitoring in Math.* Zugriff am 21.03.2012. Verfügbar unter http://k12.wa.us/SpecialEd/pubdocs/RTI/CBM_Manual_Math_9_29_05.pdf.

Fuchs, L.S. (1982). *Special Education Practice in Evaluating Student Progress toward Goals*. Minneapolis, Minn: Institute for Research on Learning Disabilities.

Fuchs, L.S. (2003). Assessing Intervention Responsiveness: Conceptual and Technical Issues. *Learning Disabil Res Pract, 18* (3), 172–186.

Fuchs, L.S. (2004). The Past, Present, and Future of Curriculum-Based Measurement Research. *School Psychology Review, 33* (2), 188–192.

Fuchs, L.S., Deno, S.L. & Mirkin, P.K. (1984). The Effects of Frequent Curriculum-Based Measurement and Evaluation on Pedagogy, Student Achievement, and Student Awareness of Learning. *American Educational Research Journal, 21* (2), 449–460.

Fuchs, L.S., Fuchs, D. & Courey, S.J. (2005). Curriculum-Based Measurement of Mathematics Competence: From Computation to Concepts and Applications to Real-Life Problem Solving. *Assessment for Effective Intervention, 30* (2), 33–46.

Fuchs, L.S., Fuchs, D. & Hamlett, C.L. (1989a). Effects of Alternative Goal Structures within Curriculum-Based Measurement. *Exceptional Children, 55* (5), 429–438.

Fuchs, L.S., Fuchs, D. & Prentice, K. (2004). Responsiveness to Mathematical Problem-Solving Instruction: Comparing Students at Risk of Mathematics Disability With and Without Risk of Reading Disability. *Journal of Learning Disabilities, 37* (4), 293–306.

Fuchs, L.S., Fuchs, D., Hamlett, C.L. & Allinder, R. (1991). The contribution of skills analysis to curriculum-based measurement in spelling. *Exceptional Children, 57,* 443–452.

Fuchs, L.S., Fuchs, D., Hamlett, C.L. & Stecker, P.M. (1990). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review, 19,* 6–22.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Phillips, N.B. & Bentz, J. (1994). Classwide curriculum-based measurement: helping general educators meet the challenge of student diversity. *Exceptional Children, 60* (6), 518–537.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Thompson, A., Roberts, R., Kupek, P. et al. (1994). Technical features of a mathematics concepts and applications curriculum-based measurement system. *Diagnostica, 19*(4), 23–49.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Walz, L. & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*, 27–48.

Fuchs, L.S., Fuchs, D., Karns, K., Hamlett, C. & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal, 36* (3), 609-646.

Fuchs, L.S., Fuchs, D., Phillips, N.B., Hamlett, C.L. & Karns, K. (1995). Acquisition and transfer effects of classwide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review, 24,* 604–620.

Fuchs, L.S., Hamlett, C.L. & Fuchs, D. (1998). *Monitoring basic skills progress: Basic math computation (computer program manual)* (2. Aufl.). Austin, Texas: PRO-ED.

Fuchs, L.S., Hamlett, C.L. & Fuchs, D. (1999). *Monitoring basic skills progress: Basic math concepts and applications (computer program manual)*. Austin, Texas: PRO-ED.

Fuchs, L.S., Hamlett, C.L. & Stecker, P.M. (1991). Effects of Curriculum-Based Measurement and Consultation on Teacher Planning and Student Achievement in Mathematics Operations. *American Educational Research Journal, 28* (3), 617–641.

Fuson, K.C. (1988). *Children's counting and concepts of number* (Springer series in cognitive development). New York, NY, Berlin: Springer.

Gagné, R.M. (1962) The acquisition of knowledge. *Psychology Review, 69*(4), 355-365.

Garnett, K. & Fleischner, J.E. (1983). Automatization and Basic Fact Performance of Normal and Learning Disabled Children. *Learning Disability Quarterly, 6* (2), 223.

Garnett, K. (1992). Developing fluenca with basic number facts: Intervention for students with learning disabilities. *Learning Disabil Res Pract, 7,* 210–216.

Gasteiger, H. (2007). *Stand der mathematischen Kompetenzdiagnosen am Übergang von Kindertagesstätten und Grundschule und zukünftige Perspektiven.* Zugriff am 11.03.2012. Verfügbar unter http://www.bildung-brandenburg.de/transkigs/fileadmin/user/redakteur/Berlin/ExpertiseKompetenzdiagnosen_Gasteiger_10-2007.pdf.

Gaupp, N. (2003). *Dyskalkulie. Arbeitsgedächtnisdefizite und Defizite numerischer Basiskompetenzen rechenschwacher Kinder*. Berlin: Logos-Verl.

Gaupp, N., Zoelch, C. & Schumann-Hengsteler, R. (2004). Defizite numerischer Basiskompetenzen bei rechenschwachen Kindern der 3. und 4. Klassenstufe. *Zeitschrift für Pädagogische Psychologie, 18* (1), 31–42.

Geary, D.C. & Brown, S.C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and

mathematically disabled children. *Developmental Psychology, 27* (3), 398–406.

Geary, D.C. & Hoard, M.K. (2005). Learning disabilities in arithmetic and mathematics: Theoretical and empirical perspectives. In J. I. D. Campbell (Hrsg.), *Handbook of mathematical cognition* (S. 253–267). New York: Psychology Press.

Geary, D.C. (1990). Acomponential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 40,* 244–259.

Geary, D.C. (2004). Mathematics and Learning Disabilities. *Journal of Learning Disabilities, 37* (1), 4–15.

Geary, D.C., Bow-Thomas, C. & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: Acomparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology, 54,* 372–391.

Geary, D.C., Brown, S.C. & Samaranayake, V.A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27* (5), 787–797.

Geary, D.C., Hamson, C.O. & Hoard, M.K. (2000). Numerical and Arithmetical Cognition: A Longitudinal Study of Process and Concept Deficits in Children with Learning Disability. *Journal of Experimental Child Psychology, 77* (3), 236–263.

Gelman, R. & Gallistel, C.R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University.

Gelman, R. & Meck, E. (1983). Preschoolers' counting: Principle before skill. *Cognition, 13,* 343–359.

Gelman, R. & Meck, E. (1986). The notion of principle: The case of counting. In J. Hiebert (Hrsg.), *Conceptual and procedural knowledge. The case of mathematics* (2. Aufl., S. 29–57). Hillsdale, NJ: Erlbaum.

Gelman, R. (2000). The epigenesist of mathematical thinking. *Journal of Applied Developmental Psychology, 21,* 27–37.

Gelman, R., Meck, E. & Merkin, S. (1986). Young children's numerical competence. *Cognitive Development, 1,* 1–29.

Gersten, R. & Chard, D.J. (1999). Number Sense. Rethinking Arithmetic Instruction for Students with Mathematical Disabilities. *The Journal of Special Education, 33*(1), 18-28.

Gersten, R. & Dimino, J.A. (2006). RTI (Response to Intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly, 41* (1), 99–108.

Gersten, R., Chard, D.J., Jayanthi, M., Baker, S.K., Morphy, P. & Flojo, J. (2009). Mathematics Instruction for Students With Learning Disabilities: A Meta-Analysis of Instructional Components. *Review of Educational Research, 79* (3), 1202–1242.

Gersten, R., Jordan, N. & Flojo, J.R. (2005). Early Identification and Interventions for Students With Mathematics Difficulties. *Journal of Learning Disabilities, 38* (4), 293–304.

Gerster, H.-D. (2009). Schwierigkeiten bei der Entwicklung arithmetischer Konzepte. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Handbuch Rechenschwäche. Lernwege, Schwierigkeiten*

und Hilfen bei Dyskalkulie (Studium Paedagogik, 2. Aufl., 248–268). Weinheim: Beltz.

Goldhammer, F. & Hartig, J. (2012). Interpretation von Testresultaten und Testeichung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch, 2. Aufl., S. 173–201). Berlin, Heidelberg: Springer.

Goldman, S.R., Pellegrino, J.W. & Mertz, D.L. (1988). Extended Practice of Basic Addition Facts: Strategy Changes in Learning-Disabled Students. *Cognition and Instruction, 5* (3), 223–265.

Gölitz, D., Roick, T. & Hasselhorn, M. (2006). *Deutscher Mathematiktest für vierte Klassen (DEMAT 4+)*. Göttingen: Hogrefe Verlag.

Grassmann, M., Klunter, M., Köhler, E., Mirwald, E., Raudies, M. & Thiel, O. (2002). *Kinderleistungen - Lehrererwartungen* (Potsdamer Studien zur Grundschulforschung, Bd. 30). Potsdam: Universitätsverlag Potsdam.

Grassmann, M., Mirwald, E., Klunter, M. & Veith, U. (1995). Arithmetische Kompetenz von Schulanfängern - Schlußfolgerungen für die Gestaltung des Anfangsunterrichtes. *Sachunterricht und Mathematik in der Primarstufe, 23* (7), 302-303, 314-321.

Griffin, S.A., Case, R., & Siegler, R.S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure . In K. McGilly (Hrsg.), *Classroom lessons: Integrating cognitive theory and classroom practice* (S. 25-49). Cambridge, MA: MIT Press.

Groen, G.J. & Parkman, J.M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79,* 329–343.

Grube, D. & Barth, U. (2004). Rechenleistung bei Grundschülern. *Zeitschrift für Pädagogische Psychologie, 18* (3-4), 245–248.

Grube, D. (2005). Entwicklung des Rechnens im Grundschulalter. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Mathematikleistungen* (Tests und Trends, Bd. 4, S. 105–124). Göttingen: Hogrefe Verlag.

Grube, D. (2006). *Entwicklung des Rechnens im Grundschulalter. Basale Fertigkeiten, Wissensabruf und Arbeitsgedächtniseinflüsse* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 52). Münster: Waxmann (Univ., Habil.-Schr.--Göttingen, 2004).

Grube, D. (2008). Rechenschwäche. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (Bd. 10, 1. Aufl., S. 642–652). Hogrefe Verlag.

Grube, D. (2009). Kognitive Bedingungen der Rechenschwäche. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Handbuch Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie* (2. Aufl., S. 181–196). Weinheim: Beltz.

Grube, D., Barth, U. & Hasselhorn, M. (2002). Rechenhefte – Tests zur Erfassung von Routinerechnen und basalem arithmetischen Faktenwissen. Unveröffentlichtes Testverfahren, Georg-August-Universität Göttingen.

Grünke, M. (2006). Zur Effektivität von Fördermethoden bei Kindern und Jugendlichen mit Lernstörungen: Eine Synopse vorliegender Metaanalysen. Kindheit und *Entwicklung, 15*, 238–253.

Guttmann, E. (1937). Congenital arithmetic disability and acalculia (Henschen). Brirish Journal of Medical Psychology, 16, 16-35.

Hagen, T. & Hillenbrand, C. (2012). Effektive Lernförderung in der Schuleingangsphase. *Zeitschrift für Heilpädagogik, 63* (8), 323–334.

Hardesty, F.P. & Priester, H.J. (1996). *Hamburg-Wechsler-Intelligenztest für Kinder. HAWIK*. Bern: Huber.

Haring, N., Liberty, K. & White, O. (1980). Rules for data-based strategy decisions in instructional programs: Current research and instructional implications. In W. Sailor, B. Wilcox & L. Brown (Hrsg.), *Methods of instruction for severely handicapped students* (S. 159–192). Baltimore, Md: Brookes.

Harris, L. (2007). Employing formative assessment in the classroom. *Improving Schools, 10* (3), 249–260.

Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch, 2. Aufl., S. 144–171). Berlin, Heidelberg: Springer.

Hartke, B. & Diehl, K. (2013). *Schulische Prävention im Bereich Lernen. Problemlösungen mit dem RTI-Ansatz.* Stuttgart: Kohlhammer.

Hartke, B. & Koch, K. (2008): Qualitätsstandards von Prävention und Präventionsforschung. In: J. Borchert, B. Hartke & P. Jogschies (Hrsg.), *Frühe Förderung entwicklungsauffälliger Kinder und Jugendlicher* (37-56). Stuttgart: Kohlhammer.

Hartke, B. (2005). Schulische Prävention - welche Maßnahmen haben sich bewährt? *Zeitschrift für Heilpädagogik, 56* (12), 470–481.

Hartke, B. (2010). Lernen fördern. In B. Hartke, K. Koch & K. Diehl (Hrsg.), *Förderung in der schulischen Eingangsstufe* (Schulpädagogik, S. 19–54). Stuttgart: Kohlhammer.

Hartmann, E. (2008). Konzeption und Diagnostik von schriftsprachlichen Lernstörungen im Responsiveness-to-Intervention-Modell: eine kritische Würdigung. *Vierteljahreszeitschrift für Heilpädagogik und ihre Nachbargebiete, 77,* 123–137.

Hascher T. & Astleitner, H. (2007). Blickpunkt Lernprozess. In M. Gläser-Zikuda & T. Hascher (Hrsg.), *Lernprozesse dokumentieren, reflektieren und beurteilen. Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis* (S. 25–43). Bad Heilbrunn: Klinkhardt.

Hasselhorn, M. & Gold, A. (2006). Erfolgreiches Lernen als gute Informationsverarbeitung. In M. Hasselhorn & A. Gold (Hrsg.), *Pädagogische Psychologie. Erfolgreiches Lernen und Lehren* (Kohlhammer Standards Psychologie, 1. Aufl., S. 66–122). Stuttgart: Kohlhammer.

Hasselhorn, M. & Gold, A. (2009). Pädagogische Psychologie. Erfolgreiches Lernen und Lehren (2., durchgesehene Aufl.). Stuttgart: Kohlhammer.

Hasselhorn, M. & Grube, D. (2003). Das Arbeitsgedächtnis: Funktionsweise, Entwicklung und Bedeutung für kognitive Leistungsstörungen. *Sprache – Stimme – Gehör, 27*, 31-37.

Hasselhorn, M., Mähler, C. & Grube, D. (2008). Lernstörungen in Teilleistungsbereichen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie. [Lehrbuch]* (Grundlagen Psychologie, 6. Aufl., S. 769–778). Weinheim: Beltz.

Hasselhorn, M., Roick, T. & Gölitz, D. (2005). Stabilitäten und prognostische Validitäten von Mathematikleistungen: Eine Längsschnittstudie mit der DEMAT-Reihe in der Grundschule. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Mathematikleistungen* (Tests und Trends, Bd. 4, 187–198). Göttingen: Hogrefe Verlag.

Hattie, J., Beywl, W. & Zierer, K. (2013). *Lernen sichtbar machen*. Baltmannsweiler: Schneider-Verl. Hohengehren.

Havighurst, R.J. (1981). *Developmental tasks and education* (3. Aufl.). New York, NY: Longman.

Hein, J., Bzufka, M. & Neumärker, K.-J. (2000). The specific disorder of arithmetical skills. Prevalence studies in a rural and an urban population sample and their clinico-neuropsychological validation. *European Child & Adolescent Psychiatry, 33,* I87-I101.

Heller, K. & Hany, E. (2002). Standardisierte Schulleistungsmessungen. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (Beltz Pädagogik, 2. Aufl., S. 87–102). Weinheim: Beltz.

Helmke, A. & Weinert, F.E. (1997). Bedingungsfaktoren schulischer Leistungen. In F.E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie, Themenbereich D, Serie 1, Bd. 3, S. 71-176). Göttingen: Hogrefe Verlag.

Helmke, A. (1997). Individuelle Bedingungsfaktoren des Schulleistung: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 203-216). Weinheim: Beltz.

Helmke, A. (2009).Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts (2. Auflage). Seelze: Klett-Kallmeyer.

Helwig, R., Anderson, L. & Tindal, G. (2002). Using a concept-grounded, curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *Journal of Special Education, 36* (2), 102-112.

Hengartner, E. & Röthlisberger H. (1995). Rechenfähigkeit von Schulanfängern. In H. Brügelmann, H. Balhorn & L. Füssenich (Hrsg.), *Am Rande der Schrift. Zwischen Sprachenvielfalt und Analphabetismus* (Libelle Lesen und Schreiben, Bd. 6, S. 66–86). Lengwil am Bodensee: Libelle.

Hessisches Kultusministerium. (1995). *Rahmenplan Grundschule.* Zugriff am 01.09.2012. Verfügbar unter http://www.kultusministerium.hessen.de/irj/servlet/prt/portal/prtroot/slimp.CMReader/HKM_15/HKM_Internet/med/8e2/8e25e9fb-a45b-901b-e592-697ccf4e69f2,22222222-2222-2222-2222-222222222222,true.

Hoge, R. D. & Coladarci, T. (1989). Teacher-Based Judgments of Academic Achievement: A Review of Literature. *Review of Educational Research, 59* (3), 297–313.

Hojnoski, R. L., Gischlar, K. L. & Missall, K. N. (2009). Improving Child Outcomes With Data-Based Decision Making: Graphing Data. *Young Exceptional Children, 12* (4), 15–30.

Hope, J.A. & Sherrill, J.M. (1987). Characteristics of Unskilled and Skilled Mental Calculators. *Journal for Research in Mathematics Education, 18* (2), 98-111.

Hosp, M. K., Hosp, J. L. & Howell, K. W. (2007). *The ABCs of CBM. A practical guide to curriculum-based measurement* (The Guilford practical intervention in the schools series). New York: Guilford Press.

Huber, C. & Grosche, M. (2012). Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik, 63* (8), 312–322.

Humbach, M. (2008). *Arithmetische Basiskompetenzen in der Klasse 10. Quantitative und qualitative Analysen.* Berlin: Dr. Köster

Humphreys, L.G. (1960). Investigations of the simplex. *Psychometrika, 25* (4), 313–323.

Huntley-Fenner, G. & Cannon, E. (2000). Preschoolers' Magnitude Comparisons are Mediated by a Preverbal Analog Mechanism. *Psychological Science, 11* (2), 147–152.

Huntley-Fenner, G. (2001). Children’s understanding of number is similar to adults• and rats: Numerical estimation by 5…7-year olds. *Cognition, 78,* B27-B40.

Ingenkamp, K. & Lissmann, U.(2008). *Lehrbuch der pädagogischen Diagnostik* (6. Aufl.). Weinheim: Beltz.

Irvin, P.S., Park, B.J., Anderson, D., Alonzo, J. & Tindal, G. (2011). A *Cross-validation of easyCBM Reading Cut Scores in Washington: 2009-2010,* University of Oregon, Behavioral Research and Teaching. Zugriff am 05.07.2012. Verfügbar unter http://www.brtprojects.org/publications/dl/1033.

Jacobs, C. & Petermann, F. (2003). Dyskalkulie - Forschungsstand und Perspektiven. *Kindheit und Entwicklung, 12* (4), 197–211.

Jacobs, C. & Petermann, F. (2005). *Diagnostik von Rechenstörungen* (Kompendien psychologische Diagnostik, Bd. 7). Göttingen: Hogrefe Verlag.

Jacobs, C. & Petermann, F. (2007). *Rechenstörungen*. Göttingen: Hogrefe Verlag.

Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H. (1999). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreib-Schwierigkeiten (BISC)*. Göttingen: Hogrefe Verlag.

Jenkins, J.R. & O'Connor, R.E. (2000).Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. Danielson & D.P. Hallahan (Hrsg.), Identification of learning disabilities: Research to practice. (99-149). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Jiban, C.L. & Deno, S.L. (2007). Using Math and Reading Curriculum-Based Measurements to Predict State Mathematics Test Performance: Are Simple One-Minute Measures Technically Adequate? *Assessment for Effective Intervention, 32* (2), 78–89.

Jiban, C.L., Deno, S.L., & Foegen, A. (2009). Developing measures for monitoring progress in elementary grade mathematics: An investigation of desirable characteristics (Technical Report 5). Minneapolis, MN: University of Minnesota, College of Education and Human Development, Research Institute on Progress Monitoring.

Jiménez Gonzalez, J. & Garcia Espínel, A. (1999). Is IQ–achievement discrepancy relevant in the definition of arithmetic learning disabilities? *Learning Disability Quarterly, 22,* 291–301.

Jitendra, A.K., Sczesniak, E. & Deatline-Buchman, A. (2005). An exploratory validation of curriculum-based mathematical word problem-solving tasks as indicators of mathematics proficiency for third graders. *School Psychology Review, 34,* 358–371.

Jones, E. & Krause, J. (1988). The effectiveness of data-based instruction by student teachers in classrooms for pupils with mild learning handicaps. *Teacher Education and Special Education, 11,* 9–19.

Jordan, N., Hanich, L. & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal Experimental Child Psychology, 85,* 103–119.

Kamata, A., Nese, J.F.T., Patarapichayatham, C. & Lai, C.F. (2012). Modeling Nonlinear Growth With Three Data Points: Illustration With Benchmarking Data. *Assessment for Effective Intervention, 20* (10), 1–12.

Kaufmann, S. & Wessolowski, S. (2005). Zählendes Rechnen. *Grundschule – Mathematik, 7,* 18–23.

Kaufmann, S. & Wessolowski, S. (2006). *Rechenstörungen. Diagnose und Förderbausteine* (1. Aufl.). Seelze: Kallmeyer.

Kaufmann, S. (2003). *Früherkennung von Rechenstörungen in der Eingangsklasse der Grundschule und darauf abgestimmte remediale Maßnahmen*. Frankfurt am Main: Peter Lang.

Kelley, B., Hosp, J. L. & Howell, K. W. (2008). Curriculum-Based Evaluation and Math: An Overview. *Assessment for Effective Intervention, 33* (4), 250–256.

Kesting, F. (2005). *Mathematisches Vorwissen zu Schuljahresbeginn bei Grundschülern der ersten drei Schuljahre*. Hildesheim, Münster.

Kingston, N. & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice, 30* (4), 28–37.

Klafki, W. (1994). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik* (4. Aufl.). Weinheim: Beltz.

Klauer, K.J. (1972). Zur Theorie und Praxis des binomialen Modells lehrzielorientierter Tests. In K.J. Klauer, R. Frikke, M. Herbig, H. Rupprecht & F. Schott (Hrsg.), *Lehrzielorientierte Tests*. Düsseldorf: Schwann.

Klauer, K.J. (1978). Diagnostik im Lehr-Lern-Prozeß. In K. J. Klauer (Hrsg.), *Handbuch der pädagogischen Diagnostik* (Schwann-Handbuch, Bd. 4, S. 857–872). Düsseldorf: Schwann.

Klauer, K.J. (1992). In Mathematik mehr leistungsschwache Mädchen, im Lesen und Rechtschreiben mehr leistungsschwache

Jungen?: Zur Diagnostik von Teilleistungsschwächen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 24,* 48–65.

Klauer, K.J. (1994a). *Diagnose- und Förderblätter 2. Rechenfertigkeiten 2. Schuljahr*. Berlin: Cornelsen.

Klauer, K.J. (1994b). *Diagnose- und Förderblätter 3. Rechenfertigkeiten 3. Schuljahr*. Berlin: Cornelsen.

Klauer, K.J. (1994c). *Diagnose- und Förderblätter 4. Rechenfertigkeiten 4. Schuljahr*. Berlin: Cornelsen.

Klauer, K.J. (2006). Erfassung des Lernfortschritts durch curriculumbasierte Messung. *Heilpädagogische Forschung, 32* (1), 16–26.

Klauer, K.J. (2011). Lernverlaufsdiagnostik - Konzept, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik, 3* (3), 207–224.

Kleber, E., Kleber, G. & Hans, O. (1999). *Differentieller Leistungstest - KG - Test zur Erfassung des Leistungsverhaltens bei konzentrierter Tätigkeit im Grundschulalter* (2., korr.). Göttingen: Hogrefe Verlag.

Klieme, E., Neubrand, M. & Lüdke, O. (2001). Mathematische Grundbildung: Textkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al., (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 141-191). Opladen: Leske + Budrich.

Knapstein, K. & Spiegel, H. (1995). Testaufgaben zur Erhebung arithmetischer Vorkenntnisse zu Beginn des 1. Schuljahres. In G.

N. Müller & E. Wittmann (Hrsg.), *Mit Kindern rechnen* (Beiträge zur Reform der Grundschule, Bd. 96, S. 65–73). Frankfurt am Main: Arbeitskreis Grundschule.

Knopp, E. (2010a). Curriculum-based Measurement - eine Möglichkeit zur Prävention von Schwierigkeiten im Anfangsunterricht Mathematik? *Sonderpädagogische Förderung heute, 55* (1), 61-82.

Knopp, E. (2010b). *Theoretische Grundlagen, Konzeption und Güte des Inventars "Rechenfische". Ein Verfahren zur Dokumentation von Fortschritten beim Erlernen arithmetischer Kenntnisse im Anfangsunterricht Mathematik*. München: Verlag Dr. Hut.

Koch, K. (2005). Probleme im Bereich des mathematischen Lernens. In S. Ellinger & M. Wittrock (Hrsg.), *Sonderpädagogik in der Regelschule. Konzepte - Forschung - Praxis* (S. 279–298). Stuttgart: Kohlhammer.

Koch, K. (2008). Förderung mathematischer Kompetenzen. In M. Fingerle & S. Ellinger (Hrsg.), *Sonderpädagogische Förderprogramme im Vergleich. Orientierungshilfen für die Praxis* (Heil- und Sonderpädagogik, S. 85–108). Stuttgart: Kohlhammer.

Kovaleski, J.F. & Pedersen, J. (2008). Best practices in data analysis teaming. In A. Thomas & J. Grimes (Hrsg.), *Best practices in school psychology, V* (5. Aufl., S. 115–130). Bethesda, MD: National Association of School Psychologists.

Krajewski, K. & Schneider W. (2006). Mathematische Vorläuferfertigkeiten im Vorschulalter und ihre Vorhersagekraft für

die Mathematikleistungen bis zum Ende der Grundschulzeit. *Psychologie und Unterricht, 53,* 246–262.

Krajewski, K. (2003). *Vorhersage von Rechenschwäche in der Grundschule*. Hamburg, Würzburg: Dr. Kovac.

Krajewski, K., Küspert, P., Schneider, W. & Visé, M. (2002). *Deutscher Mathematiktest für erste Klassen. DEMAT 1+.* Göttingen: Hogrefe Verlag.

Krajewski, K., Liehm, S. & Schneider, W. (2004). *Deutscher Mathematiktest für zweite Klassen. DEMAT 2+.* Göttingen: Hogrefe Verlag.

Krapp, A. (1993). Diagnose und Prognose. In B. Weidenmann, A. Krapp, M. Hofer, G. L. Huber & H. Mandl (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (3. Aufl., S. 565-630). Weinheim: Beltz.

Kretschmann, R. (2003). Prävention und pädagogische Interventionen bei Beeinträchtigungen des Lernens. In A. Leonhardt & F.B. Wember (Hrsg.), *Grundfragen der Sonderpädagogik. Bildung – Erziehung – Behinderung* (S. 465-503). Weinheim: Beltz.

Kultusministerium Sachsen-Anhalt. (2007). *Fachlehrplan Grundschule. Mathematik.* Zugriff am 01.09.2012. Verfügbar unter http://www.bildung-lsa.de/pool/RRL_Lehrplaene/Entwuerfe/lpgsmathe.pdf.

Kultusministerkonferenz (2004a). *Standards Lehrerbildung KMK-Arbeitsgruppe.* Zugriff am 12.12.2012. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf.

Kultusministerkonferenz (2004b). *Bildungsstandards im Fach Mathematik für den Primarbereich*. München: Luchterhand. Zugriff am 11.12.2012. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Mathe-Primar.pdf.

Kurdek, L. & Sinclair, R. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93,* 451–455.

Küspert, P. & Schneider, W. (1998). Würzburger Leise Leseprobe. WLLP. Göttingen: Hogrefe Verlag.

Landerl, K., Bevan, A. & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9-year-old students. *Cognition, 93* (2), 99–125.

Langer, W. (2009). *Mehrebenenanalyse. Eine Einführung für Forschung und Praxis* (Studienskripten zur Soziologie, 2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden.

Lee, Y.-S., Lembke, E.S., Moore, D., Ginsberg, H. & Pappas, S. (2007). *Identifying Technically Adequate Early Mathematics Measures.* Zugriff am 10.12.2012. Verfügbar unter http://www.wirelessgeneration.com/pdf/white-papers/Math_Reseach_TechnicalPaper_2008_01.pdf.

Lembke, E.S. & Foegen, A. (2005). *Identifying indicators of early mathematics proficiency in Kindergarten and Grade 1 (Technical Report 6)*. Minneapolis, MN: University of Minnesota, College of

Education and Human Development, Research Institute on Progress Monitoring.

Lembke, E.S. & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and grade 1 students. *Learning Disabilities Research & Practice, 24*, 12-20.

Lembke, E.S. & Stecker, P.M. (2007). *Curriculum-Based Measurement in Mathematics: An Evidence-Based Formative Assessment Procedure*. Portsmouth, NH: Center on Instruction.

Lembke, E.S., Foegen, A., Whittaker, T. A. & Hampton, D. (2008). Establishing Technically Adequate Measures of Progress in Early Numeracy. *Assessment for Effective Intervention, 33* (4), 206–214.

Lenhard, W. (2010). *Auswertungshilfe zur Berechnung von Testkennwerten von Screeningverfahren,* Universität Würzburg: Lehrstuhl Psychologie IV. Zugriff am 04.07.2012. Verfügbar unter http://www.i4.psychologie.uni-wuerzburg.de/en/staff/dr_wolfgang_lenhard/materialien/.

Lewis, C., Hitch, G.J. & Walker, P. (1994). The Prevalence of Specific Arithmetic Difficulties and Specific Reading Difficulties in 9- to 10-year-old Boys and Girls. *J Child Psychol & Psychiat, 35* (2), 283–292.

Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.

Lin, P. & Humphreys, L.G. (1977). Predictions of Academic Performance in Graduate and Professional School. *Applied Psychological Measurement, 1* (2), 249–257.

Lorenz, J.H. & Radatz, H. (1993). *Handbuch des Förderns im Mathematikunterricht*. Hannover: Schroedel-Schulbuchverl.

Lorenz, J.H. (2003). *Lernschwache Rechner fördern. Ursachen der Rechenschwäche ; Frühhinweise auf Rechenschwäche ; diagnostisches Vorgehen* (Lehrer-Bücherei Grundschule, 1. Aufl.). Berlin: Cornelsen Scriptor.

MacMillan, D. L., Siperstein, G. N. & Gresham, F. M. (1998). A challenge to the viability of mild mental retardation as a diagnostic category. *Exceptional Children, 62,* 356–371.

Mahlau, K., Diehl, K., Voß, S. & Hartke, B. (Hrsg.). (2011). *Lernen nachhaltig fördern Klasse eins. Fortbildungseinheiten zur Gestaltung einer präventiven und integrativen Grundschule – Stand 15. März 2011*. Rostock: Universität Rostock.

Maier, U. (2010). Formative Assessment - ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitung für Erziehungswissenschaft, 13* (2), 293–308.

Mannhaupt, G. (2005). *MÜSC: Münsteraner Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten*. Berlin: Cornelsen.

Marx, H. & Lenhard, W. (2010). Diagnostische Merkmale von Screeningverfahren. In M. Hasselhorn & W. Schneider (Hrsg.), *Frühprognose schulischer Kompetenzen* (Tests und Trends, Bd. 9, Bd. 9, S. 68–84). Göttingen: Hogrefe Verlag.

Marx, H. (1992). Methodische und inhaltliche Argumente für und wider eine frühe Identifikation und Prädiktion von Lese-Rechtschreibschwierigkeiten. *Diagnostica, 38,* 249–268.

Mazzocco, M.M.M. & Thompson, R.E. (2005). Kindergarten Predictors of Math Learning Disability. *Learning Disabil Res Pract, 20* (3), 142–155.

Mazzocco, M.M.M. (2005). Challenges in Identifying Target Skills for Math Disability Screening and Intervention. *Journal of Learning Disabilities, 38* (4), 318–323.

Mazzocco, M.M.M., Feigenson, L. & Halberda, J. (2011). Preschoolers' Precision of the Approximate Number System Predicts Later School Mathematics Performance. *PLoS ONE, 6* (9), e23749.

McCloskey, M. & Macaruso, P. (1995). Representing and using numerical information. *American Psychologist, 50*, 351–363.

McCloskey, M., & Macaruso, P. (1995). Representing and using numerical information. *American Psychologist, 50*, 351-363

McCloskey, M., Caramazza, A. & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4* (2), 171–196.

McCurdy, B.L. & Shapiro, E.S. (1992). A Comparison of Teacher-, Peer-, and Self-Monitoring with Curriculum-Based Measurement in Reading Among Students with Learning Disabilities. *The Journal of Special Education, 26* (2), 162–180.

McIntosh, A., Reys, B.J., & Reys, R.E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics, 12*, 2—8.

McLean, J.F. & Hitch, G.J. (1999). Working memory impairments impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74,* 240–260.

McMillan, J.H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation, 7* (8). Zugriff am 27.04.2012. Verfügbar unter http://pareonline.net/getvn.asp?v=7&n=8.

Methe, S. A., Begeny, J. C. & Leary, L. L. (2011). Development of Conceptually Focused Early Numeracy Skill Indicators. *Assessment for Effective Intervention, 36* (4), 230–242.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63* (2), 81–97.

Miller, M.D., Linn, R.L. & Gronlund, N.E. (2008). Measurement and Assessment in Teaching (10. Aufl.). Upper Saddle River, NJ: Merrill/Prentice Hall..

Ministerium für Bildung, F. F. u. K. S. (2009). *Kernlehrplan Mathematik. Grundschule.* Zugriff am 01.09.2012. Verfügbar unter http://www.saarland.de/dokumente/thema_bildung/KLPGSMathematik.pdf.

Ministerium für Bildung, F. u. J. (2002). *Rahmenplan Grundschule. Allgemeine Grundlegung. Teilrahmenplan Mathematik.* Zugriff am 01.09.2012. Verfügbar unter http://grundschule.bildung-rp.de/fileadmin/user_upload/grundschule.bildung-rp.de/Downloads/Rahmenplan/Rahmenplan.pdf.

Ministerium für Kultus, J. u. S. B.-W. (2004). *Bildungsplan 2004 Grundschule.* Zugriff am 01.09.2012. Verfügbar unter http://www.bildung-staerkt-menschen.de/service/downloads/Bildungsstandards/GS/GS_M_bs.pdf.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2009). *Kernlehrpläne - Kompetenzen.* Zugriff am 01.09.2012. Verfügbar unter http://www.standardsicherung.schulministerium.nrw.de/lehrplaene/lehrplaene-gs/mathematik/lehrplan-mathematik/kompetenzen/.

Missall, K.N., Mercer, S.H., Martinez, R.S. & Casebeer, D. (2012). Concurrent and Longitudinal Patterns and Trends in Performance on Early Numeracy Curriculum-Based Measures in Kindergarten Through Third Grade. *Assessment for Effective Intervention, 37* (2), 95–106.

Moosbrugger, H. & Kelava, A. (Hrsg.). (2012). *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch, 2. Aufl.). Berlin, Heidelberg: Springer.

Moosbrugger, H. (2012). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch, 2. Aufl., S. 103–117). Berlin, Heidelberg: Springer.

Moser Opitz, E. (2001). *Zählen, Zahlbegriff, Rechnen. Theoretische Grundlagen und eine empirische Untersuchung zum mathematischen Erstunterricht in Sonderklassen* (Beiträge zur Heil- und Sonderpädagogik, Bd. 27). Bern: Haupt.

Moser Opitz, E. (2004). Dyskalkulie: Krankheit, Erfindung, Mythos, Etikett … ? Auseinandersetzung mit einem geläufigen, aber ungeklärten Begriff. *Vierteljahreszeitschrift für Heilpädagogik und ihre Nachbargebiete, 72,* 179–190.

Moser Opitz, E. (2007). *Rechenschwäche/Dyskalkulie. Theoretische Klärungen und empirische Studien an betroffenen Schülerinnen und Schülern* (Beiträge zur Heil- und Sonderpädagogik, Bd. 31, 1. Aufl.). Bern: Haupt.

Moser Opitz, E., Reusser, L., Moeri Müller, M., Anliker, B., Wittich, C. & Freesemann, O. (2010). *Basisdiagnostik Mathematik für die Klassen 4–8*. Bern: Huber.

Müller, C. & Hartmann, E. (2009). Lernfortschritte im Unterricht erheben – Möglichkeiten und Grenzen des curriculumbasierten Messens. *Schweizerische Zeitschrift für Heilpädagogik, 10,* 36–42.

Müller, C. & Hartmann, E. (2010). *Freiburger Aufgaben zur Lernfortschrittsdiagnostik (CBM) Mathematik*. Freiburg: Universität Freiburg (unveröffentlicht).

Müller, G.N. & Wittmann, E. (Hrsg.). (2004). *Das Zahlenbuch 1*. Klett.

Müller, R. (1984). *Deutscher Lesetest zur Frühdiagnose von Lesestörungen (DLF 1-2)*. Weinheim: Beltz.

Murphy, M.M., Mazzocco, M.M.M., Hanich, L. & Early, M.C. (2007). Cognitive Characteristics of Children With Mathematics Learning Disability (MLD) Vary as a Function of the Cutoff Criterion Used to Define MLD. *Journal of Learning Disabilities, 40* (5), 458–478.

Nachtigall, C. & Suhl U. (2002). *Der Regressionseffekt: Mythos und Wirklichkeit.* Zugriff am 08.12.2012. Verfügbar unter

http://www.metheval.uni-jena.de/materialien/reports/report_2002_02.pdf.

National Center on Response to Intervention (NCRTI) (2010). *Essential Components of RTI – A Closer Look at Response to Intervention.* Zugriff am 13.03.2012. Verfügbar unter http://www.rti4success.org/pdf/rtiessentialcomponents_042710.pdf.

National Center on Response to Intervention (NCRTI) (2012). *Progress Monitoring Tools.* Zugriff am 27.04.2012. Verfügbar unter http://www.rti4success.org/progressMonitoringTools.

National Institute for Literacy (2006). *What is Scientifically Based Research? A Guide for Teachers.* Zugriff am 14.03.2012. Verfügbar unter http://lincs.ed.gov/publications/pdf/science_research.pdf.

National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the national mathematics advisory panel,* U.S. Department of Education. Zugriff am 02.04.2012. Verfügbar unter http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf.

Natriello, G. (1987). The Impact of Evaluation Processes on Students. *Educational Psychologist, 22* (2), 155–175.

Naumann, C.L. (1999). *Schriftwortschatz: Untersuchungen und Wortlisten zum orthographischen Lexikon bei Schülern und Erwachsenen* (4. Aufl.). Weinheim: Beltz.

Nese, J.F.T., Biancarosa, G., Anderson, D., Lai, C.F., Alonzo, J. & Tindal, G. (2012). Within-year oral reading fluency with CBM: a comparison of models. *Reading and Writing, 25,* 887–915.

Nichols, S.L. & Berliner, D.C. (2007). *Collateral damage. How high-stakes testing corrupts America's schools*. Cambridge, Mass: Harvard Education Press.

Niedersächsisches Kultusministerium (2006). *Kerncurriculum für die Grundschule Schuljahrgänge 1-4. Mathematik.* Zugriff am 01.09.2012. Verfügbar unter http://db2.nibis.de/1db/cuvo/datei/kc_gs_mathe_nib.pdf.

Niemeyer, W. (1976). *Bremer Lesetest für 3. Und 4. Klassen (BLT 3-4)*. Bremen: Herbig.

Noddings, N. (2004). High stakes testing: Why? *Theory and Research in Education, 2* (3), 263–269.

Nunnery, J.A., Ross, S.M. & McDonald, A. (2006). A randomized experimental evaluation of the impact of accelerated reader/reading renaissance implementation on reading achievement in grades 3 to 6. *Journal of Education for Students Placed at Risk, 111*, 1-18.

Nußbeck, S. (2007). Evidenzbasierte Praxis: Ein Konzept für sonderpädagogisches Handeln? *Sonderpädagogik, 37,* 146–155.

OECD (2010). PISA 2009 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften (Band 1). München: Bertelsmann.

Okamoto, Y. & Case, R. (1996). Number Knowledge Test (NKT). Ann Arbor, MI: Society for Research in Child Development.

Ostad, S. (1997). Developmental differences in addition strategies: a comparison of mathematically disabledand mathematically normal children. *British Journal of Educational Psychology, 67,* 345–357.

Ostad, S. (1998). Developmental differences in solving simple arithmetic number-fact problems: A comparison of mathematically normal and mathematically disabled children. *Mathematical Cognition, 4,* 1–19.

Oxford Centre for Evidence-based Medicine (2012). *Levels of Evidence (March 2009).* Zugriff am 14.03.2012. Verfügbar unter http://www.cebm.net/index.aspx?o=1025.

Parkman, J. M. & Groen, G. J. (1971). Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology, 89* (2), 335–342.

Parmar, R. (2008). Lernbehinderung – Ein Forschungsüberblick zur Situation in den USA am Beispiel von Lesen und Mathematik. Vierteljahreszeitschrift für Heilpädagogik und ihre Nachbargebiete, 77, 43–56.

Parsons, S. & Bynner, J. (2005). *Does numeracy matter more?* Zugriff am 11.03.2012. Verfügbar unter http://scholar.google.de/scholar?cluster=17129067779343627177&hl=de&as_sdt=0&as_vis=1.

Passolunghi, M. & Siegel, L.S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in

arithmetic problem solving. *Journal of Experimental Psychology, 80,* 44–57.

Pawlik, K. (1982). *Diagnose der Diagnostik*. Stuttgart: Klett.

Peard, R. (2004). School Mathematical Achievement as a Predictor of Success in a First Year University Mathematics Foundation Unit. In I. Putt (Hrsg.), *Mathematics Education Research Group of Australasia 27th Conference Proceedings* (S. 422-429). Townsville: Mathematics Education Research Group of Australasia.

Pellegrino, J.W. & Goldman, S.R. (1987). Information processing and elementary mathematics. *Journal of Learning Disabilities, 57* (20), 23–32.

Pellegrino, J.W., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academic Press.

Perrez, M. (1991). Prävention, Gesundheits- und Entfaltungsförderung: Systematik und allgemeine Aspekte. In M. Perrez & U. Baumann (Hrsg.), *Klinische Psychologie. Band 2: Intervention* (S. 80-98). Stuttgart: Huber.

Phillips, N.B., Hamlett, C.L., Fuchs, L.S. & Fuchs, D. (1993). Combining classwide curriculum-based measurement and peer tutoring to help general educators provide adaptive education. *Learning Disabil Res Pract, 8,* 148–156.

Piaget, J. & Szeminska, A. (1975). *Die Entwicklung des Zahlbegriffs beim Kinde*. Stuttgart: Klett.

Piaget, J. (1964). *Rechenunterricht und Zahlbegriff. Die Entwicklung des kindlichen Zahlbegriffes und ihre Bedeutung für den Rechenunterricht*. Braunschweig: Westermann.

Plickat, H.H. (1987). *Deutscher Grundwortschatz* (3. Aufl.). Weinheim: Beltz.

Pospeschill, M. (2010). *Testtheorie, Testkonstruktion, Testevaluation*. München: Reinhardt.

Prenzel, M., Drechsel, B., Carstensen, C.H. & Ramm, G. (2004). PISA 2003 - eine Einführung. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 13-46). Münster: Waxmann.: Beltz.

Raudenbush, S.W., Bryk, A.S & Congdon, R. (2011). HLM 7 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

Reschly, D.J. (2005). LD identification: primary intervention, secondary intervention, then what? *Journal of Learning Disabilities, 38*, 510-515.

Reschly, D.J., Hosp, J.L. & Schmied, C. M. (2003). And miles to go.... State SLD Requirements and Authoritative Recommendations. Nashville, TN: National Research Center on Learning Disabilities.

Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist, 44* (2), 162–169.

Resnick, L. B. (1992). From protoquantities to operators: building mathematical competence on a foundation of everyday knowledge. In G. Leinhardt, R. Putnam & R. A. Hattrup (Hrsg.), *Analysis of*

arithmetic for mathematics teaching (S. 374–429). Hillsdale, NJ: Erlbaum.

Resnick, L. B. (2000). Developing mathematical knowledge. In P. K. Smith & A. D. Pellegrini (Hrsg.), *Psychology of education. Major themes* (Bd. 3, S. 223–238). London: Routledge/Falmer.

Reys, B.J. (1991). *Developing number sense: Addenda series, grades 5-8*. Reston, VA: National Council of Teachers of Mathematics.

Richardson, K. (2002). *Assessing math concepts*. Rowley, MA: Math Perspectives.

Ricken, G. (2009). Dyskalkulie. In A. Lohaus & H. Domsch (Hrsg.), Psychologische Förder- und Interventionsprogramme für das Kindes- und Jugendalter (S. 113-128). Heidelberg: Springer.

Ricken, G. (2011). „Kalkulie“ – ein entwicklungspsychologisch begründeter Förderansatz. In K. Mahlau, K. Diehl, S. Voß & B. Hartke (Hrsg.), *Lernen nachhaltig fördern Klasse eins. Fortbildungseinheiten zur Gestaltung einer präventiven und integrativen Grundschule – Stand 15. März 2011* (S. 243–255). Rostock: Universität Rostock.

Ricken, G., Fritz, A. & Balzer, L. (2011). Mathematik und Rechnen – Test zur Erfassung von Konzepten im Vorschulalter (MARKO-D) – ein Beispiel für einen niveauorientierten Ansatz. *Empirische Sonderpädagogik, 3*, 256-271.

Ricken, G., Fritz, A. & Balzer, L. (im Druck). *MARKO-D – Mathematik und Rechnen – Test zur Erfassung von Konzepten im Vorschulalter*. Göttingen: Hogrefe Verlag.

Rinkens, H.-D. & Eilerts, K. (2001). *Feldstudie zur beginnenden Rechenfertigkeit von Erstklässlern.* Zugriff am 11.03.2012. Verfügbar unter http://www.rinkens-hd.de/_data/ErstklaesslerFaeh.pdf.

Rinkens, H.-D. (2004). *Arithmetische Fähigkeiten am Schulanfang.* Zugriff am 11.03.2012. Verfügbar unter http://www.rinkens-hd.de/_data/AritFaeh.pdf.

Roick, T. & Hasselhorn, M. (2005). Der Kettenrechner 3-4. Zusätzliche Differenzierung durch komplexe arithmetische Faktenaufgaben. In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), Diagnostik von Mathematikleistungen. Jahrbuch der pädagogisch-psychologischen Diagnostik, N. F. Band 4 (233-250). Göttingen: Hogrefe.

Roick, T. (2007). Kognitive Determinanten unterschiedlicher Rechenleistungen in der Primarstufe: Arbeitsgedächtnismerkmale und Aufmerksamkeitsaspekte. Dissertation, Georg-August-Universität Göttingen.

Roick, T., Göhlitz, D. & Hasselhorn, M. (2004). *Deutscher Mathematiktest für dritte Klassen (DEMAT 3+)*. Göttingen: Hogrefe Verlag.

Ross, S. (1989). Parts, wholes, and place value: A developmental view. *Arithmetic Teacher, 36*, 47-51.

Rost, D.H. (2007). *Interpretation und Bewertung pädagogisch-psychologischer Studien* (2.Aufl.). Weinheim: Beltz.

Ruiz-Primo, M.A. & Furtak, E.M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *J. Res. Sci. Teach., 44* (1), 57–84.

Sacher, W. (2006). Schulische Leistungsdiagnose. Grundlagen und Probleme. *Schulmagazin 5 bis 10, 74*(3), 5-8.

Sächsisches Staatsministerium für Kultus. (2009). *Lehrplan Grundschule. Mathematik.* Zugriff am 01.09.2012. Verfügbar unter http://195.37.90.111/apps/lehrplandb//downloads/lehrplaene/lp_gs_mathematik_2009.pdf.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B. & Richardson, W.S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ, 312* (7023), 71–72.

Salvia, J., Ysseldyke, J.E. & Bolt, S. (2007). *Assessment in special and inclusive education* (10. Aufl.). Boston: Houghton Mifflin Company.

Sandall, S. R., Schwartz, I. S. & Lacroix, B. (2004). Interventionists' Perspectives about Data Collection in Integrated Early Childhood Classrooms. *Journal of Early Intervention, 26* (3), 161–174.

Sawyer, R. J., Graham, S. & Harris, K. R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology, 84* (3), 340–352.

Scherer, P. & Moser Opitz, E. (2010). *Fördern im Mathematikunterricht der Primarstufe* (Mathematik Primar- und Sekundarstufe). Heidelberg: Spektrum Akademischer Verlag.

Schermelleh-Engel, K. & Schweizer, K. (2012). Multitrait-Multimethod-Analysen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch, 2. Aufl., S. 345–362). Berlin, Heidelberg: Springer.

Schermelleh-Engel, K. & Werner, C. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch, 2. Aufl., S. 114–133). Berlin, Heidelberg: Springer.

Schipper, W. (1996). Kompetenz und Heterogenität im arithmetischen Anfangsunterricht. *Die Grundschulzeitschrift, 10* (96), 10–15.

Schipper, W. (1998). Schulanfänger verfügen über hohe mathematische Kompetenzen - Eine Auseinandersetzung mit einem Mythos. In A. Peter-Koop (Hrsg.), *Das besondere Kind im Mathematikunterricht der Grundschule* (S. 119–140). Offenburg: Mildenberger.

Schipper, W. (2003). Kompetenzentwicklung beim schriftlichen Rechnen. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie ; ein Handbuch* (Beltz-Handbuch, S. 98–115). Weinheim: Beltz.

Schmidt, R. (1982). *Zahlenkenntnisse von Schulanfängern. Ergebnisse einer zu Beginn des Schuljahres 1981/82 durchgeführten Untersuchung*. Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung.

Schmidt, S. (2009). Arithmetische Kenntnisse am Schulanfang. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Handbuch Rechenschwäche.*

Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie (Studium Paedagogik, 2. Aufl., S. 77–97). Weinheim: Beltz.

Schmithorst, V.J. & Brown, R.D. (2004). Empirical validation of the triple-code model of numerical processing for complex math operations using functional MRI and group Independent Component Analysis of the mental addition and subtraction of fractions. *NeuroImage, 22* (3), 1414–1420.

Schrader, F.-W. & Helmke, A. (2002). Alltägliche Leistungsbeurteilung durch Lehrer. Achievement evaluation by teachers. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (Beltz Pädagogik, 2. Aufl., S. 45–58). Weinheim: Beltz.

Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 95–100). Weinheim: Beltz.

Schuchardt, K., Kunze, J., Grube, D. & Hasselhorn, M. (2006). Arbeitsgedächtnisdefizite bei Kindern mit schwachen Rechen- und Schriftsprachleistungen. *Zeitschrift für Pädagogische Psychologie, 20* (4), 261–268.

Schuchardt, K., Roick, T., Mähler, C. & Hasselhorn, M. (2008). Unterscheidet sich die Struktur des Arbeitsgedächtnisses bei Schulkindern mit und ohne Lernstörung? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 40* (3), 147–151.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Hrsg.), *Perspectives of curriculum evaluation* (S. 39–83). Chicago, IL: Rand McNally.

Selter, C. (1995). Zur Fiktivität der "Stunde Null" im arithmetischen Anfangsunterricht. *Mathematische Unterrichtspraxis, 2,* 11–19.

Senate and House of Representatives of the United States of America. (2004). Individuals with Disabilities Education Improvement Act of 2004. IDEA. Zugriff am 27.04.2012. Verfügbar unter http://nichcy.org/wp-content/uploads/docs/PL108-446.pdf.

Shalev, R.S., Auerbach, J., Manor, O. & Gross-Tsur, V. (2000). Developmental dyscalculia: prevalence and prognosis. *European Child & Adolescent Psychiatry, 9* (S2), 58-64.

Shalev, R.S., Manor, O. & Gross-Tsur, V. (2005). Developmental dyscalculia: a prospective six-year follow-up. *Developmental Medicine & Child Neurology, 47* (2), 121–125.

Shalev, R.S., Manor, O., Kerem, B., Ayali, M., Badichi, N., Friedlander, Y. et al. (2001). Developmental Dyscalculia Is a Familial Learning Disability. *Journal of Learning Disabilities, 34* (1), 59–65.

Shapiro, E.S. (1996). *Curriculum-based measurement procedures to develop multiple-skill mathematics computation probes: Evaluation of random and stratified stimulus-set arrangements* (2. Aufl.). New York

Shapiro, E.S. (2004). *Academic skills problems. Direct assessment and intervention* (The Guilford school practitioner series, 3. Aufl.). New York: Guilford Press.

Shapiro, E.S., Edwards, L. & Zigmond, N. (2005). Progress monitoring of mathematics among students with learning disabilities. *Assessment for Effective Intervention, 30* (2), 15–32.

Shapiro, E.S., Keller, M.A., Lutz, J.G., Edwards Santoro, L. & Hintze, J. M. (2006). Curriculum-Based Measures and Performance on State Assessment and Standardized Tests: Reading and Math Performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24* (1), 19–35.

Shepard, L.A. (2010). Commentary: Evaluating the Validity of Formative and Interim Assessment. *Educational Measurement: Issues and Practice, 28* (3), 32–37.

Shin, J., Espin, C. A., Deno, S.L. & McConnell, S. (2004). Use of hierarchical linear modeling and curriculum-based measurement for assessing academic growth and instructional factors for students with learning difficulties. *Asia Pacific Education Review, 5* (2), 136–148.

Shinn, M.R. & Bamonto, S. (1998). Advanced Applications of Curriculum-Based Measurement: "Big Ideas" and Avoiding Confusion. In M. R. Shinn (Hrsg.), *Advanced applications of curriculum-based measurement* (The Guilford school practitioner series, S. 1–31). New York: Guilford Press.

Shinn, M.R. (1995). Using Curriculum-Based Measurement in a Problem-Solving Model. In A. Thomas & J. Grimes (Hrsg.), *Best practices in school psychology, II* (2. Aufl., S. 547–568). Bethesda, MD: National Association of School Psychologists.

Shinn, M.R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Hrsg.), *Best practices in school psychology, IV* (4. Aufl., S.

671–698). Bethesda, MD: National Association of School Psychologists.

Shinn, M.R., Bamonto, S., Cornachione, C., Parker, C., Peterson, K., Thurber, R. et al. (1997). *Curriculum-based measurement and problem-solving assessment: Training modules* (5. Aufl.). Eugene, OR

Shinn, M.R., Walker, H.M. & Stoner, G. (2006). *Interventions for academic and behavior problems* (4. Aufl.). Bethesda, MD: NASP.

Siegel, L.S. (1989a). IQ Is Irrelevant to the Definition of Learning Disabilities. *Journal of Learning Disabilities, 22* (8), 469–478.

Siegel, L.S. (1989b). Why We Do Not Need Intelligence Test Scores in the Definition and Analyses of Learning Disabilities. *Journal of Learning Disabilities, 22* (8), 514–518.

Siegler, R.S. & Jenkins, E. (1989). *How children discover new strategies* (Bd. 1989). Hillsdale, NJ: Erlbaum.

Siegler, R.S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology, 116,* 250–264.

Simmons, D.C., Coyne, M.D., Kwok, O.-M., McDonagh, S., Harn, B.A. & Kame'enui, E.J. (2008). Indexing Response to Intervention: A Longitudinal Study of Reading Risk From Kindergarten Through Third Grade. *Journal of Learning Disabilities, 41* (2), 158–173.

Simpson, R.L., LaCava, P.G. & Sampson Graner, P. (2004). The No Child Left Behind Act: Challenges and Implications for Educators. *Intervention in School and Clinic, 40* (2), 67–75.

Sophian, C. (1998). A developmental perspective on children's counting. In C. Donlan (Hrsg.), *The development of mathematical skills* (Studies in developmental psychology, S. 27–46). Hove: Psychology Press.

Souvignier, E. & Förster, N. (2011). Effekte prozessorientierter Diagnostik auf die Entwicklung der Lesekompetenz leseschwacher Viertklässler. *Empirische Sonderpädagogik, 3*(3), 243-255.

Souvignier, E., Förster, N. & Salaschek, M. (2010). *Ergebnisbericht des "quop"-Forschungsprojekts zur Lernverlaufsdiagnostik für das Schuljahr 2009/10.* Zugriff am 06.04.2012. Verfügbar unter http://quop.de/fileadmin/uploads/quop/literature/Souvignier_Wissenschaftlicher_Ergebnisbericht_quop_2010.pdf.

Sowder, J. (1988). Mental computation and number comparison: Their roles in the development of number sense and computational estimation. In J. Heibert & M. Behr (Hrsg.), Research Agenda for Mathematics Education: Number Concepts and Operations in the Middle Grades (192-197). Hillsdale, NJ: Lawrence, Erlbaum & Reston.

Sowder, J.T. (1992). Making sense of numbers in school mathematics. G. Leinhardt, R.T. Putnam, & R.A. Hattrup (Hrsg.), *Analysis of arithmetic for mathematics teaching* (S. 1-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Spellings, M. (2005). Key Policy Letters Signed by the Education Secretary or Deputy Secretary. Zugriff am 12.03.2012. Verfügbar unter http://www2.ed.gov/policy/elsec/guid/secletter/051121.html.

Staatsinstitut für Schulqualität und Bildungsforschung München. (2000). *Grundschullehrplan Jgst. 1-2.* Zugriff am 01.09.2012. Verfügbar unter http://www.isb.bayern.de/isb/download.aspx?DownloadFileID=e3e56d177fef177dcfcd7910ffa68941.

Stanovich, K.E. & Siegel, L.S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86* (1), 24–53.

Stanovich, P.J. & Stanovich, K.E. (2003). *Using research and reason in education: How teachers can use scientifically based research to make curricular and instructional decisions*. Washington, D. C.: United States Department of Education.

Starkey, P. & Cooper, R.G. (1980). Perception of numbers by human infants. *Science, 210* (4473), 1033–1035.

Starkey, P. & Cooper, R.G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology, 13* (4), 399–420.

Starkey, P., Spelke, E. & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36,* 97–127.

Stecker, P.M. & Fuchs, L.S. (2000). Effecting Superior Achievement Using Curriculum-based Measurement: The Importance of Individual Progress Monitoring. *Learning Disabil Res Pract, 15* (3), 128–134.

Stecker, P.M., Fuchs, L.S. & Fuchs, D. (2005). Using Curriculum-Based Measurement to Improve Student Achievement: Review of Research. *Psychol. Schs., 42* (8), 795–819.

Stern, E. (1994). Die Erweiterung des mathematischen Verständnisses mit Hilfe von Textaufgaben. *Die Grundschule, 3,* 23–25.

Stern, E. (1997). Erwerb mathematischer Kompetenzen: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 157–170). Weinheim: Beltz.

Stern, E. (1998). *Die Entwicklung des mathematischen Verständnisses im Kindesalter*. Lengerich: Pabst.

Stern, E. (2003). Früh übt sich - Neuere Ergebnisse aus der LOGIK-Studie zum Lösen mathematischer Textaufgaben. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie* (Beltz-Handbuch, S. 116–130). Weinheim: Beltz.

Strathmann, A.M. & Klauer, K.J. (2008). Diagnostik des Lernverlaufs. Eine Pilotstudie am Beispiel der Entwicklung der Rechtschreibkompetenz. *Sonderpädagogik, 38* (1), 5–24.

Strathmann, A.M. & Klauer, K.J. (2010). Lernverlaufsdiagnostik: Ein Ansatz zur längerfristigen Lernfortschrittsmessung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42* (2), 111–122.

Strathmann, A.M. & Klauer, K.J. (2012). *Lernverlaufsdiagnostik Mathematik. LVD-M*. Göttingen: Hogrefe Verlag.

Strathmann, A.M., Klauer, K.J. & Greisbach, M. (2010). Lernverlaufsdiagnostik - Dargestellt am Beispiel der Entwicklung der Rechtschreibkompetenz in der Grundschule. *Empirische Sonderpädagogik, 2* (1), 64–77.

Strauss, M. & Curtis, L. (1981). Infant Perception of Numerocity. *Child Development, 52,* 1146–1152.

Suchodoletz, W. v. (2005). Früherkennung einer Lese-Rechtschreibstörung. In W. v. Suchodoletz (Hrsg.), *Früherkennung von Entwicklungsstörungen* (S. 191-222). Göttingen: Hogrefe Verlag.

Swanson, H. & Sachse-Lee, C. (2001). A Subgroup Analysis of Working Memory in Children with Reading Disabilities: Domain-General or Domain-Specific Deficiency? *Journal of Learning Disabilities, 34* (3), 249–263.

Swanson, H. (1999). Reading research for students with LD: A meta-analysis in intervention outcomes. *Journal of Learning Disabilities, 32,* 504–532.

Swanson, H. (2000). What instruction works for students with learning disabilities ? Summarizing the results from a meta-analysis of intervention studies. In R. Gersten, E. Schiller & S. Vaughn (Hrsg.), *Contemporary special education research: Synthesis of the knowledge base on critical instructional issues* (S. 1–30). Mahwah, NJ: Erlbaum.

Swanson, H., Hoskyn, M. & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes.* New York: Guilford Press.

Swets, J.A., Dawes, R.M. & Monahan, J. (2000). Better Decisions through Science. *Scientific American* (283), 82–87.

Temple, C.M. & Sherwood, S. (2002). Representation and retrieval of arithmetical facts: Developmental difficulties. *The Quarterly Journal of Experimental Psychology Section A, 55* (3), 733–752.

The Center for Public Education. (2006). *A guide to standardized testing – The nature of assessment*. Zugriff am 27.04.2012. Verfügbar unter http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/A-guide-to-standardized-testing-The-nature-of-assessment.

Thurber, R.S., Shinn, M.R. & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review, 31* (4), 498–513.

Thüringer Ministerium für Bildung, W. u. K. (2010). *Lehrplan für die Grundschule und die Förderschule mit dem Bildungsgang Grundschule. Mathematik.* Zugriff am 01.09.2012. Verfügbar unter http://thillm.rz.tu-ilmenau.de/get.aspx?mid=1262&did=13959&file=lp_gs_Ma_2010.pdf&usage=2&sid=44ebf4d3-a4c5-4d4a-80e9-f2a22aafa04d.

Tilly, W. (2008). The Evolution of School Psychology to Science-Based Practice: Problem Solving and the Three-Tiered Model. In A. Thomas & J. Grimes (Hrsg.), *Best practices in school psychology, V* (5. Aufl., S. 17–36). Bethesda, MD: National Association of School Psychologists.

Tindal, G., Germann, G. & Deno, S.L. (1983). Descriptive research on the Pine County norms: A compilation of findings (Research Report No. 132). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Marston, D. & Deno, S. L. (1983). *The reliability of direct and repeated measurement (Research Report No. 109)*. Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Trafton, P. (1992). Using number sense to develop mental computation and computational estimation. In C. Irons (Hrsg.), *Challenging Children to Think when they Compute* (S. 78-92). Brisbane: Centre for Mathematics and Science Education, Queensland University of Technology.

Trafton, P.R. & Thiessen, D. (1999). *Learning through Problems: Number Sense and Computational Strategies. A Resource for Primary Teachers*. Portsmouth, NH: Heinemann.

Tucker, J.A. (1985). Curriculum-Based Assessment: An Introduction. *Exceptional Children, 52* (3), 199–204.

Tupaika, J. (2003). *Schulversagen als komplexes Phänomen: Ein Beitrag zur Theorieentwicklung*. Bad Heilbrunn/Obb.: Klinkhardt.

van Eimeren, L. & Ansari, D. (2009). Rechenschwäche – eine neurokognitive Perspektive. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Handbuch Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie* (Studium Paedagogik, 2. Aufl., S. 25–33). Weinheim: Beltz.

van Heuvel-Panhuizen, M. (1995). Leistungsmessung im aktiv-entdeckenden Mathematikunterricht. In H. Brügelmann, H. Balhorn & L. Füssenich (Hrsg.), *Am Rande der Schrift. Zwischen Sprachenvielfalt und Analphabetismus* (Libelle Lesen und Schreiben, Bd. 6, S. 87–107). Lengwil am Bodensee: Libelle.

van Loosbroek, E. & Smitsman, A. (1990). Visual perception of numerosity in infancy. *Developmental Psychology, 26,* 916–922.

van Luit, J.E.H., van de Rijt, B.A.M. & Hasemann, K. (1994). *Osnabrücker Test zur Zahlbegriffsentwicklung (OTZ)*. Weinheim: Beltz.

VanDerHeyden, A. M. & Burns, M. K. (2005). Using Curriculum-Based Assessment and Curriculum-Based Measurement to Guide Elementary Mathematics Instruction: Effect on Individual and Group Accountability Scores. *Assessment for Effective Intervention, 30* (3), 15–31.

VanDerHeyden, A.M. & Burns, M.K. (2008). Examination of the Utility of Various Measures of Mathematics Proficiency. *Assessment for Effective Intervention, 33* (4), 215–224.

VanDerHeyden, A.M., Broussard, C., Fabre, M., Stanley, J., Legendre, J. & Creppell, R. (2004). Development and validation of curriculum-based measures of math performance for preschool children. *Journal of Early Intervention, 27*, 27-41.

VanDerHeyden, A.M., Witt, J.C. & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45* (2), 225–256.

VanDerHeyden, A.M., Witt, J.C., Naquin, G. & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30* (3), 363–382.

Vaughn, S. & Fuchs, L.S. (2003). Redefining Learning Disabilities as Inadequate Response to Instruction: The Promise and Potential Problems. *Learning Disabil Res Pract, 18* (3), 137–146.

Vaughn, S. & Roberts, G. (2007). Secondary Interventions in Reading Providing Additional Instruction for Students At Risk. *Exceptional Children, 39* (5), 40–46.

Vaughn, S., Linan-Thompson, S. & Hickman, P. (2003). Response to Instruction as a Means of Identifying Students with Reading/Learning Disabilities. *Exceptional Children, 69* (4), 391–409.

Vellutino, F.R., Scalon, D. & Lyon, G.R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition for reading disability. *Journal of Learning Disabilities, 33,* 223–238.

Vellutino, F.R., Scanlon, D.M., Small, S. & Fanuele, D.P. (2006). Response to Intervention as a Vehicle for Distinguishing Between Children With and Without Reading Disabilities: Evidence for the Role of Kindergarten and First-Grade Interventions. *Journal of Learning Disabilities, 39* (2), 157–169.

von Aster, M.G. (2005). Wie kommen Zahlen in den Kopf? Ein Modell der normalen und abweichenden Entwicklung

zahlenverarbeitender Hirnfunktionen. In M.G. von Aster & J.H. Lorenz (Hrsg.), *Rechenstörungen bei Kindern. Neurowissenschaft, Psychologie, Pädagogik ; mit 9 Tabellen* (S. 13–33). Göttingen: Vandenhoeck & Ruprecht.

von Aster, M.G., Deloche, G., Dellatolas, G. & Meier, M. (1997). Zahlenverarbeitung und Rechnen bei Schulkindern der 2. und 3. Klassenstufe: Eine vergleichende Studie französischsprachiger und deutschsprachiger Kinder. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 29,* 151–166.

von Aster, M.G., Kucian, K. & Martin, E. (2006). Gehirnentwicklung und Dyskalkulie. *Sprache Stimme Gehör, 30,* 1–6.

von Aster, M.G., Schweiter, M. & Weinhold Zulauf, M. (2007). Rechenstörungen bei Kindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 39* (2), 85–96.

Voß, S., Blumenthal, Y., Diehl, K., Ehlers, K., Mahlau, K. & Hartke, B. (2012). Erste Evaluationsergebnisse des Projekts „Rügener Inklusionsmodell (RIM) – Präventive und Integrative Schule auf Rügen (PISaR)“ – Ein Zwischenbericht. Rostock: Universität Rostock.

Wagner, E. & Hartke, B. (2006). *Rechenfische*. Versuchsversion. Rostock: Universität.

Wagner, H.-J. & Born, C. (1994). *Diagnostikum: Basisfähigkeiten im Zahlenraum 0 bis 20. DBZ 1*. Göttingen: Hogrefe Verlag.

Walter, J. (2008a). Adaptiver Unterricht erneut betrachtet: Über die Notwendigkeit systematischer formativer Evaluation von Lehr- und Lernprozessen und die daraus resultierende Diagnostik und Neudefinition von Lernstörungen nach dem RTI-Paradigma. *Zeitschrift für Heilpädagogik, 59,* 202–215.

Walter, J. (2008b). Curriculumbasiertes Messen (CBM) als lernprozessbegleitende Diagnostik: Erste deutschsprachige Ergebnisse zur Validität, Reliabilität und Veränderungssensibilität eines robusten Indikators zur Lernfortschrittsmessung beim Lesen. *Heilpädagogische Forschung, 34,* 62–79.

Walter, J. (2009a). Theorie und Praxis Curriculumbasierten Messens (CBM) in Unterricht und Foerderung. *Zeitschrift für Heilpädagogik, 60* (5), 162–170.

Walter, J. (2009b). Eignet sich die Messtechnik "MAZE" zur Erfassung von Lesekompetenzen als lernprozessbegleitende Diagnostik? *Heilpädagogische Forschung, 35,* 62–75.

Walter, J. (2010a). *Lernfortschrittsdiagnostik Lesen. LDL - ein curriculumbasiertes Verfahren* (Deutsche Schultests). Göttingen: Hogrefe Verlag.

Walter, J. (2010b). Lernfortschrittsdiagnostik am Beispiel der Lesekompetenz (LDL). Messtechnische Grundlagen sowie Befunde über zu erwartende Zuwachsraten während der Grundschulzeit. *Heilpädagogische Forschung, 36* (4), 162–176.

Walter, J. (2011a). Die Messung der Entwicklung der Lesekompetenz im Dienste der systematischen formativen Evaluation von Lehr- und Lernprozessen. *Zeitschrift für Heilpädagogik, 62,* 204–217.

Walter, J. (2011b). Die Entwicklung eines auch computerbasiert einsetzbaren Instruments zur formativen Messung von Lesekompetenz. *Heilpädagogische Forschung, 37,* 106–126.

Warrington, E.K. (1982). The fractionation of arithmetical skills: A single case study. *The Quarterly Journal of Experimental Psychology Section A, 34* (1), 31–51.

Wayman, M., Wallace, T., Wiley, H.I., Ticha, R. & Espin, C.A. (2007). Literature Synthesis on Curriculum-Based Measurement in Reading. *The Journal of Special Education, 41* (2), 85–120.

Wehrmann, M. (2003). *Qualitative Diagnostik von Rechenschwierigkeiten im Grundlagenbereich Arithmetik*. Berlin.

Wehrmann, M. (2007): Qualitative Diagnostik der Rechenschwäche. Bedeutung der differenzierten Diagnostik im Lernprozess. In G. Schulte-Körne (Hrsg.), *Legasthenie und Dyskalkulie: aktuelle Entwicklungen in Wissenschaft, Schule und Gesellschaft* (333-337). Bochum: Winkler.

Weißhaupt, S. & Peucker, S. (2009). Entwicklung arithmetischen Vorwissens. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Handbuch Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie* (Studium Paedagogik, 2. Aufl., S. 52–76). Weinheim: Beltz.

Weißhaupt, S., Peucker, S. & Wirtz, M. (2006). Diagnose mathematischen Vorwissens im Vorschulalter und Vorhersage von Rechenleistungen und Rechenschwierigkeiten in der Grundschule. *Psychologie in Erziehung und Unterreicht, 53,* 236–245.

Werner, B. (2007). Diagnose und Förderung im Bereich der Zahlbegriffsentwicklung. In J. Walter & F.B. Wember (Hrsg.), *Sonderpädagogik des Lernens* (Handbuch Sonderpädagogik, Bd. 2, 1. Aufl., 571–590). Göttingen: Hogrefe Verlag.

Wesson, C.L. (1991). Curriculum-based Measurement and 2 Models of follow-up Consultation. *Exceptional Children, 57* (3), 246–256.

Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik, 3* (3), 225–242.

Wiley, I.H. & Deno, S.L. (2005). Oral Reading and Maze Measures as Predictors of Success for English Learners on a State Standards Assessment. *Remedial and Special Education, 26* (4), 207–214.

Wilson, K. & Swanson, H. (2001). Are Mathematics Disabilities Due to a Domain-General or a Domain-Specific Working Memory Deficit? *Journal of Learning Disabilities, 34* (3), 237–248.

Winer, B.J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

Wininger, S. (2005). Using Your Tests to Teach: Formative Summative Assessment. *Teaching of Psych., 32* (3), 164–166.

Winter, F. (2006). *Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen*. Baltmannsweiler: Schneider Hohengehren.

Wittmann, E.C. & Müller, G.N. (2004). *Das Zahlenbuch 1.*. Leipzig: Ernst Klett.

Woodcock, R.W. & Johnson, M.B. (1989). *WJ-R Tests of Cognitive Ability*. Itasca, IL: Riverside Publishing.

Wright, J. (2001). *Curriculum-Based Measurement: A Manual for Teachers.* Zugriff am 14.05.2012. Verfügbar unter http://www.jimwrightonline.com/pdfdocs/cbaManual.pdf.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358* (6389), 749–750.

Xu, F. & Arriaga, R. I. (2007). Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology, 25* (1), 103–108.

Xu, F. & Spelke, E. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74,* B1-B11.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition, 89,* B15-B25.

Yeh, S.S. (2006). Can rapid assessment reduce the pressure? *Teachers college Record, 108*(4), 621-661.

Yeo, S. (2010). Predicting Performance on State Achievement Tests Using Curriculum-Based Measurement in Reading: A Multilevel Meta-Analysis. *Remedial and Special Education, 31* (6), 412–422.

Abkürzungsverzeichnis

Abb.	Abbildung
bzw.	beziehungsweise
CBA	curriculum-based assessment
CBM	curriculum-based measurement bzw. curriculumbasiertes Messverfahren
d. h.	das heißt
DIBS	dynamic idicators of basic skills
et al.	et alii bzw. et aliae
evtl.	eventuell
f.	folgende (eine Seite)
ff.	folgende (mehr als eine Seite)
ggf.	gegebenenfalls
GOM	general outcome measure
Herv. d. Verf.	Hervorhebung des Verfassers
HLM	Hierarchisch-lineare Modellierung
in Vorb.	In Vorbereitung
ISER	Institut für Sonderpädagogische Entwicklungsförderung und Rehabilitation
Max.	Maximum
MBSP	Monitoring Basic Skills Progress
Min.	Minimum
MM	mastery measure
MW	Mittelwert
MZP	Messzeitpunkt
N	Stichprobenumfang
p	Signifikanzniveau
PR	Prozentrang
o. S.	ohne Seitenangabe
r	Reliabilitätskoeffizient (auf Basis einer Korrelation)
RW	Rohwert
S.	Seite
SD	Standardabweichung

SBM	skills-based measure
sog.	sogenannter, sogenannte bzw. sogenanntes
u. a.	unter anderem
u. ä.	und ähnliches
UG	Untersuchungsgruppe
usw.	und so weiter
UT	Untertest
vgl.	vergleiche
z. B.	zum Beispiel

Tabellenverzeichnis

Tabelle 1: Zusammenfassung der prozentualen Anteile der Schüler unterhalb Kompetenzstufe I und auf Kompetenzstufe I in dem Bereich mathematischer Kompetenzen aus PISA 2000, 2003, 2006 und 2009 (Klieme, Neubrand & Lüdtke, 2001; Prenzel et al., 2004; Frey et al., 2007; OECD, 2010)..... 10

Tabelle 2: Klassifikationssystem zur Einschätzung von Evidenzen gemäß dem Deutschen Crochane Zentrum (2011, o. S.)..... 128

Tabelle 3: Vorgeschlagene Kategorisierung von Evidenzkriterien für (sonder-)pädagogische Förderprogramme nach Hartke (2005) sowie Fingerle und Ellinger (2008) (entnommen aus: Hartke, 2005, S. 17; Fingerle & Ellinger, 2008, S 9f.)..... 131

Tabelle 4: Überblick über die bekanntesten in den USA verfügbare Early Numeracy CBM (Hosp et al., 2007) 159

Tabelle 5: Übersicht über Studien zur Paralleltest-Reliabilität von CBM-Verfahren im Bereich Mathematik (entnommen aus: Christ et al., 2008, S. 202)..... 181

Tabelle 6: Übersicht über Studien zur internen Konsistenz von CBM-Verfahren im Bereich Mathematik (entnommen aus: Christ et al., 2008, S. 202)..... 182

Tabelle 7: Übersicht über die curriculumbasierten Messungen im ersten Schuljahr (Niveaustufe 1 sowie 2)..... 219

Tabelle 8: Einordnung der CBM in die Curricula und Rahmenlehrpläne der deutschen Bundesländer 222

Tabelle 9: Im CBM „Addition im Zwanzigerraum“ enthaltene Aufgabenstrukturen..... 239

Tabelle 10: Im CBM „Subtraktion im Zwanzigerraum“ enthaltene Aufgabenstrukturen 244

Tabelle 11: Übersicht über die Untersuchungsgruppen sowie deren Besonderheiten und Zweck 255

Tabelle 12: Eingesetzte Versionen der CBM „Arithmetische Basiskompetenzen“ in den Regionen Rügen (UG 1) und Nordvorpommern (UG 2) innerhalb des ersten Schuljahres (Niveaustufe 1 und 2) 263

Tabelle 13: Eingesetzte Versionen der CBM „Zahlenlesen im Zwanzigerraum“, „Zahlenreihe im Zwanzigerraum“ und „Mengenvergleich im Zwanzigerraum“ in den Regionen Rügen und Nordvorpommern in der ersten Hälfte des ersten Schuljahres (Niveaustufe 1) 263

Tabelle 14: Eingesetzte Versionen der CBM „Zahlzerlegung im Zwanzigerraum“, „Addition im Zwanzigerraum“ und „Subtraktion im Zwanzigerraum“ in den Regionen Rügen und Nordvorpommern in der zweiten Hälfte des ersten Schuljahres (Niveaustufe 2) 264

Tabelle 15: Leistungseinschätzung auf Grundlage erreichter Prozentrangwerte 267

Tabelle 16: Angaben zum Einsatz der verschiedenen CBM-Versionen innerhalb der Substudie 3 in der Region Rostock 269

Tabelle 17: Übersicht über in der vorliegenden Arbeit durchgeführte statistische Analysen zur Einschätzung der Güte der erarbeiteten CBM 283

Tabelle 18: Interpretation des Korrelationskoeffizienten r nach Bühl (2006, S. 342) 287

Tabelle 19: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Zahlenlesen im Zwanzigerraum“ über die vier

Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 311

Tabelle 20: Deskriptive Statistiken des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen 314

Tabelle 21: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Zahlenreihe im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 317

Tabelle 22: Deskriptive Statistiken des CBM „Zahlenreihe im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen 319

Tabelle 23: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 322

Tabelle 24: Deskriptive Statistiken des CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen 324

Tabelle 25: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr in der Region Nordvorpommern 328

Tabelle 26: Deskriptive Statistiken des CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr in der Region Rügen..... 332

Tabelle 27: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf

Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 335

Tabelle 28: Deskriptive Statistiken des CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen 337

Tabelle 29: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 341

Tabelle 30: Deskriptive Statistiken des CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen 343

Tabelle 31: Deskriptive Statistiken der Versionen 1 und 2 des CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 346

Tabelle 32: Deskriptive Statistiken des CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen 348

Tabelle 33: Ergebnisse der Analysen hinsichtlich der Splithalf-Reliabilität des CBM „Zahlenlesen im Zwanzigerraum“ in der Region Rostock..... 353

Tabelle 34: Ergebnisse der Analysen hinsichtlich der Splithalf-Reliabilität des CBM „Zahlenreihe im Zwanzigerraum“ in der Region Rostock..... 354

Tabelle 35: Ergebnisse der Analysen hinsichtlich der Splithalf-Reliabilität des CBM „Mengenvergleich im Zwanzigerraum“ in der Region Rostock 355

Tabelle 36: Ergebnisse der Analysen hinsichtlich der Splithalf-Reliabilität des CBM „Arithmetische Basiskompetenzen“ in der Region Rostock..... 356

Tabelle 37: Ergebnisse der Analysen hinsichtlich der Retest-Reliabilität des CBM „Arithmetische Basiskompetenzen“ in der Region Nordvorpommern 359

Tabelle 38: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Zahlenlesen im Zwanzigerraum“ zu den vier Messzeitpunkten in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern 362

Tabelle 39: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Zahlenreihe im Zwanzigerraum“ zu den vier Messzeitpunkten in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern 363

Tabelle 40: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Mengenvergleich im Zwanzigerraum“ zu den vier Messzeitpunkten in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 364

Tabelle 41: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Arithmetische Basiskompetenzen“ zu den neun Messzeitpunkten im ersten Schuljahr in der Region Nordvorpommern .. 366

Tabelle 42: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Zahlzerlegung im Zwanzigerraum“ zu den fünf Messzeitpunkten in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern 367

Tabelle 43: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Addition im Zwanzigerraum“ zu den fünf Messzeitpunkten in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern 368

Tabelle 44: Korrelationskoeffizienten und Ergebnisse der t-Tests der Version 1 und Version 2 des CBM „Subtraktion im Zwanzigerraum“ zu den fünf Messzeitpunkten in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern 369

Tabelle 45: Interkorrelation (nach Pearson) der CBM der Niveaustufe 1 zum MZP1 in der fünften Schulwoche 372

Tabelle 46: Interkorrelation (nach Pearson) der CBM der Niveaustufe 1 zum MZP 2 in der achten Schulwoche 373

Tabelle 47: Interkorrelation (nach Pearson) der CBM der Niveaustufe 1 zum MZP 3 in der zwölften Schulwoche 373

Tabelle 48: Interkorrelation (nach Pearson) der CBM der Niveaustufe 1 zum MZP 4 in der 16. Schulwoche .. 374

Tabelle 49: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP 5 in der 21. Schulwoche .. 374

Tabelle 50: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP 6 in der 26. Schulwoche .. 375

Tabelle 51: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP7 in der 30. Schulwoche ... 376

Tabelle 52: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP 8 in der 34. Schulwoche .. 376

Tabelle 53: Interkorrelation (nach Pearson) der CBM der Niveaustufe 2 zum MZP 9 in der 38. Schulwoche .. 377

Tabelle 54: Multitrait-Multimethod-Matrix des CBM „Zahlenlesen im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), des Diagnoseverfahrens Kalkulie Teil 1 (Fritz et al., 2007) sowie des Münsteraner Screenings (MÜSC; Mannhaupt, 2005) zum ersten Messzeitpunkt in der 5. Schulwoche (N = 436)..... 379

Tabelle 55: Multitrait-Multimethod-Matrix des CBM „Zahlenreihe im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), des Diagnoseverfahrens Kalkulie Teil 1 (Fritz et al., 2007) sowie des Münsteraner Screenings (MÜSC; Mannhaupt, 2005) zum ersten Messzeitpunkt in der 5. Schulwoche (N = 435)..... 379

Tabelle 56: Multitrait-Multimethod-Matrix des CBM „Mengenvergleich im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), des Diagnoseverfahrens Kalkulie Teil 1 (Fritz et al., 2007) sowie des Münsteraner Screenings (MÜSC; Mannhaupt, 2005) zum ersten Messzeitpunkt in der 5. Schulwoche (N = 435)..... 380

Tabelle 57: Multitrait-Multimethod-Matrix des CBM „Arithmetische Basiskompetenzen“, des CBM „Silbenlesen“ der Niveaustufe 1 (Diehl, 2010a), des Diagnoseverfahrens Kalkulie Teil 1 (Fritz et al., 2007) sowie des Münsteraner Screenings (MÜSC; Mannhaupt, 2005) zum ersten Messzeitpunkt in der 5. Schulwoche (N = 435)..... 381

Tabelle 58: Multitrait-Multimethod-Matrix des CBM „Arithmetische Basiskompetenzen“, des CBM „Silbenlesen“ der

Niveaustufe 2 (Diehl, 2010b), des Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) sowie der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) zum neunten Messzeitpunkt in der 38. Schulwoche (N = 364)..... 382

Tabelle 59: Multitrait-Multimethod-Matrix des CBM „Zahlzerlegung im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b), des Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) sowie der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) zum neunten Messzeitpunkt in der 38. Schulwoche (N = 365)..... 383

Tabelle 60: Multitrait-Multimethod-Matrix des CBM „Addition im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b), des Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) sowie der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) zum neunten Messzeitpunkt in der 38. Schulwoche (N = 366)..... 384

Tabelle 61: Multitrait-Multimethod-Matrix des CBM „Subtraktion im Zwanzigerraum“, des CBM „Silbenlesen“ der Niveaustufe 2 (Diehl, 2010b), des Deutschen Mathematiktest für erste Klassen (Krajewski et al., 2002) sowie der Würzburger Leise Leseprobe (Küspert & Schneider, 1998) zum neunten Messzeitpunkt in der 38. Schulwoche (N = 362)..... 385

Tabelle 62: Korrelationsmatrix nach Pearson der CBM der Niveaustufe 1 und 2 mit dem Inventar Rechenfische (Wagner & Hartke, 2006) zum vierten bzw. fünften Messzeitpunkt in der 21. Schulwoche 386

Tabelle 63: Korrelationsmatrix nach Pearson der CBM der Niveaustufe 2 mit dem Inventar Rechenfische (Wagner

& Hartke, 2006) zum achten Messzeitpunkt in der 34. Schulwoche 387

Tabelle 64: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Zahlenlesen im Zwanzigerraum“ (PR ≤ 25)..... 392

Tabelle 65: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Zahlenreihe im Zwanzigerraum“ (PR ≤ 25) 394

Tabelle 66: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Mengenvergleich im Zwanzigerraum“ (PR ≤ 25) ... 396

Tabelle 67: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Arithmetische Basiskompetenzen““ (PR ≤ 25)..... 398

Tabelle 68: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Zahlzerlegung im Zwanzigerraum““ (PR ≤ 25) 400

Tabelle 69: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Addition im Zwanzigerraum““ (PR ≤ 25) 402

Tabelle 70: Ergebnisse zur Vorhersagegüte von Minderleistungen im Rechnen zum Ende der Klasse eins (PR ≤ 15 im DEMAT 1+; Krajewski et al., 2002) durch das CBM „Subtraktion im Zwanzigerraum““ (PR ≤ 25)..... 405

Tabelle 71: Ergebnisse der logistischen Regressionsanalyse der CBM der Niveaustufe 1 des ersten Messzeitpunkts in der fünften Schulwoche zur Vorhersage der Zugehörigkeit zur Gruppe der Risikokinder im Bereich Mathematik, operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 1+ (Krajewski et al., 2002) zum Ende der Klassenstufe 1 (N = 387)..... 406

Tabelle 72: Ergebnisse der logistischen Regressionsanalyse der CBM der Niveaustufe 2 des fünften Messzeitpunkts in der 21. Schulwoche zur Vorhersage der Zugehörigkeit zur Gruppe der Risikokinder im Bereich Mathematik, operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 1+ (Krajewski et al., 2002) zum Ende der Klassenstufe 1 (N = 362) 408

Tabelle 73: Ergebnisse der logistischen Regressionsanalyse der CBM der Niveaustufe 1 des ersten Messzeitpunkts in der fünften Schulwoche zur Vorhersage der Zugehörigkeit zur Gruppe der Risikokinder im Bereich Mathematik, operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 2+ zum Ende der Klassenstufe 2 (N = 388)..... 410

Tabelle 74: Ergebnisse der logistischen Regressionsanalyse der CBM der Niveaustufe 2 des fünften Messzeitpunkts in der 21. Schulwoche zur Vorhersage der Zugehörigkeit zur Gruppe der Risikokinder im Bereich Mathematik, operationalisiert durch einen Prozentrang kleiner bzw. gleich 15 im DEMAT 1+ (Krajewski et al., 2002) zum Ende der Klassenstufe 2 (N = 377) 411

Tabelle 75: Interkorrelationen der CBM „Zahlenlesen im Zwanzigerraum“ über die MZP 1 – 4 (listenweiser Fallausschluss fehlender Werte, N = 402)..... 414

Tabelle 76: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM “Zahlenlesen im Zwanzigerraum” 416

Tabelle 77: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM “Zahlenlesen im Zwanzigerraum” 419

Tabelle 78: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Zahlenlesen im Zwanzigerraum“ 419

Tabelle 79: Interkorrelationen der CBM „Zahlenreihe im Zwanzigerraum“ über die MZP 1 – 4 (listenweiser Fallausschluss fehlender Werte, N = 395)..... 420

Tabelle 80: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM “Zahlenreihe im Zwanzigerraum” 422

Tabelle 81: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM “Zahlenreihe im Zwanzigerraum” 424

Tabelle 82: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Zahlenreihe im Zwanzigerraum“ 424

Tabelle 83: Interkorrelationen der CBM „Mengenvergleich im Zwanzigerraum“ über die MZP 1 – 4 (listenweiser Fallausschluss fehlender Werte, N = 390)..... 425

Tabelle 84: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM “Mengenvergleich im Zwanzigerraum” 427

Tabelle 85: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM “Mengenvergleich im Zwanzigerraum” 429

Tabelle 86: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Mengenvergleich im Zwanzigerraum“ 429

Tabelle 87: Interkorrelationen der CBM „Arithmetische Basiskompetenzen“ über die MZP 1 – 9 (listenweiser Fallausschluss fehlender Werte, N = 307)..... 430

Tabelle 88: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM “Arithmetische Basiskompetenzen” 432

Tabelle 89: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM “Arithmetische Basiskompetenzen” 434

Tabelle 90: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Arithmetische Basiskompetenzen“ 434

Tabelle 91: Interkorrelationen der CBM „Zahlzerlegung im Zwanzigerraum“ über die MZP 5 – 9 (listenweiser Fallausschluss fehlender Werte, N = 335)..... 435

Tabelle 92: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM „Zahlzerlegung im Zwanzigerraum” 437

Tabelle 93: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM „Zahlzerlegung im Zwanzigerraum“ 439

Tabelle 94: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Zahlzerlegung im Zwanzigerraum“ 439

Tabelle 95: Interkorrelationen der CBM „Addition im Zwanzigerraum“ über die MZP 5 – 9 (listenweiser Fallausschluss fehlender Werte, N = 331)..... 440

Tabelle 96: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM “Addition im Zwanzigerraum” 442

Tabelle 97: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM „Addition im Zwanzigerraum“ 443

Tabelle 98: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Addition im Zwanzigerraum“ 444

Tabelle 99: Interkorrelationen der CBM „Subtraktion im Zwanzigerraum“ über die MZP 5 – 9 (listenweiser Fallausschluss fehlender Werte, N = 333)..... 445

Tabelle 100: Ergebnisse des unkonditionierten Hierarchisch-linearen Modells für das CBM „Subtraktion im Zwanzigerraum“ 447

Tabelle 101: Ergebnisse des allgemeinen Hierarchisch-linearen Modells für das CBM „Subtraktion im Zwanzigerraum“ 449

Tabelle 102: Übersicht über die erklärte Varianz im Ausgangsniveau und im Lernzuwachs des CBM „Subtraktion im Zwanzigerraum“ 449

Tabelle 103: Einordnung ermittelter durchschnittlicher Anstiege je Woche in bereits veröffentlichte Referenzwerte (AIMSweb.com, 2012; Clarke & Shinn, 2004; easyCBM.com, 2012b; L.S. Fuchs & D. Fuchs, Hamlett, Walz & Germann, 1993; Lembke et al., 2008; Lembke & Foegen, 2009; Shapiro et al., 2005)..... 507

Tabelle 104: Überblick über zentrale Befunde der innerhalb der Studie erarbeiteten CBM-Verfahren 520

Abbildungsverzeichnis

Abbildung 1: Kernkomponenten des Response-to-Intervention-Ansatzes (adaptiert und modifiziert nach National Center on Response to Intervention, 2010)..... 23

Abbildung 2: Versuchsaufbau zur Darstellung der Additionsaufgabe 1 + 1 und der Subtraktionsaufgabe 2 – 1 für Säuglinge nach Wynn (1992) (entnommen aus: Wynn, 1992, S. 749) 32

Abbildung 3: „Entwicklungsmodell früher mathematischer Kompetenzen“ nach Krajewski und Schneider (2006) (entnommen aus: Krajewski & Schneider, 2006, S. 250) 44

Abbildung 4: Modell der mathematischen Kompetenzentwicklung nach Fritz et al. (2007) (entnommen aus: Fritz et al., 2007, S. 15) 49

Abbildung 5: Triple-code-model nach Dehaene (übersetzt und vereinfacht nach Dehaene, 1992, S. 31)..... 62

Abbildung 6: Verarbeitung einer Information gemäß dem Drei-Speicher-Modell (Atkinson & Shiffrin, 1968)..... 68

Abbildung 7: Arbeitsgedächtnismodell nach Baddeley (2000) (entnommen aus: Hasselhorn & Gold, 2009, S. 74) . 69

Abbildung 8: Allgemeine und inhaltsbezogene mathematischen Kompetenzen im Mathematikunterricht der Grundschule (Kultusministerkonferenz, 2004, S. 7) 73

Abbildung 9: Zweistufiges Diagnoseverfahren nach Moser Opitz (entnommen aus Moser Opitz, 2009, S. 294)..... 100

Abbildung 10:Risiko Indices für Lernbehinderung (learning disability) in den USA für die Jahre 1974 bis 1998 (Daten entnommen aus: Donovan & Cross, 2002, S. 49) .. 116

Abbildung 11: Flussdiagramm zur schematischen Darstellung des Einsatzes diagnostischer Maßnahmen im RtI in Anlehnung an L.S. Fuchs und D. Fuchs (2007) sowie NCRTI (2010) (entnommen und modifiziert aus: L.S. Fuchs & D. Fuchs, 2007, S. 26) 126

Abbildung 12: Hypothetisches Beispiel für ein mittels mastery measures ermitteltes Lernverlaufsprofil in Anlehnung an Hosp et al. (2007) (entnommen aus: Hosp et al., 2007, S. 13) 147

Abbildung 13: Schematische Darstellung des Zusammenhangs von curriculum-based asessment (CBA), curriculum-based measurements (CBM), skills-based measures (SBM), general outcome measures (GOM), dynamic indicators of basic skills (DIBS) und mastery measures (MM) in Anlehnung an Tucker (1985), Deno (1985), Hosp, Hosp und Howell (2007), L.S. Fuchs und Deno (1994), Shinn (1995) sowie VanDerHeyden & Burns (2005) 149

Abbildung 14: Auswertung von Schülerlösungen anhand der Anzahl korrekter Ziffern (entnommen aus: L.S. Fuchs & D. Fuchs, 2005, S. 14)..... 151

Abbildung 15: Beispiel für die grafische Darstellung von CBM-Ergebnissen (schwarze Linie) im Vergleich zu einer festgelegten Ziellinie (gestrichelte Linie) und zu einem zuvor definierten Ziel (schwarzes Kreuz) (entnommen und modifiziert aus: L.S. Fuchs & D. Fuchs, 2005, S. 66) 154

Abbildung 16: Hypothetisches Beispiel für den Lernverlauf eines Schülers (schwarze Linie) erhoben mittels CBM sowie der dazugehörigen Trendlinie (graue Linie) im Vergleich zu einer festgelegten Ziellinie (gestrichelte Linie) und zu einem zuvor definierten Ziel (schwarzes Kreuz)

(entnommen und modifiziert aus: L.S. Fuchs & D. Fuchs, 2005, S. 74)..... 156

Abbildung 17: Hypothetisches Beispiel für den Lernverlauf eines Schülers (schwarze Linie) erhoben mittels CBM sowie der dazugehörigen Trendlinie (graue Linie) im Vergleich zu einer festgelegten Ziellinie (gestrichelte Linie) und zu einem zuvor definierten Ziel (schwarzes Kreuz) (entnommen und modifiziert aus: L.S. Fuchs & D. Fuchs, 2005, S. 75)..... 157

Abbildung 18: Beispielseite 1 eines CBM des Typs concepts & application nach L.S. Fuchs und D. Fuchs (2007) (entnommen aus: L.S. Fuchs & D. Fuchs, 2007, S. 25) 162

Abbildung 19: Beispielseite 2 eines CBM des Typs concepts & application nach L.S. Fuchs und D. Fuchs (2007) (entnommen aus: L.S. Fuchs & D. Fuchs, 2007, S. 25) 163

Abbildung 20: Beispielseite 3 eines CBM des Typs concepts & application nach L.S. Fuchs und D. Fuchs (2007) (entnommen aus: L.S. Fuchs & D. Fuchs, 2007, S. 25) 164

Abbildung 21: Ausschnitt der Internetseite des NCRTI zum Überblick über CBM-Verfahren (Stand 12.01.2013) für den Bereich der Grundschulmathematik, welche durch das Review-Komitee des National Center on Response to Intervention begutachtet wurden (National Center on Response to Intervention, 2012) 187

Abbildung 22: Ausschnitt einer möglichen Ordnung von Teilfertigkeiten innerhalb einer Lernhierarchie am Beispiel der Subtraktion im Zahlenraum (ZR) bis 20 (in Anlehnung an Klauer, 1978)..... 205

Abbildung 23: Beispiel zur Veranschaulichung der Itemanordnung in Parallelformen von Mathematik-CBM: zufällige Itemanordnung (links) vs. stratifizierte (nach inhaltlichen Aspekten geordnete) Items (rechts) (entnommen und modifiziert nach Christ & Vining, 2006, S. 390) 207

Abbildung 24: Beispielitems des CBM „Zahlenlesen im Zwanzigerraum“ 225

Abbildung 25: Beispielitems des CBM „Zahlenreihe im Zwanzigerraum“ 226

Abbildung 26: Beispielitems des CBM „Mengenvergleich im Zwanzigerraum“ 228

Abbildung 27: In den CBM „Arithmetische Basiskompetenzen“ abgefragte Kompetenzen verdeutlicht anhand von Beispielitems 233

Abbildung 28: Beispielitems des CBM „Zahlzerlegung im Zwanzigerraum“ 234

Abbildung 29: Prozentuale Anteile richtiger Lösungen von Additionsaufgaben im Zahlenraum bis 20, bei einer Anzahl an Probanden zwischen 146 und 284 Erstklässlern (entnommen aus: Rinkens & Eilerts, 2004, S. 7) 236

Abbildung 30: Beispielitems des CBM „Addition im Zwanzigerraum“ 239

Abbildung 31: Prozentuale Anteile richtiger Lösungen von Subtraktionsaufgaben im Zahlenraum bis 20, bei einer Anzahl an Probanden zwischen 141 und 287 Erstklässlern (entnommen aus: Rinkens & Eilerts, 2004, S. 13) 241

Abbildung 32: Beispielitems des CBM „Subtraktion im Zwanzigerraum“ 244

Abbildung 33: Überblick über die zu den verschiedenen Messzeitpunkten eingesetzten Testverfahren 261

Abbildung 34: Beispiel einer mithilfe der vorprogrammierten Microsoft-Excel-Datei automatisch generierten grafischen Ausgabe..... 266

Abbildung 35: Teilaspekte der Validität eines Testverfahrens (angelehnt an: Bühner, 2006; Moosbrugger & Kelava, 2012)..... 287

Abbildung 36: schematische Darstellung der Multitrait-Multimethod-Matrix zweier Traits (mathematische und Lesekompetenz) erhoben mittels zweier Methods (CBM und weiteres standardisiertes Verfahren)..... 290

Abbildung 37: Hypothetische Klassifikation anhand eines Screenings (Prädiktor) sowie eines Kriteriums untersuchter Personen in Anlehnung an Marx und Lenhard (2010) 297

Abbildung 38: Datenstruktur einer wiederholten Messung (z. B. mittels CBM) im Rahmen des hierarchisch-linearen Mehrebenenmodells (entnommen und modifiziert aus Langer, 2009, S. 223) 301

Abbildung 39: Datenstruktur einer wiederholten Messung mittels CBM im Rahmen des unkonditionierten hierarchisch-linearen Mehrebenenmodells (entnommen und modifiziert aus Ditton, 1998, S. 67) 303

Abbildung 40: Datenstruktur einer wiederholten Messung (z. B. mittels CBM) im Rahmen des allgemeinen hierarchisch-linearen Mehrebenenmodells (entnommen und modifiziert aus Langer, 2009, S. 225)..... 306

Abbildung 41: Rohwertverteilungen der Versionen 1 und 2 des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern 312

Abbildung 42: Rohwertverteilungen des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen 315

Abbildung 43: Rohwertverteilungen der Versionen 1 und 2 des CBM „Zahlenlesen im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern 318

Abbildung 44: Rohwertverteilungen des CBM „Zahlenreihe im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen 320

Abbildung 45: Rohwertverteilungen der Versionen 1 und 2 des CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Nordvorpommern (Ausschnitt) 323

Abbildung 46: Rohwertverteilungen des CBM „Mengenvergleich im Zwanzigerraum“ über die vier Messzeitpunkte in der ersten Hälfte des ersten Schuljahres in der Region Rügen (Ausschnitt) 326

Abbildung 47: Rohwertverteilungen der Versionen 1 und 2 des CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr in der Region Nordvorpommern 330

Abbildung 48: Rohwertverteilungen des CBM „Arithmetische Basiskompetenzen“ über die neun Messzeitpunkte im ersten Schuljahr in der Region Rügen..... 333

Abbildung 49: Rohwertverteilungen der Versionen 1 und 2 des CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 336

Abbildung 50: Rohwertverteilungen des CBM „Zahlzerlegung im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen 339

Abbildung 51: Rohwertverteilungen der Versionen 1 und 2 des CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 342

Abbildung 52: Rohwertverteilungen des CBM „Addition im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen 344

Abbildung 53: Rohwertverteilungen der Versionen 1 und 2 des CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Nordvorpommern..... 347

Abbildung 54: Rohwertverteilungen des CBM „Subtraktion im Zwanzigerraum“ über die fünf Messzeitpunkte in der zweiten Hälfte des ersten Schuljahres in der Region Rügen 349

Abbildung 55: Ergebnisse der Beantwortung des Items 1) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 370

Abbildung 56: Ergebnisse der Beantwortung des Items 3) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 371

Abbildung 57: Ergebnisse der Beantwortung des Items 10) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 450

Abbildung 58: Ergebnisse der Beantwortung des Items 12) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 452

Abbildung 59: Ergebnisse der Beantwortung des Items 13) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 26) 453

Abbildung 60: Ergebnisse der Beantwortung des Items 4) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 454

Abbildung 61: Ergebnisse der Beantwortung des Items 5) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 455

Abbildung 62: Ergebnisse der Beantwortung des Items 6) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 456

Abbildung 63: Ergebnisse der Beantwortung des Items 7) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 457

Abbildung 64: Ergebnisse der Beantwortung des Items 8) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 458

Abbildung 65: Ergebnisse der Beantwortung des Items 9) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N =37) 459

Abbildung 66: Ergebnisse der Beantwortung des Items 14) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 460

Abbildung 67: Ergebnisse der Beantwortung des Items 2) des Aussagenkatalogs zur Einschätzung der CBM sowie deren Einsatz (N = 37) 461

Curriculumbasierte Messverfahren im mathematischen Erstunterricht

Die formative Leistungserfassung stellt ein wesentliches Element schulischer Prävention von Rechenschwierigkeiten dar. Als besonders innovativ und nachweislich wirksam zur Steigerung von schulischen Leistungen haben sich die in der US-amerikanischen Literatur beschriebenen curriculum-based measurements bzw. Curriculumbasierten Messverfahren (CBM) erwiesen und sind daher in den Fokus der Forschung in Deutschland gerückt. Die vorliegende Studie stellt den Konzeptions- und Evaluationsprozess von Verfahren zur Lernfortschrittsdokumentation nach diesem Ansatz differenziert, am Beispiel des Anfangsunterrichts Mathematik in der Grundschule dar.

Stefan Voß
Jahrgang 1985. 2009 Beendigung des Studiums der Sonderpädagogik. Seit 2010 wissenschaftlicher Mitarbeiter am Institut für Sonderpädagogische Entwicklungsförderung & Rehabilitation der Universität Rostock im Projekt „Rügener Inklusionsmodell“. Von 2010 bis 2013 Promotion zum Thema „Curriculumbasierte Messverfahren im mathematischen Erstunterricht“.

978-3-8381-3810-7