

**EIN ALLGEMEINES KOMPETENZORIENTIERTES
INFORMATIONEN- UND VORGEHENSMODELL
FÜR DIE STRUKTURIERUNG VON
DOMÄNENINHALTEN, ASSESSMENTS UND FEEDBACK**

Dissertation

zur
Erlangung des akademischen Grades
Doktor -Ingenieur (Dr.-Ing.)

Fakultät für Informatik und Elektrotechnik
der Universität Rostock

eingereicht von
John-Harry Wieken
Dipl.-Inform.
geb. am 22.02.1961 in Hannover

betreuender Hochschullehrer: Prof. Dr.-Ing. habil. Peter Forbrig



Rostock, 29.07.2016

Gutachter

1. Prof. Dr.-Ing. habil. Peter Forbrig,
Universität Rostock, Institut für Softwaretechnik
2. O. Univ.-Prof. Dr. Christian Stary, Johannes-Kepler-Universität Linz,
Institut für Wirtschaftsinformatik – Communications Engineering
3. Prof. Dr.-Ing. Alke Martens,
Universität Rostock, Institut für Praktische Informatik

Verteidigung: 12.12.2016, Universität Rostock

Abstract

With increasing requirements in dealing with complex situations and manage professional challenges, educational processes are in need to focus on preparing learners for such circumstances. Since knowledge is growing exponentially and relevant parts undergo frequent changes, increasing emphasis is put on the term “competence” up to the point that it has become the main goal in educational processes.

Therefore, the assessment of competencies is a prerequisite to evaluate the success of an educational process. It is essential to provide sound feedback to both the teachers as well as learners. Communicating success and failure in a substantiated way is one of the most critical factors for success of the underlying educational process. To achieve this, it is necessary to describe the relationship between educational objectives, competencies, domain contents and the assessment as well as results of the learners. This description has to take shape in a structured, traceable and explanatory model that can provide evidence-based feedback on an individual as well as on classroom level.

The development of the so called Core-Situation-Personal (CSP) information-model incorporates these factors. It is a meta-model using the Unified Modeling Language (UML) and is implemented in a database management system. Various integrated sub-models describe parts such as educational goals, competence hierarchies, domain contents, assessment structures and items, manifest and latent variables, psychometric models and item responses. A complementary process model is added and two graphical notations have been developed to visualize critical parts on instance level. The model has been tested by using an assessment in mathematics covering linear and exponential growth along with corresponding function types.

As a prerequisite for building this model, well suited definitions for terms and concepts like competence, performance and qualifications, which are under discussion, have been selected and adapted. While the focus is not on large-scale assessments like PISA or TIMSS but rather on classroom-level ideas, different approaches like Wilson’s “Four Building Blocks” and the “Evidence-Centered Design” as well as the DFG-Program “Kompetenzmodelle” have been adapted.

It turns out that only a domain based concept of competence will be useful while involvement in broader approaches like key competencies does not seem promising. However, this neither restricts the usability of the model in different domains nor the interchangeability of description between domains.

The specified requirements involve the flexibility to model different domains, traceability to provide explanations for results, the integration of psychometric models so results are not necessarily limited to single assessments, an understandable language-oriented content description, reusability and an economy of resources.

Based on these requirements the three-tier CSP Information Model is developed, where the three tiers are CORE (external and internal goals, competencies and competence hierarchies, competence levels, domain content, facts and objects, item prototypes), SITUATION (assessments, assessment structures, items, response types, manifest scoring model, psychometric model) and PERSONAL (actual responses, typed responses, personal data). Competencies may be modeled in different

hierarchies and combined in matrices with themselves as well as with competence levels in taxonomies.

A complementary process model with 15 steps is added to facilitate the application of the model. Each step defines fundamental goals in terms of information structures and gives hints to necessary actions. In particular, to facilitate the content modeling a graphical language-oriented approach is proposed to visualize the meaning of the content in a pseudo-lingual manner. Another critical piece, namely the mapping of manifest response types to competence-specific competence levels has been visualized to provide transparency.

In a first test the model is applied to a mathematical domain with 4 groups including 76 learners in a 75-minute e-assessment. This assessment is based on a web-based questionnaire and the results are documented and evaluated using the CSP as well as Access, Excel and R. Learner's responses are used to calculate manifest scores that are comparable to the results of a paper-and-pencil-Assessment as well as to estimate competence traits (with the psychometric RASCH-model).

A first group of results show:

- **A significant correlation between manifest results of learners in the CSP-based e-assessment and the parallel paper-and-pencil assessment designed by teachers themselves.**
- **A highly significant correlation between the CSP-based manifest scores and the estimated traits.**

This indicates that scoring based on the CSP e-assessment and the classical paper-and-pencil approach is similar. Although this has to be analyzed in more detail an indication for comparable judgement can be found. The correlation between trait and manifest results seems to indicate that estimated competencies are valid beyond the actual assessments.

As expected from their design, parts of the assessment items are grouped according to specific domain contents or competencies. Different manifest and latent variables are defined and analyzed, *e.g.* exponential or linear functions and the meaning of their parameters to describe different aspects of growth.

Hereby, a multidimensional analysis is carried out to show the possibility for delivering both manifest results and estimated traits related to specific aspects with even a small amount of data.

This multidimensional (multi-variable) analysis facilitates the use of these traits to support the validity of manifest scores for these groups and the detailed aspects of the domain.

As a third aspect of evaluation, individual and group feedback is addressed. The variables of the multidimensional view are traced back to the CORE-model descriptions.

- **Strengths or weaknesses in these parts of the contents and competencies may be responsible for the results and can be included in a detailed feedback. Thus single variables of the multidimensional view may be used as "evidence" for the achievement of specific parts of the domain.**

To highlight the possibility to provide dedicated support an exemplary description of parts of the education process itself (which may be enriched with additional educational material) are described in a sub-model of the CSP. This is linked to the specific variables as used before.

- **Through the description in the CSP it becomes possible to give further teaching and learning advice on specific domain subjects or expertise. This may help learners as well as the teacher.**

Thus, based on the CSP it is possible to get a manifest score, an estimated trait of competence and formative feedback from the same assessment.

The effort that is needed to use the CSP has two aspects. Initially, the CSP needs additional effort to describe the domain content as well as to develop the assessment itself which may consume a few weeks. But while the domain description is an additional time consuming input the reusability of the assessment as a whole or at least part of it will be worth the increased effort. The second aspect is the effort necessary for correction which is remarkably less than for a classic paper-and-pencil-Assessment. Here a reduction of 90% in groups like the 76 learners in the example seems not to be unrealistic. Effort reductions is expected to increase or decrease with a respective change in group size.

Obviously the one, relatively small test presented does not provide firm evidence of the results presented regarding both the manifest and psychometric outcome and its comparison to classical paper-and-pencil-Assessments. Yet, there are good indicators for correlation and although differences remain, it is open to discussion which of the two assessment types delivers more valid results with regards to educational goals.

One great benefit of using the CSP is the structured description of major parts of what has to be achieved within the educational process. There is a traceable structure which shows the relationship between different parts of the domains that may also be used to build assessment items related to the domain content and the underlying competencies. These assessments may produce manifest scores and psychometric traits.

Thus the CSP may support the possibility to focus on the goals of the educational process as well as the measurement of the achievement of these goals. The externalization of goals, competencies and content in relationship to an assessment is a solid ground for any feedback given to the educational process stakeholders.

A further aspect is the economy of resources. The model may offer substantial savings particularly by using it in multiple groups of learners, reuse assessments or parts of it and centralize the usage as much as possible. A stronger support by a dedicated software environment is recommended, yet this goes beyond the scope of this work.

Moreover, the model has to be tested with other domains which are possibly not as suitable for a structured domain content description as mathematics. Here, further research is needed.

INHALTSVERZEICHNIS

1	EINLEITUNG.....	1
2	ZIEL, AUFGABENSTELLUNG UND VORGEHEN	3
2.1	ZIEL	3
2.2	AUFGABENSTELLUNG.....	5
2.3	VORGEHEN.....	7
3	BILDUNGSZIELE UND -PROZESSE	8
3.1	BILDUNGSZIELE	8
3.2	DER BEGRIFF ASSESSMENT	9
3.3	KOMPETENZ – PERFORMANZ - QUALIFIKATION.....	12
3.3.1	<i>Kompetenz nach Weinert</i>	12
3.3.2	<i>Performanz</i>	13
3.3.3	<i>Qualifikation</i>	14
3.3.4	<i>Übersicht</i>	16
3.4	DER KOMPETENZBEGRIFF	18
3.4.1	<i>Historie</i>	18
3.4.1.1	Sozialwissenschaftliche Kompetenz	18
3.4.1.2	Pädagogische Kompetenz.....	19
3.4.2	<i>Nicht kognitive Fähigkeiten</i>	20
3.4.3	<i>Individualkompetenz</i>	23
3.4.4	<i>Handlungskompetenz</i>	23
3.4.5	<i>Kompetenz im engeren Sinne</i>	24
3.4.6	<i>Kontextabhängigkeit</i>	26
3.4.7	<i>Domänen</i>	27
4	MESSPROZESS	29
4.1	ASSESSMENTS.....	29
4.1.1	<i>Assessments im Zusammenhang mit Kompetenz und Qualifikation</i>	29
4.1.2	<i>Funktion und Bewertungsform</i>	30
4.1.3	<i>Messung und Messbarkeit</i>	32
4.1.3.1	Skala	32
4.1.3.2	Objektivität.....	32
4.1.3.3	Reliabilität	33
4.1.3.4	Validität	33
4.1.3.5	e-Assessments.....	35
4.2	KOMPETENZDIAGNOSTIK	36
4.2.1	<i>Forschungstraditionen</i>	36
4.2.2	<i>Kompetenzmodellarten</i>	37
4.2.2.1	Kompetenzstrukturmodelle	38
4.2.2.2	Kompetenzniveaumodelle.....	38
4.2.2.3	Kompetenzentwicklungsmodelle	41
4.2.3	<i>Quantitative Kompetenzdiagnostik</i>	41
4.2.4	<i>Individualfeedback</i>	42
4.2.5	<i>Psychometrische Modelle</i>	43
4.2.5.1	Grundidee des RASCH-Modells	43
4.2.5.2	Voraussetzungen.....	45
4.2.5.3	Einhaltung der Modellvoraussetzungen.....	46
4.2.5.3.1	Konstruktive Schritte	46
4.2.5.3.2	Graphischer Test.....	46

4.2.5.3.3	Andersen LR-Test	47
4.2.5.3.4	Wald-Test.....	48
4.2.5.3.5	Weitere Tests	48
4.3	ASSESSMENT-MODELLE.....	48
4.3.1	<i>Assessment Triangle</i>	48
4.3.2	<i>Evidence-Centered Assessment Design (ECD)</i>	49
4.3.2.1	Domänenanalyse	50
4.3.2.2	Domänenmodellierung.....	51
4.3.2.3	Konzeption des Assessment-Modells	51
4.3.2.4	Entwicklung des Assessments (Assessment implementation).....	52
4.3.2.5	Einsatz des Assessments (Assessment delivery)	52
4.3.3	<i>Die Four Building Blocks</i>	52
4.4	EINORDNUNG IN DEN BILDUNGSPROZESS	53
4.5	PÄDAGOGISCHE PROBLEME.....	57
4.5.1	<i>Bildungsideal und qualitative Ansätze</i>	57
4.5.2	<i>Input- und Prozessorientierung</i>	59
4.5.3	<i>Realitätsnähe</i>	59
4.5.4	<i>Ansätze der Leistungsbewertung</i>	60
5	ANSATZ.....	62
5.1	ARCHITEKTUR.....	62
5.2	ANFORDERUNGEN	62
5.2.1	<i>Domäneninhalt</i>	62
5.2.2	<i>Assessment</i>	63
5.2.3	<i>Evidenz</i>	64
5.2.4	<i>Ressourcen</i>	64
5.3	EINORDNUNG IN EINE METAMODELL-ARCHITEKTUR.....	65
5.4	IDEE DES CORE-SITUATION-PERSON-MODELLS (CSP)	68
5.5	VORGEHEN	69
6	CSP - INFORMATIONSMODELL.....	71
6.1	ARCHITEKTUR DES CSP-MODELLS	71
6.1.1	<i>Strukturelle Modellierungsentscheidungen</i>	71
6.1.2	<i>Die CSP-Ebenen</i>	72
6.2	DAS CORE-MODELL	73
6.2.1	<i>Überblick</i>	73
6.2.1.1	Allgemeiner Teil des CSP-Modells	73
6.2.1.2	Die Bereiche des CORE-Modells	74
6.2.2	<i>Modellbereich Content</i>	75
6.2.2.1	Ausgangspunkt natürliche Sprache	76
6.2.2.2	Anforderungen an den Modellbereich Content	77
6.2.2.3	Formale Beschreibungsmethoden.....	78
6.2.2.3.1	Prädikatenlogik	78
6.2.2.3.2	Formale Sprachen	78
6.2.2.3.3	Ontologien	78
6.2.2.3.4	Folgerungen für die Auswahl	80
6.2.2.4	Semi-formale Darstellung.....	80
6.2.2.5	Auswahl von NIAM als Basis	81
6.2.2.6	Modellierung des Contents	83
6.2.2.7	Das konzeptionelle Modell für den Bereich Content.....	87
6.2.2.8	Bereich Content – Wissenstyp prozedurales Wissen.....	88
6.2.2.9	Bezug zu den detaillierten Anforderungen	90
6.2.2.9.1	Bewertbarkeit (1).....	90

6.2.2.9.2	Sprachliche Nähe (2).....	91
6.2.2.9.3	Beispiele (3).....	91
6.2.2.9.4	Wissensinseln (4).....	91
6.2.2.9.5	Widersprüche (5).....	91
6.2.2.9.6	Graphische semi-formale Darstellung (6A).....	92
6.2.2.9.7	Formale Umsetzbarkeit (6B).....	92
6.2.2.9.8	Abbildbarkeit in einem Metamodell (6C).....	92
6.2.3	Modellbereich Item.....	92
6.2.4	Modellbereich Skill (Fähigkeiten).....	94
6.2.5	Modellbereich Competence Specifications.....	97
6.2.6	Modellbereich Kompetenzdimension.....	98
6.2.7	Modellbereich Feedback.....	104
6.3	DAS SITUATION-MODELL.....	106
6.3.1	Modellbereich Assessment.....	107
6.3.2	Modellbereich Score.....	110
6.3.3	Modellbereich Statistik.....	112
6.4	DAS PERSONAL-MODELL.....	114
7	CSP - VORGEHENSMODELL.....	117
7.1	ÜBERBLICK.....	117
7.2	BEZÜGE ZU ANDEREN ASSESSMENT-MODELLEN.....	120
7.2.1	Bezug zum Assessment Triangle.....	120
7.2.2	Bezug zum ECD-Modell.....	120
7.2.3	Bezug zu den Four Building Blocks.....	121
7.3	BEZUG ZUM INFORMATIONSMODELL.....	122
7.4	SCHRITT 0 – INITIALIZATION.....	124
7.5	SCHRITT 1 - EXTERNAL ANALYSIS.....	125
7.6	SCHRITT 2 - COMPETENCE MODELING.....	126
7.7	SCHRITT 3 - CONTENT MODELING.....	129
7.8	SCHRITT 4 - ABSTRACT ASSESSMENT MODELING.....	132
7.8.1	Operationalisierung zu Items.....	133
7.8.2	Verwendung von Skills.....	133
7.8.2.1	Competence Niveau Maps.....	134
7.8.2.2	Operatoren und Kompetenzniveaus.....	137
7.9	SCHRITT 5 – ASSESSMENT MODELING.....	138
7.9.1	Assessmentbeschreibung.....	138
7.9.2	Statistisches Modell.....	139
7.9.3	Item-Design.....	139
7.9.4	Item-Gruppierung.....	140
7.9.5	Response-Modeling.....	140
7.10	SCHRITT 6 - ASSESSMENT EXECUTION.....	144
7.11	SCHRITT 7 – OUTCOME MODELING.....	145
7.11.1	Raw-Error-Korrektur.....	145
7.11.2	Outcome Space.....	146
7.12	SCHRITT 8 - FEEDBACK MODELING.....	147
7.13	SCHRITT 9 – PERFORMANCE ANALYSIS.....	148
7.14	SCHRITT 10 - MANIFEST SCORE.....	151
7.14.1	Gesamtscore pro Person.....	151
7.14.2	Analyse für Skills.....	152
7.14.3	Analyse für Facts.....	155
7.15	SCHRITT 11 – STATISTICAL ANALYSIS.....	156
7.16	SCHRITT 12 – COMPETENCE ESTIMATION.....	158

7.17	SCHRITT 13 – PERSONAL FEEDBACK.....	159
7.18	SCHRITT 14 – INTEGRATION	164
8	ANWENDUNG UND ERGEBNISSE	165
8.1	THEORETISCHE ABLEITUNG UND RAHMENBEDINGUNGEN	165
8.1.1	<i>Domäneninhalt.....</i>	165
8.1.2	<i>Korrektur der Durchführungsprobleme(Probleme)</i>	166
8.1.3	<i>Korrektur der RAW ERROR TYPEs</i>	166
8.1.4	<i>Analyse der Antworten (Response Modeling)</i>	167
8.2	ERGEBNISSE IM VERGLEICH ZU KLASSISCHEN KLAUSUREN	167
8.2.1	<i>Überblick</i>	167
8.2.2	<i>Vergleich der kompletten Assessments mit konventioneller Gewichtung.....</i>	169
8.2.3	<i>Vergleich der kompletten Assessments mit dichotomer Bewertung (ohne Gewichtung)</i>	169
8.2.4	<i>Vergleich ausgewählter Items mit Gewichtung</i>	170
8.2.5	<i>Vergleich ausgewählter Items (dichotom)</i>	170
8.3	ERGEBNISSE AUS DER STATISTISCHEN ANALYSE	170
8.3.1	<i>Erste statistische Analyse mit dem RASCH-Modell.....</i>	170
8.3.2	<i>Zweite statistische Analyse mit dem RASCH-Modell.....</i>	172
8.3.3	<i>Vergleiche der statistischen mit der manifesten Analyse.....</i>	172
8.4	ERGEBNISSE AUF KOMPETENZEBENE	173
8.4.1	<i>Ergebnisse auf Skill- und Faktenebene</i>	173
8.4.2	<i>Scoring mit Anforderungsniveaus</i>	174
8.5	INDIVIDUELLE ANALYSE.....	175
8.6	FEEDBACK.....	177
8.7	ZEITBEDARF.....	180
9	ZUSAMMENFASSUNG.....	182
10	AUSBLICK.....	185
10.1	WEITERE AUFGABEN UND DOMÄNEN	185
10.2	WEITERE TEILMODELLE	186
10.2.1	<i>Qualifikationen.....</i>	186
10.2.2	<i>Competence Intensive Prozess (CIP).....</i>	189
10.2.3	<i>Idee eines Modellbereichs Situation Process Model.....</i>	193
10.3	SOZIO-TECHNISCHE SYSTEME.....	194
A	ANHANG.....	196
A.1	VORGEHENSWEISE BEI DER STATISTISCHEN ANALYSE.....	196
A.2	ERLÄUTERUNGEN ZUR ERSTEN STATISTISCHEN ANALYSE	197
A.3	ERLÄUTERUNGEN ZUR ZWEITEN STATISTISCHEN ANALYSE	203
A.4	ERLÄUTERUNGEN ZU DEN TYPEN DES CSP-INFORMATIONSMODELLS	205
A.4.1	<i>Modellbereich CONTENT</i>	205
A.4.2	<i>Modellbereich ITEM</i>	207
A.4.3	<i>Modellbereich SKILL.....</i>	208
A.4.4	<i>Modellbereich COMPETENCE SPECIFICATION</i>	210
A.4.5	<i>Modellbereich COMPETENCE DIMENSION</i>	211
A.4.6	<i>Modellbereich FEEDBACK</i>	213
A.4.7	<i>Modellbereich ASSESSMENT</i>	216
A.4.8	<i>Modellbereich SCORE</i>	221
A.4.9	<i>Modellbereich STATISTIK.....</i>	224
A.4.10	<i>Modellbereich PERSONAL.....</i>	227

B.	FEHLERANALYSE	230
B.1	ANALYSE DER KOMBINATION ZWEITER ITEMS	230
B.2	ANALYSE EINES EINZELNEN OFFENEN ITEMS.....	233
B.3	ANALYSE EINZELNER ASSESSMENT-TEILNEHMER	236
B.4	NUTZUNG DES INDIVIDUALFEEDBACKS	237
B.4.1	<i>Feedback für die Assessment-Teilnehmer.....</i>	<i>237</i>
B.4.2	<i>Feedback für die Lehrperson.....</i>	<i>240</i>
C.	ABBILDUNGSVERZEICHNIS	242
D.	ASSESSMENT.....	245
E.	LITERATURVERZEICHNIS.....	262

Abkürzungsverzeichnis

BMBF	Bundesministerium für Bildung und Forschung
BPM	Business Process Modeling
BPMN	Business Process Model and Notation
CIP	Competence Intensive Process
CNM	Competence Niveau Map
CSP	Core-Situation-Personal
CTT	Classical Test Theory
DB	Datenbank
DFG	Deutsche Forschungsgemeinschaft
DIF	Differential Item Functioning
DQR	Deutscher Qualifikationsrahmen
ebd.	ebenda
ECD	Evidence-centered Design
i.e. S.	im engeren Sinne
EQR	Europäischer Qualifikationsrahmen
ICC	Item Characteristic Curve
IQB	Institut für Qualitätssicherung im Bildungswesen
IRT	Item Response Theory
KMK	Kultusministerkonferenz
MOF	Meta Object Facility
NAC	National Academy Press
NIAM	Nijssen Information Analysis Method
OMG	Object Management Group
OWL	Web Ontology Language
PISA	Programme for International Student Assessment
RDF	Resource Description Framework
SOLO	Structure of Observed Learning Outcome
SuS	Schülerinnen und Schüler
TIMSS	Trends in International Mathematics and Science Study
UML	Unified Modeling Language
WFMC	Workflow Management Coalition
YAWL	Yet Another Workflow Language

1 Einleitung

Im Umgang mit komplexen Problemstellungen ist die Verfügbarkeit von Kompetenzen, Qualifikationen, Fähigkeiten, Strategien und Wissen, um einige der in diesem Zusammenhang betrachteten Lösungsaspekte zu nennen, wesentlich. Diese Lösungsaspekte führen zusammen mit einer adäquaten Zielsetzung und einer entsprechenden zielgerichteten Anwendung zu erfolgreichen Problemlösungen in vielen Bereichen der Erfahrungsumwelt.

Sie können insbesondere im Zusammenhang mit beruflichen Anforderungen, aber auch mit Studium, Ausbildung und Weiterbildung, mit Inhalten im Rahmen der schulischen Bildung, wie auch im Umgang mit komplexen sozio-technischen Systemen stehen. Die Überprüfung der Verfügbarkeit benötigter Lösungsaspekte bei einzelnen Personen ist dann eine zentrale Fragestellung. So kann eine Messung beispielsweise in der Situation eines Einstellungsprozesses oder eines Aufnahmeverfahrens von Interesse sein. Sie ist aber auch ein Teil von Bildungsprozessen, deren Ziel gerade der Erwerb derartiger Aspekte zum Umgang mit komplexen Problemstellungen ist. Deren Messung ist dabei eine zentrale Basis jedes Feedbacks für Lernende und Lehrende, sei es als formatives Ergebnis in Form einer Beobachtung, als Bewertung oder Beurteilung. Daher sollen in dieser Arbeit derartige Überprüfungen, speziell in Form von Assessments, am Beispiel von Bildungsprozessen betrachtet werden.

Den zentralen Lösungsaspekt stellen in diesem Umfeld derzeit die Kompetenzen dar. Das Ziel von Bildungs- und Lernprozessen ist somit der Aufbau von Kompetenzen bei den Lernenden. Ob es sich dabei auch um Wissen und Können, um Fähigkeiten oder Fertigkeiten, um Qualifikationen oder Strategien handelt, die in privaten oder beruflichen Lebenssituationen genutzt werden können, ist selbst eine begrifflich nicht leicht zu lösende Problematik.

Wesentlich ist aber, dass eine Zielvorstellung formuliert wird. Daraus ergibt sich dann die Frage, wie das Erreichen des Ziels gemessen werden kann. So sind in schulischen, universitären, beruflichen und anderen Bereichen zahlreiche Wege zur Bestimmung des Erfolgs von Bildungsprozessen gesucht und beschritten worden. Viele dieser Wege beruhen auf Leistungsmessungen, die in Form von schriftlichen oder mündlichen Prüfungen, von manuell oder elektronisch beantworteten Fragebögen, von Beobachtungsprotokollen oder Entwicklungsberichten vorliegen. Die meisten der so ermittelten Ergebnisse sind zunächst Beobachtungen, aus denen durch Zuordnung von Messgrößen Bewertungen und schließlich Beurteilungen werden.

Bei einer Messung ist die Frage nach der Messbarkeit der Zielerreichung und der Definition dieser Ziele für unterschiedliche Anspruchsgruppen relevant, für den Lernenden, für den Lehrenden und für andere dritte Stakeholder des Bildungsprozesses. Je präziser allen diesen Personen ein Feedback gegeben werden kann, umso besser kann der Prozess in seinem weiteren Verlauf individuell angepasst und am Ende der Erfolg in Form der Zielerreichung beurteilt werden.

Gerade das detaillierte Feedback für Lernende und Lehrende während des Prozesses kann als ein zentraler Erfolgsfaktor für den Prozess selbst angesehen werden. Problematisch ist dabei allerdings, dass ein solches Feedback zum einen einer detaillierten Zieldefinition und Inhaltsbeschreibung bedarf, zum anderen mit erheblichem Aufwand verbunden sein kann.

Leistungsüberprüfungen, Assessments, sind oft komplex und bedürfen eines hohen Korrektur- und Analyseaufwandes. Auf der anderen Seite existieren bei Multiple Choice Tests Bedenken

hinsichtlich der Überprüfung komplexer Themen oder sie bedürfen zumindest eines hohen Design- und Normierungsaufwandes.

Ein Grundproblem ist bei vielen Assessments bereits, dass keine operationalisierte Beschreibung der zu überprüfenden Aspekte vorliegt. Diese ist aber die Voraussetzung für ein detailliertes Feedback, wenn es sich an Zielen und Inhalten orientieren soll. Daher wird in dieser Arbeit ein Modell für eine solche detaillierte Beschreibung vorgestellt, das nicht an bestimmte Themen oder Domänen gebunden ist. Dieses Modell wird mit einer Beschreibung von modularen, wiederverwendbaren Assessmentinhalten gekoppelt. Mittels eines e-Assessments mit unterschiedlich offenen Fragestellungen wird die Ähnlichkeit der Ergebnisse zu einer klassischen Klausur exemplarisch getestet. Insbesondere wird die Erstellung kompetenzbezogener, inhaltlich detaillierter Feedbacks angestrebt, die durch eine Kombination klassischer Bewertung (Scoring) und statistischer Analyse gestützt werden können.

Insgesamt ergibt sich so in der Durchführung des eigentlichen Assessments eine erhebliche Zeiteinsparung, während umgekehrt mehr Aufwand in die Beschreibung der Ziele, Kompetenzen, Inhalte und in die Konstruktion des Assessments einzubringen ist. Daher wird diese Modellierung unter dem Aspekt der Wiederverwendbarkeit auf allen Ebenen, von der Zielformulierung bis zum konkreten Assessment durchgeführt. Damit können zu einmal beschriebenen Domänen neue Assessments konstruiert werden, ohne dass eine erneute Modellierung erforderlich ist. Es können aber auch Teile oder komplette Assessments wiederverwendet werden.

Das Modell wird mittels der Beschreibung eines Ausschnittes der Mathematik der 11. Klassenstufe eines Gymnasiums und der Auswertung eines entsprechenden Assessments dargestellt und evaluiert.

2 Ziel, Aufgabenstellung und Vorgehen

2.1 Ziel

Zunächst sind die Hintergründe der augenblicklichen Orientierung der Bildungsprozesse zu betrachten. Die Output-Orientierung ist eine zentrale Neuausrichtung der Pädagogik und als Folge der ersten Pisa-Studien und folgender Evaluationen zu sehen. Durch die Neuentwicklung der Bildungsstandards wurde die Output-Orientierung ebenso zentral (Saldern 2011, 73) wie der Leitbegriff Kompetenz (BMBF 2009, 9).

Kompetenzen sind eng mit der Umsetzung der Output-Orientierung verbunden, wie beispielsweise Schott und Ghanbari betonen: „Im Rahmen der Forderung nach einer verstärkten Output-Orientierung des Bildungswesens und nach überprüfbaren Bildungsstandards hat der Kompetenzbegriff zentrale Bedeutung erlangt.“ (Schott und Ghanbari 2008, 13). Entwicklungen in der beruflichen Bildung haben zu einer Betonung von Konzepten wie Output-Orientierung, Kompetenzvermittlung, insbesondere Handlungskompetenz, und in der Folge zu den Versuchen einer entsprechenden Messung dieser Kompetenzen geführt. „Die Modellierung und Messung von Kompetenzen spielt eine entscheidende Rolle bei der Optimierung von Bildungsprozessen sowie der Qualitätssicherung und Qualitätsentwicklung des Bildungswesens.“ (Leutner, Klieme, et al. 2013, 2).

Nun wird die Erschließung der Kompetenzen nicht allein über deren Messung zu realisieren sein, aber ohne eine Messung ihres Effektes bleiben alle Maßnahmen im Diffusen. Klieme und Hartig schreiben bereits: „Modellierung und Messung von Kompetenzen sind sicherlich nicht die einzigen Aufgaben, die sozial- und erziehungswissenschaftliche Forschung [...] zum Kompetenzbegriff bewältigen muss. Aber ohne Fortschritte in diesem Bereich sind differenziertere Ergebnisse etwa in der Erforschung von Schul- und Unterrichtseffekten nicht zu erwarten.“ (Klieme und Hartig 2007, 26).

Zur Motivation einer Kompetenzmodellierung für eine Messung sei die Expertise zur Entwicklung nationaler Bildungsstandards zitiert: „Kompetenzmodelle stellen damit die Grundlage für Operationalisierungen von Bildungszielen dar, die den Output des Bildungssystems über das Erstellen von Testverfahren [...] empirisch zu überprüfen erlauben. Die Rolle von Kompetenzmodellen besteht darin, zwischen abstrakten Bildungszielen und konkreten Aufgabensammlungen zu vermitteln.“ (BMBF, Zur Entwicklung nationaler Bildungsstandards 2003, 71)

Eine Messung muss ihre Ergebnisse an vorgegebenen Maßstäben, also Zielen, ausrichten. Bei der Erarbeitung des Modells muss also eine kompetenzorientierte Dokumentation der Ziele eines Bildungsprozesses erfolgen. Damit erklärt sich auch, dass Kompetenzmodelle die Rolle übernehmen, die zuvor Lehrziele hatten. Beide beschreiben die zu erwerbenden Fähigkeiten als Ziel des Bildungsprozesses. (Schott und Ghanbari 2008, 14).

Wenn spätestens seit den PISA und TIMSS-Assessments die zentrale Überprüfung eines Leistungsstandes in den Fokus der Bildungsprozesse gerückt ist, bedeutet dies einerseits, dass diese großen zentralen Assessments Gegenstand der Diskussion der Bildungsprozesse im Ganzen sind, andererseits stellt sich die Frage, wie im normalen Betrieb von Bildungseinrichtungen auf „classroom-level“ objektive, reliable und valide Assessments stattfinden sollen. Auf diesem

„Level“ finden bereits, nicht-standardisierte, Assessments statt, um beispielsweise individuelles Feedback zum Lernprozess zu erhalten. Zu klären ist dann die Frage, welche die zu erwerbenden und überprüfenden Kompetenzen sind und wie diese zuverlässig und mit sinnvollen, überprüfbaren und praktisch nutzbaren Ergebnissen gemessen werden sollen. (Klieme, Hartig und Rauch 2008, 13)

Hierbei ist zwischen einer Beobachtung, einer Bewertung und einer Beurteilung der Ergebnisse derartiger Assessments zu unterscheiden. Gerade die Beobachtung der Ergebnisse kann allen am Bildungsprozess Beteiligten ein wertvolles Feedback für dessen Justierung liefern und ist ein zentraler Erfolgsfaktor für den gesamten Prozess (Hattie 2013).

Individuelles Feedback bedeutet im Bereich der kompetenzorientierten Zielsetzung die Messung von Kompetenzen und deren Nutzung zu einem Feedback für die am Bildungsprozess beteiligten Personen. Wichtig erscheint das Feedback für den einzelnen Lernenden wie auch die Lehrenden. Feedback zeigt sich in den „Top“- Faktoren erfolgreichen Lernens, wie sie John Hattie in seiner umfassenden Analyse sichtbar gemacht hat, so beispielsweise in

- „Selbsteinschätzung des eigenen Leistungsniveaus“ (Platz 1) (Hattie 2013, 47) oder
- „Formative Evaluation des Unterrichts“ (Platz 3) (Hattie 2013, 193).

Generell taucht „Feedback“ als eigenständiger Erfolgsfaktor auf Platz 10 (Hattie 2013, 193) auf. Die Arbeit von Hattie basiert auf einer umfangreichen Anzahl von Studien zu Bildungssystemen und strukturiert deren Ergebnisse. Man findet die Bedeutung von Feedback aber auch an anderen Stellen. So weist von Saldern darauf hin, dass sie im Rahmen der Leistungsfeststellung wesentlich ist. „Sie dient der Erfassung der Fortschritte der Schülerinnen und Schüler in Bezug auf ihre Sach-, Methoden-, Sozial- und Persönlichkeitskompetenz.“ (Saldern 2011, 85).

Auch wenn, wie in Abschnitt 4.2.4 festgestellt wird, Feedback im Rahmen der Individualdiagnostik als komplex angesehen wird, schließt das deren Sinnhaftigkeit nicht aus. Insbesondere gilt dies auch für das Feedback für die Lehrenden wie von Hattie (Hattie 2013, 206) betont wird.

So wird hier nicht wie bei den großen Qualitätsuntersuchungen wie PISA und TIMSS nach einem Feedback für die Verantwortlichen im politischen Bereich wie auch für andere am Bildungsprozess im Großen beteiligten Interessengruppen gesucht. Vielmehr geht es um das Feedback im Kleinen, („classroom-level“). Denn es besteht die Notwendigkeit einer individualdiagnostischen Leistungsrückmeldung, die insbesondere für formative Evaluationen eine selbstgesteuerte und einfache Anwendung ermöglichen soll (Pant 2013, 76).

Auch in der Praxis werden in Leitfäden für den Unterricht hinsichtlich der Leistungsbeurteilung

- Transparenz,
- zügiges Feedback,
- individuelles Feedback,
- Trennung der Phasen

gefordert (Meyer 2004, 117ff.).

Dabei ist eine Note zu wenig aussagekräftig. Eine inhaltlich detaillierte Aussage ist wesentlich und wird in der Realität versucht durch Zusatzinformationen zu erhalten (Saldern 2011, 85). In dieser Arbeit sollen Noten nicht generell in Frage gestellt werden. Allerdings bedürfen sie eines

ergänzenden, detaillierteren und problemorientierten Feedbacks. Schnelles, häufiges und zielgerichtetes Feedback erfordert eine detaillierte Beobachtung und Beurteilung durch die Lehrkraft und erhöht den ohnehin großen Aufwand bei der Bewertung von Leistungsüberprüfungen, insbesondere in höheren Klassenstufen. Daher müssen ressourcensparende Assessments präferiert werden, wenn sie ähnliche Ergebnisse wie andere Verfahren liefern.

Eine zentrale Rolle hat gerade im Zusammenhang mit den großen Assessments die Verwendung von e-Assessments und statistischen Modellen erlangt. Schwierigkeiten bei quantitativen insbesondere statistischen Modellen ergeben sich aber aus dem fehlenden Verständnis und der daraus resultierenden fehlenden Akzeptanz der Lehrenden, so dass diese Modelle nicht immer als hilfreich empfunden werden.

Fraglich ist daher, inwiefern diese statistischen Modelle auf „classroom-level“ und für formative Assessments eingesetzt werden können. Sinnvoll ist es in jedem Fall, wenn „...formative und summative Assessments in einem bestimmten Kompetenzbereich auf denselben theoretischen Annahmen, d.h. auf den gleichen Kompetenzmodellen.“ (Bögeholz und Eggert 2013, 62) beruhen. Die Basis bilden dann eine gemeinsame Beschreibung der zu beobachtenden Inhalte und ein Verständnis der Lehrenden für diese Beschreibungen.

Die strukturelle Beschreibung von Systemen und Abläufen ist ein Schwerpunkt und eine Stärke der Informatik, insbesondere im Bereich der Analyse und Definition von Anwendungssystemen. Diese Arbeit verfolgt daher das Ziel, mittels derartiger existierender und zu entwickelnder Strukturen, die eine Anpassung an die pädagogisch-psychologische Bedürfnisse erfordern, eine detailliertere und strukturierte Dokumentation von Kompetenzen zu ermöglichen.

2.2 Aufgabenstellung

Das Ziel ist, Beschreibungen zu finden, die das Ziel des Bildungsprozesses inhaltlich darstellen, sich als Basis für die Nutzung in Assessments eignen, mit statistischen und rein deterministischen Methoden ausgewertet werden und einen Bezug zwischen Assessment-Ergebnis und Bildungsziel zum Zweck der Akzeptanz und des Feedbacks herstellen können.

Die Fragestellung lautet somit:

Lässt sich ein Modell für die Kompetenzbeschreibung ermitteln, das eine Leistungskontrolle hinsichtlich der Verfügbarkeit dieser Kompetenzen in Form ressourcenschonender formativer und summativer Assessments mit einer Erklärungskomponente für das Feedback auf „classroom-level“ ermöglicht?

Im Detail bedeutet das, das Modell soll die **Modellierung und Dokumentation** von **Kompetenzstrukturen und Kompetenzniveaus**, von **Domänenstrukturen und Domäneninhalten**, von **Assessments** und deren **Evaluierung** erlauben. Die Analyse beinhaltet etwa auf „classroom-level“ die **Beobachtung und Bewertung** sowohl in **summativer Form (Scoring und Kompetenz)** als auch formativ in **inhalts- und kompetenzbezogener Form** mit dem Ziel der **Individualdiagnostik für die Lernenden** und **Gruppendiagnostik für die Lehrenden**. Die Analyse soll möglichst valide, objektive und reliable Ergebnisse **mit limitierten Ressourcen** ermöglichen und eine **Erklärungsfähigkeit und Lenkungsfunktion** besitzen.

Das Modell soll nicht nur beschrieben werden, sondern es soll auch eine Vorgehensweise entwickelt werden, wie es real zu nutzen ist, und diese Nutzung soll exemplarisch erfolgen.

Die Erklärungskomponente des Modells soll dazu genutzt werden können, individuelles Feedback als eine Kopplung an den Lernprozess zu bieten, so dass nicht nur ein individuelles Feedback möglich ist, sondern mit Hilfe des Modells auch Lernschritte, Lernpfade, Interventionen und Materialien zielgerichtet ausgewählt werden können, um den weiteren Lernprozess zu steuern. Es soll also eine Brücke zwischen Messung und Prozess geschaffen und die Basis für eine Intervention im Bildungsprozess gelegt werden.

Dabei sei bereits hier betont, dass die Konstruktion von Assessments wie sie hier erfolgt klassische Diagnoseverfahren der Leistungsbeurteilung nicht vollständig ersetzen, sondern diese ergänzen soll.

Bei der Beschreibung von komplexen Situationen haben sich nicht nur in der Informatik Modelle bewährt, also das Konzept, die wesentlichen Eigenschaften darzustellen und von als unwesentlich erachteten Aspekten zu abstrahieren. Die Informatik hat sich Modelle in vielfältiger Form zunutze gemacht. So ist die Entwicklung von Software im Wesentlichen durch Modellbildungen geprägt, die auf unterschiedlichen Abstraktionsebenen und unter unterschiedlichen Blickwinkeln die verschiedenen Aspekte eines Gesamtsystems beschreiben. Diese wesentlichen Aspekte werden dann mit vorstrukturierten Elementen beschrieben.

Hier sollen Beschreibungsformen, wie sie die Informatik kennt, genutzt, entwickelt und bereitgestellt werden, um domänenspezifische Inhalte kompetenzorientiert möglichst exakt und verständlich zugleich darzustellen.

- Im Fokus dieser Arbeit soll daher erstens eine strukturierte Dokumentation der Ziele, Kompetenzen und Inhalte eines Bildungsprozesses stehen. Als beispielhaftes System wird auf das System der Berufsbildenden Schulen zurückgegriffen und es wird an deren Bildungsprozess angeknüpft, da dieser letztlich eine Vorbereitung und Simulation realer Systeme darstellt. Dabei steht der Domäneninhalt im Fokus, den die Lernenden als Folge eines Bildungsprozesses zielführend und problemlösend in variablen Situationen kompetent nutzen können sollen. Diese Inhalte sollen als Kompetenz in entsprechenden Situationen beobachtbar sein.
- Damit entsteht die Gelegenheit, die Überprüfung von Kompetenzen und die Erzeugung eines zielgerichteten Feedbacks in die entsprechenden Assessments zu integrieren. So soll zweitens ein Vorschlag für die domäneninhaltsorientierte Entwicklung eines Assessments erarbeitet werden. Dies beinhaltet auch die Chance einer erhöhten Wiederverwendbarkeit.
- Drittens soll gezeigt werden, dass auf dieser Basis ähnliche Ergebnisse wie in klassischen Klausuren erzielt werden, aber zusätzlich eine an Inhalten und Kompetenzniveaus orientierte Bewertung der einzelnen Personen und der Gruppe möglich ist. So soll die Möglichkeit einer Unterstützung und Ergänzung der Diagnostik um ein Feedback für Lernende und Lehrende aufgezeigt werden, die eine Erklärungsfähigkeit besitzt und an Interventionen im Bildungsprozess angeknüpft werden kann.

Wesentlich ist in jedem Fall, dass die ermittelten Ergebnisse in ihrer Herkunft strukturiert nachvollziehbar sind. Dies soll mit vertretbarem Aufwand möglich sein. Daher soll auch der notwendige Aufwand in die Betrachtung einbezogen werden.

2.3 Vorgehen

Das Vorgehen für die Entwicklung eines solchen Modells beginnt mit der Diskussion der Bildungsprozesse und –ziele unter dem Blickwinkel der erwähnten Aspekte, insbesondere des Kompetenzbegriffs. Hierzu werden in Kapitel 3 zunächst die wesentlichen Begrifflichkeiten sowie die zugrundeliegende Vorstellung des Bildungsprozesses beschrieben.

Es folgt die Betrachtung von Messprozessen einschließlich der Bedeutung der IRT (Item Response Theory) als stochastisches Analysewerkzeug für diese Prozesse. Beispielhaft werden drei Assessment-Modelle vorgestellt. Anschließend werden Messprozesse als quantitative, output-orientierte Verfahren in den Bildungsprozess eingeordnet. Dies geschieht in Kapitel 4, das somit auf die grundlegenden Elemente eines Assessments sowie auf existierende Kompetenzmessmodelle eingeht. Dabei werden wie in Kapitel 3 schrittweise Anforderungen an das hier zu entwickelnde Modell festgelegt.

Daraus wird der Ansatz für das hier entwickelte Modell und die Zielsetzung konkretisiert. Den Schwerpunkt bilden die drei dann folgenden Kapitel. Zum einen wird das 3-Ebenen-Informationsmodell entwickelt und im Detail beschrieben. Es wird dabei als Core-Situation-Personal-Modell (kurz: CSP) bezeichnet. In Kapitel 5 werden die Anforderungen zusammengefasst und der grundlegende Ansatz wird dargestellt.

Das Kapitel 6 bildet einen Schwerpunkt mit der Entwicklung des CSP-Informationsmodells. Dazu wird auf die Begrifflichkeiten der Pädagogik insbesondere im Bereich der Kompetenz, Qualifikation und Fähigkeiten zurückgegriffen. Diese werden in das Modell übertragen und deren Zusammenhänge für dieses Modell festgelegt. Die Bedeutung und die Inhalte dieser Begriffe sind nicht immer eindeutig, daher müssen hier Entscheidungen getroffen werden, die nicht allen Sichtweisen entsprechen können. Ergänzend werden auch zwei graphische Darstellungen zur Beschreibung der Inhalte und zur Kompetenzbestimmung entwickelt.

In Kapitel 7 wird das dazugehörige 15-schrittige Vorgehensmodell für die Anwendung des Modells bei der Beschreibung der Domäneninhalte, Assessments und Analysen dargestellt.

Schließlich wird in Kapitel 8 das gesamte Modell in einem konkreten Assessment angewendet und die Ergebnisse werden punktuell analysiert, um die Anwendbarkeit zu demonstrieren, bevor in Kapitel 9 eine Zusammenfassung der Ergebnisse erfolgt.

3 Bildungsziele und -prozesse

3.1 Bildungsziele

Wenn es um die Messung des Erfolgs eines Bildungsprozesses geht, ist, wie bei jedem Prozess, zunächst zu klären, woran der Erfolg oder die Zielerreichung zu messen ist. Die Bedeutung einer solchen Ziel- oder Erfolgsmessung ist zentral für Prozesse, auch wenn sie nicht immer klar definiert wird (Soffer und Wand 2005, 663).

Hammer und Champy definieren aus dem Blickwinkel des Business Process Modeling (BPM) einen Prozess „[...] as a collection of activities that takes one or more kinds of input and creates an output that is of value to the customer“. (Hammer und Champy 1993, 35) Die „Kunden“ von Bildungsprozessen sind vielfältige Anspruchsgruppen. Die entsprechenden Vorgaben für diese Anspruchsgruppen und für das Schulsystem als Ganzes werden durch die Schulgesetze der einzelnen Bundesländer definiert. Dort findet man einen breiten Kanon ethischer Grundsätze, Wertvorstellungen wie Gerechtigkeit, Gleichberechtigung, Berücksichtigung ökonomischer und ökologischer Zusammenhänge, Konfliktmanagement, Selbstständigkeit und schließlich auch „[...] sich im Berufsleben zu behaupten und das soziale Leben verantwortlich mitzugestalten.“ (Niedersächsisches Schulgesetz 2015, 8). Diesen Kanon kann man als Kern zeitgemäßer schulischer Allgemeinbildung bezeichnen.

Allerdings wird in der Realität von den meisten am System Schule beteiligten gesellschaftlichen Anspruchsgruppen, insbesondere der Wirtschaft, den Universitäten, aber auch den Eltern und nicht zuletzt den Schülerinnen und Schülern selbst die „Zubringerfunktion“ von Schule für weiterführende Bildungseinrichtungen, für andere Schulformen, Universitäten oder für eine Ausbildung als zentrales Ziel gesehen (Saldern 2011, 51). Unterstrichen wird diese Zielrichtung gerade durch die Entwicklung der Bildungsstandards, die – ausdrücklich nicht im Widerspruch, sondern in Ergänzung zur Allgemeinbildung – die praktischen Aspekte dieser Zubringerfunktion definieren. „Sie decken nicht die ganze Breite eines Lernbereiches ab, sondern formulieren fachliche und fachübergreifende Basisqualifikationen, die für die weitere schulische und berufliche Ausbildung von Bedeutung sind und die anschlussfähiges Lernen ermöglichen.“ (Bildungsstandards der Kultusministerkonferenz 2005, 7).

Bei der Festlegung der Ziele werden die in den Schulgesetzen ausdrücklich formulierten und die von vielen gesellschaftlichen Gruppen implizit postulierten Ziele oft in weit gefassten Begriffen subsumiert. So wurde bereits früh vom deutschen Bildungsrat „[...] die Bewältigung von Lebenssituationen sowie Urteils- und Handlungsfähigkeit in den verschiedenen Bereichen des Lebens [...]“ als Ziel des Lernens bezeichnet (Deutscher Bildungsrat 1974, 54).

In dieser Arbeit wird im Interesse einer klareren Messbarkeit daher bewusst eine Output- oder „Outcome-Orientierung“ (Bildungsstandards der Kultusministerkonferenz 2005, 6) in den Mittelpunkt gerückt, wie sie auch im Bereich der Business Process Modellierung, wie oben zitiert, verwendet wird. Dies entspricht der breiten gesellschaftlichen Erwartung an die Ziele von Schule ebenso wie der Intention der Bildungsstandards. Die Bildungsstandards werden außerdem sukzessive zur Basis von zentralen Prüfungen wie dem Zentralabitur gemacht. Sie sind letztlich eine Reaktion auf die Folge des „PISA-Schocks“. (Schott und Ghanbari 2008, 13)

3.2 Der Begriff Assessment

Wenn die Ziele des Prozesses als bekannt angenommen werden, ist zunächst festzuhalten, dass eine Erfolgskontrolle im Bildungsprozess notwendig ist. Die bloße Definition von Absichten garantiert keine Wirkung, ohne Erfolgs- oder Wirkungskontrolle wird pädagogisches Handeln zu einer Art Prinzip „Hoffnung“ und verliert letztlich seine Berechtigung. (Heid 1994, 53f.)

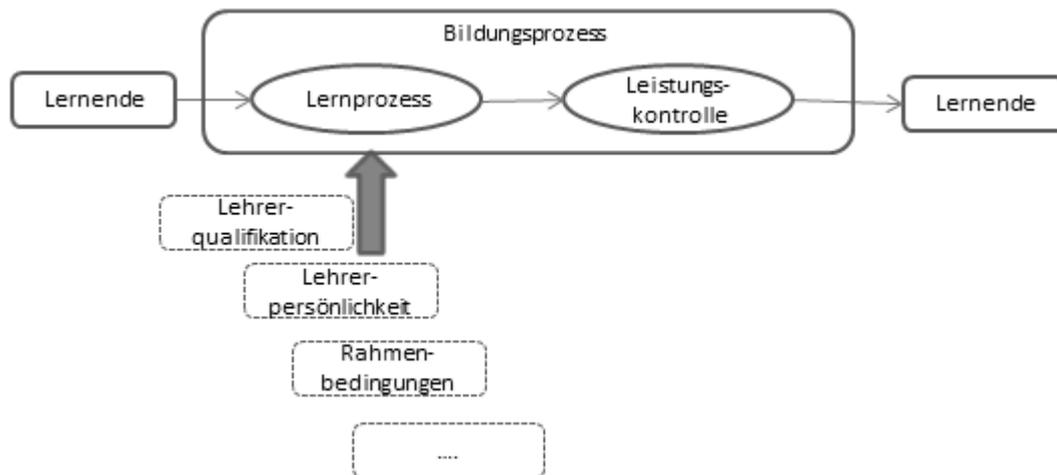


Abbildung 1 Erste Übersicht Bildungsprozess

Somit wird im Folgenden davon ausgegangen, dass sich der Bildungsprozess aus einem Lernprozess und aus Leistungskontrollen zusammensetzt, wie in Abbildung 1 dargestellt. Anknüpfungspunkt für die Leistungskontrollen sind die Zielsetzungen des Bildungsprozesses. Die durch den Bildungsprozess erzeugten Kompetenzen der Lernenden sind dessen Output. Dabei wirken externe Einflussfaktoren wie die Lehrerqualifikation und –persönlichkeit oder auch die Rahmenbedingungen und eine Reihe weiterer Komponenten des Gesamtsystems als Input mit. So hat allein Hattie 138 Einflussfaktoren aus zahlreichen Meta-Analysen in mehrjähriger Arbeit herauskristallisiert und deren Einfluss auf den Bildungsprozess, basierend auf diesen Analysen, dokumentiert (Hattie 2013).

So unterschiedlich die Zielformulierungen für den Bildungsprozess sein mögen, es bleibt unabhängig von der genauen Definition der Ziele festzuhalten, dass sie in irgendeiner Form durch eine Erfolgskontrolle gemessen werden müssen, soll der Bildungsprozess nicht nur dem oben erwähnten Prinzip „Hoffnung“ folgen - zumal auch zu klären wäre: „Hoffnung, worauf?“.

Der klassische Begriff in diesem Zusammenhang ist die Leistungskontrolle oder Leistungsbewertung. Leistungsbewertungen dienen im Wesentlichen der Ermittlung des Leistungsstandes der Lernenden, aber auch als Feedback für die Lernenden, für Lehrer, für Eltern und für andere am Bildungsprozess Beteiligte sowie letztlich auch als Erziehungsmittel. (Saldern 2011, 84ff.) Die Leistungsbewertungen stehen in der Praxis immer im Spannungsfeld von Schule an sich, von Selektion und Förderung. (Meyer 2004, 118).

Leistungsbewertungen erfolgen in der Praxis sowohl in klassischen Klausuren (paper-and-pencil-Leistungskontrollen) als auch prozessbegleitend in Form von Beobachtung der Schüler während des Unterrichtsprozesses. Hinzu kommen eine ganze Reihe weiterer Möglichkeiten wie Tests oder Projektarbeiten. Viele dieser Formen stellen für die geforderte Trennung von Lernphasen und

Bewertungsphasen (Meyer 2004, 118) (Abbildung 2) eine Herausforderung dar, da eine solche Trennung beispielsweise bei der Beobachtung und Betreuung von Projektarbeiten schwierig ist. Im Besonderen gilt dies auch für die veränderten Ansätze der Leistungsbewertung im Rahmen der „neuen Lernkultur“. Hier geht es in den Grundzügen darum, dass „[...] die Leistungsbewertung in ein engeres Wechselverhältnis zum Lernprozess [...]“ (Winter 2014, 3) gebracht werden soll.

Die Vielfalt von Leistungsbewertungen ist also groß. Um nicht immer auf die verschiedenen Arten möglicher Leistungsüberprüfungen eingehen zu müssen und nicht durch einen bestimmten Begriff eine Form zu implizieren, soll jede Form von Leistungsüberprüfung in dieser Arbeit als **Assessment** bezeichnet werden. Wichtig ist dabei, dass der Zweck des Assessments nicht unbedingt eine Bewertung oder Beurteilung beinhalten muss, allein die dokumentierte Beobachtung reicht aus.

Definition 1: Assessment

Ein Assessment stellt jede Art von Leistungsbeobachtung dar. Mit einem Assessment kann eine klassische Klausur, eine e-Klausur, ein Test, eine Projektarbeit, eine Beobachtung der Beteiligung am Unterricht oder auch ein Portfolio gemeint sein. Im Allgemeinen trifft die Vorstellung einer klassischen Klausur (paper-and-pencil) oder einer e-Klausur die Beschreibung am besten.

Gemeinsam ist allen Ansätzen von Assessments, dass bei der Beobachtung und gegebenenfalls der Bewertung und Beurteilung Transparenz hergestellt werden muss (Meyer 2004, 113ff.). In Abbildung 2 ist die Problematik der mangelnden Transparenz jeweils durch eine Wolke visualisiert. Mit mangelnder Transparenz in Form der Wolken ist gemeint, dass der Zusammenhang zwischen den Zielen und der Bewertung beziehungsweise Beurteilung nicht transparent wird.

Die schriftlichen Leistungsbewertungen (klassisch über Noten oder Notenpunkte) stellen eine summarische Bewertung dar, die nur in Ansätzen eine Analyse der Fehler und ein detailliertes Feedback mit Korrekturmöglichkeiten für den Lernenden bietet. Sie werden durch eine Reihe von Einzel**beobachtungen** aus dem Unterrichtsgeschehen ergänzt, die jeweils in Einzelbewertungen und schließlich in einer Gesamtbewertung enden. In diesem Zusammenhang soll von **Bewertung** gesprochen werden. Aus den Beobachtungen werden also durch Anlegen einer Messskala Bewertungen. Die einzelnen Bewertungen werden zu einer **Gesamtbeurteilung** zusammengefasst, die durchaus auch generelle weitere Eindrücke beinhalten kann, die im Rahmen der pädagogischen Freiheit einbezogen werden.

Zusammenfassend bedeutet dies, dass viele einzelne Beobachtungen aus dem Bildungsprozess und separate Leistungskontrollen zu einer Beurteilung eines Lernenden in einem Zeitraum, zumeist einem Schulhalbjahr führen. Die **Beurteilung** ist also die Zusammenfassung der Einzelbewertungen im Rahmen des Bildungsprozesses.

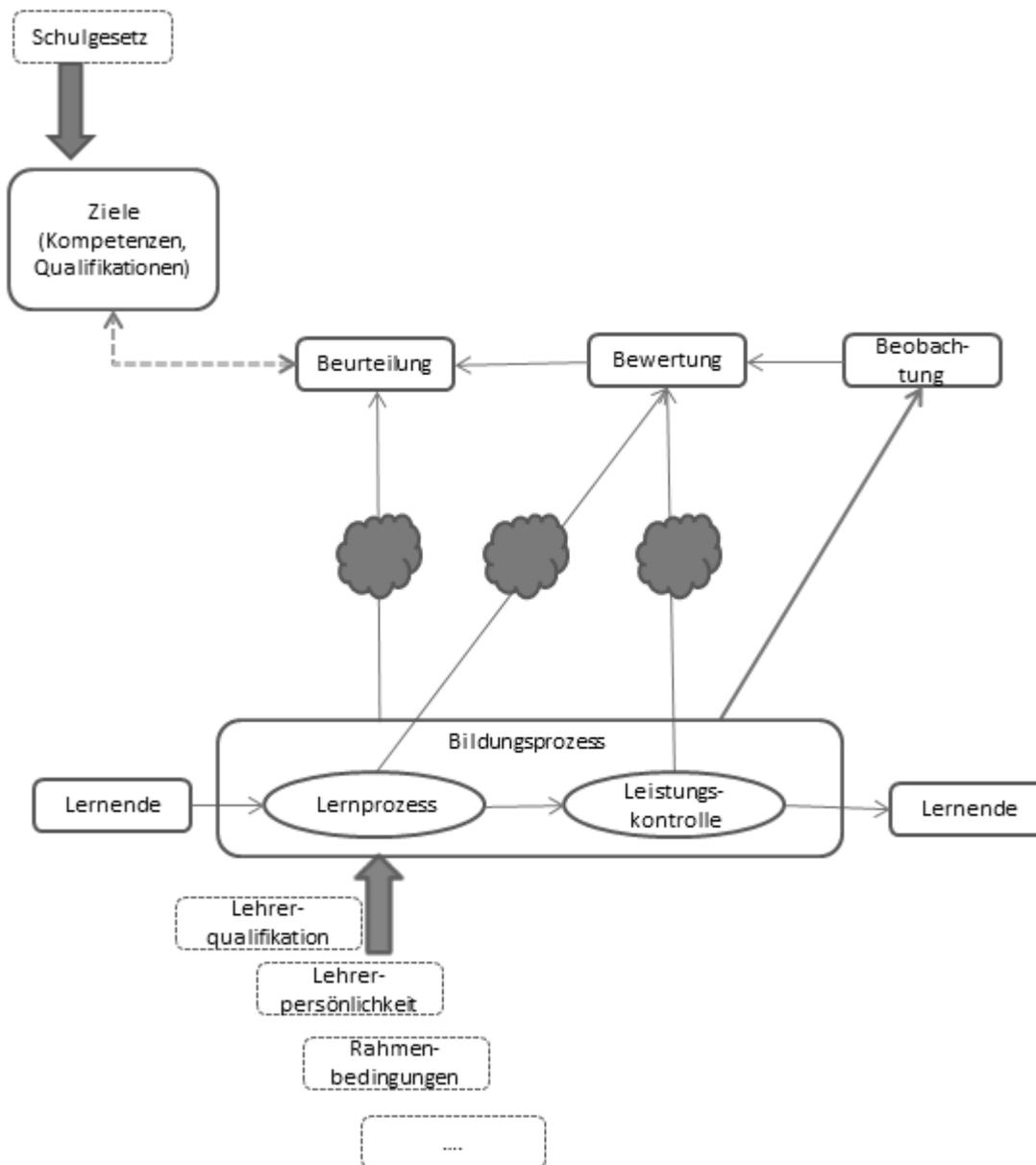


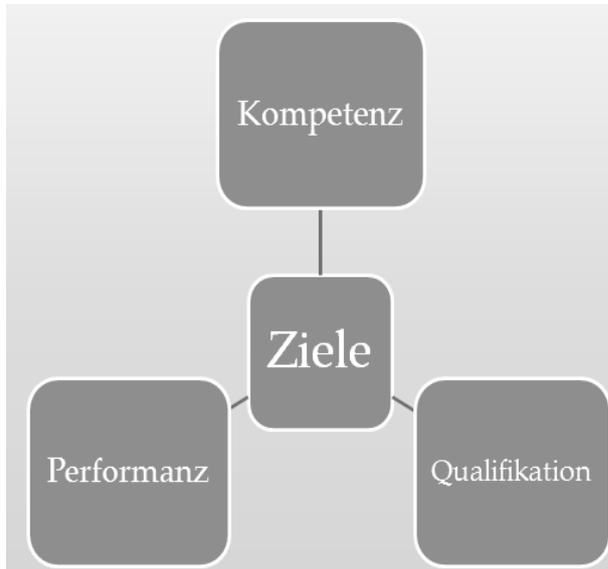
Abbildung 2 Bildungsprozess mit Bewertung und Beurteilung

Zwischen Beobachtung und Bewertung wird oft klar unterschieden, während Bewertung und Beurteilung in der Literatur nicht eindeutig definiert sind (Saldern 2011, 89) oder sogar oft synonym verwendet werden (Initiative Fortbildung e.V. 2013). Saldern unterscheidet fünf Phasen, die Beobachtung, die Beschreibung, die Bewertung, die Beurteilung und die Prognose (Saldern 2011, 90). Da es in dieser Arbeit immer auch um eine Beschreibung geht, können Beobachtung und Beschreibung zusammengefasst werden. Die Prognose ist wie die Beurteilung nicht Gegenstand dieser Arbeit. Sie kann daher als „Teil“ der Beurteilung aufgefasst werden und wird nicht direkt weiter betrachtet.

In dieser Arbeit wird also zwischen der formativen, rein **beschreibenden Beobachtung**, der **Bewertung** dieser Beobachtung, normalerweise durch einen Zahlenwert auf einer ordinalen oder kardinalen Skala, und der **Beurteilung** einer Gesamtleistung, die im schulischen Umfeld noch weiteren pädagogischen Aspekten unterliegt, unterschieden.

3.3 Kompetenz – Performanz - Qualifikation

Die Ziele wie sie im Schulgesetz definiert sind und wie sie von den Beteiligten erwartet werden wurden bereits angesprochen. In der Praxis werden diese Ziele mittels Curricula, einheitlichen Prüfungsanforderungen und anderen Mitteln interpretiert, operationalisiert und konkretisiert.



Parallel dazu werden mit Wirtschaftsvereinigungen wie der Industrie- und Handelskammer die Anforderungen an deren Abschlussprüfungen abgestimmt. Die dabei beschriebenen Inhalte können in Form von fachlichen Inhalten, Kompetenzen, Qualifikationen oder einer erwarteten Performanz dargestellt werden, sodass sich eine Reihe von Begriffen finden lässt, die anfangs unter dem Ausdruck „Lösungsaspekte für Problemstellungen“ subsummiert wurden und die im Rahmen der Zielvorstellung eines Bildungsprozesses adressiert werden können.

Abbildung 3 Zentrale Begriffe zur Zielbeschreibung

Drei zentrale Begriffe, die für Bildungsziele als Basis eines Assessments oft formuliert werden, sind in Abbildung 3 dargestellt.

Daher ist es sinnvoll zunächst diese Begrifflichkeiten inhaltlich voneinander zu trennen.

3.3.1 Kompetenz nach Weinert

Eine vielfach verwendete und zitierte Definition von Kompetenz stammt von Weinert. Danach sind Kompetenzen die „[...] bei Individuen verfügbaren und durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ (Weinert 2001, 27f.).

Diese Definition beinhaltet eine Reihe wesentlicher Aspekte. Zunächst wird deutlich, dass es sich um eine Eigenschaft von Individuen handelt, nicht von Gruppen, Institutionen oder von Situationen.

Der zweite Aspekt ist die Erlernbarkeit. Damit werden individuelle Dispositionen, die nicht durch einen Bildungsprozess erlernbar sind, ausgeschlossen, beispielsweise die Fähigkeit zu sehen oder zu hören.

Außerdem werden kognitive Fähigkeiten und Fertigkeiten erwähnt - Begriffe, die wohl auch mit „Wissen und Können“ oder „deklarativem“ und „prozeduralem“ Wissen übersetzt werden können (Wilhelm und Nickolaus 2013, 25). Eine akzeptierte Einteilung, die gelegentlich durch zusätzliche Wissensaspekte wie beispielsweise das „strategische Wissen“. (Achtenhagen und Baethge 2007, 57) oder die „Personalkompetenz“ (Kaufhold 2006, 96) erweitert wird.

Ergänzt wird die kognitive Komponente durch die motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die mit den kognitiven Komponenten verbunden sind und indirekt immer mitgemessen werden.

Die Anwendung sowohl der kognitiven als auch der anderen Komponenten wird auf „variable Situationen“ bezogen, das heißt die Kompetenz wird nicht auf eine konkrete Situation oder ein konkretes Assessment bezogen, sondern es steht deren allgemeine Nutzung im Vordergrund. Die erfolgreiche und verantwortungsvolle Nutzung zeigt die Zielgerichtetheit und die impliziten Werte (siehe auch Schulgesetze) als verborgene Ziele auf.

Der Aspekt der Beobachtbarkeit (und die daran anschließende Messung) wird bei Weinert bereits durch den Zusatz „um bestimmte Problem zu lösen“ wie auch die „Problemlösungen in variablen Situationen“ adressiert.

Die Definition von Weinert zeigt also viele wichtige Aspekte der Kompetenz. Sie wurde auch von der Expertise zur Entwicklung nationaler Bildungsstandards übernommen.

3.3.2 Performanz

Die Performanz beschreibt den beobachtbaren Einsatz der Fähigkeiten, Fertigkeiten und sonstiger Kompetenzbestandteile, die zum zielgerichteten Handeln in einer bestimmten Situation in einer bestimmten Domäne befähigen. (Saldern 2011, 36) Die „bestimmte Situation“ weist darauf hin, dass es sich um ein konkretes Assessment mit konkreten Ergebnissen für eine einzelne Person handelt, zumindest aber nicht um eine der Person in vielfältigen Situationen zugeordnete Eigenschaft.

Eine Abgrenzung zum Begriff der Kompetenz kann also darin liegen, dass in der Performanz im Regelfall alle Komponenten der individuellen Kompetenz zum Ausdruck kommen indem sie in einer konkreten Situation beobachtet werden. Dies führt zur Vorstellung von einem „Handeln in der Situation“. Abbildung 4 zeigt eine Gegenüberstellung, die die dem Individuum zugeordneten Kompetenzen der in einer konkreten Situation beobachteten Performanz zeigt.

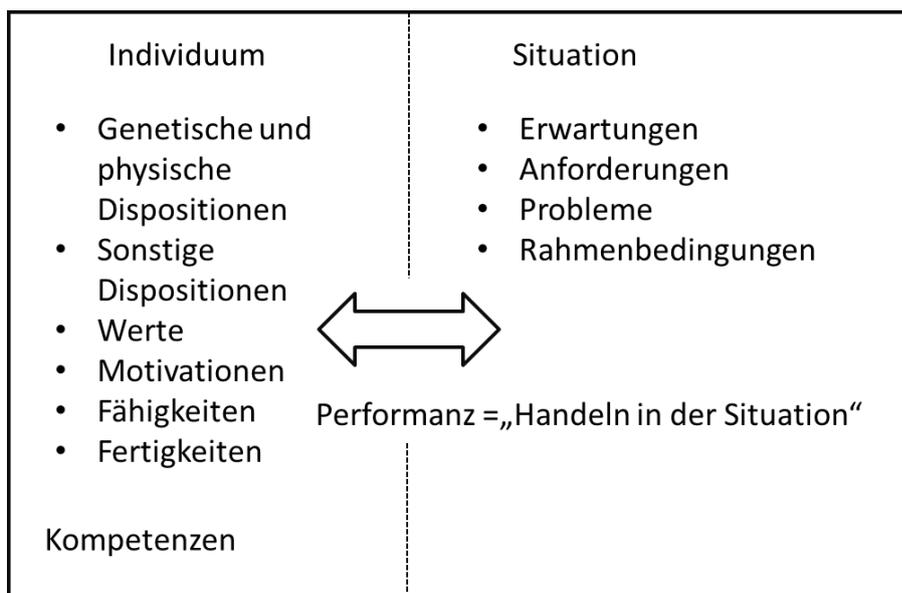


Abbildung 4 Kompetenz und Performanz (Gnahs 2007, 23)

Die der Kompetenz zugeordneten Begriffe werden im Rahmen der Kompetenzbetrachtung ausführlicher erläutert. Kompetenzen lassen sich nur durch Handeln beobachten. Das Handeln wird aber erst zur Performanz, wenn es in einer konkreten Situation geschieht. Die Situation wird durch die Probleme, Anforderungen, Erwartungen sowie die gegebenen Rahmenbedingungen beschrieben. Dies kann ganz klassisch eine Klausur sein, in der (in einem kaum dokumentierten Prozess) aus den von Zielen abgeleitete Anforderungen in Form von Problemstellungen, also von formulierten Aufgaben, im Mittelpunkt stehen.¹ Die Erwartungen werden in den Erwartungshorizonten beschrieben und die Bewertung erfolgt durch einen Abgleich der Leistungen der Assessment-Teilnehmer mit diesen Erwartungen. Dauer, Räume und Hilfsmittel stellen Beispiele für die Rahmenbedingungen eines solchen Assessments dar.

Durch den Abgleich mit den Erwartungen ergibt sich die meist aus Einzelbewertungen, Punkten, Prozentzahlen und sonstigen Beobachtungen abgeleitete Bewertung der einzelnen Person. Diese Bewertung ist das Maß für deren Performanz im konkreten Assessment.

Kritisch können verschiedene Aspekte betrachtet werden.

- Die Anforderungen sind zumeist kaum transparent dokumentiert. Ihre Herleitung bleibt der Erfahrung überlassen. Damit ist die Validität des Assessments ebenfalls dieser Erfahrung überlassen.
- Die Antworten der Lernenden im Assessment stellen zunächst eine Beobachtung dar. Diese erlaubt gegebenenfalls Rückschlüsse auf deren Kompetenz. Eine Bewertung ist aber nicht transparent an diese Kompetenz gekoppelt.
- Die Bewertung geschieht mittels des Erwartungshorizontes. Inwieweit dieser konkret genug gestaltet und die Bewertungsbedingungen immer einheitlich sind, wäre zumindest zu hinterfragen. Damit ist die Objektivität des Assessments im Einzelfall zu prüfen.

Im Rahmen dieser Arbeit soll der Begriff der Performanz daher wie oben geschildert auf das konkrete Ergebnis eines Assessments beschränkt werden.

Definition 2: Performanz

Performanz stellt in einem bestimmten Assessment die Bewertung der in dem Assessment erfolgten Beobachtungen in Form der Antworten der Assessment-Teilnehmer dar. Diese Bewertung kann sich an Qualifikationen, an Kompetenzen oder an anderen Inhalten orientieren, muss aber keine unmittelbare und allein tragfähige Basis für Aussagen über diese darstellen.

3.3.3 Qualifikation

Der dritte zentrale Begriff in diesem Zusammenhang ist die Qualifikation. So hat der DQR (Deutscher Qualifikationsrahmen) abgeleitet vom EQR (Europäischer Qualifikationsrahmen) das Ziel, deutsche Qualifikationen in Europa vergleichbarer zu machen. (BMBF, DQR - Deutscher Qualifikationsrahmen 2011, 3). Am DQR sind die verschiedenen Niveaustufen der Bildungsabschlüsse wie Master, Bachelor, Meisterprüfung oder Ausbildungsabschlüsse orientiert. Sie sollen sich allerdings an Kompetenzen orientieren, die formal, aber auch nicht formal erworben

¹ Zur Verwendung von Aufgaben bei der Kompetenzbeschreibung siehe auch (Schott und Ghanbari 2008).

worden sein können (BMBF, DQR - Deutscher Qualifikationsrahmen 2011, 4). Damit sind sie eine Folge von (gemessenen) Kompetenzen und eng mit diesem Begriff verbunden.

So definiert der DQR Qualifikation wie folgt:

„Qualifikation bezeichnet das formale Ergebnis eines Beurteilungs- und Validierungsprozesses, bei dem eine dafür zuständige Institution festgestellt hat, dass die individuellen Lernergebnisse vorgegebenen Standards entsprechen.“ (BMBF, DQR - Deutscher Qualifikationsrahmen 2011, 9)

Wesentlich ist, dass es sich um ein formales Ergebnis handelt, dass von einer Institution entsprechend vorgegebener Standards festgestellt wird. In der pädagogischen Diskussion sind verschiedene Relationen der Begriffe Qualifikation und Kompetenz erfolgt. Sie werden teilweise im Rahmen der Fokussierung auf Kompetenzen als Gegenbegriffe gesehen, wobei auch hier eine eher subjektorientierte von einer institutionellen Sichtweise abgegrenzt wird.

Wie bei der Gegenüberstellung der Kompetenz und der Performanz findet man auch in diesem Zusammenhang die Kompetenz dem Subjekt, also dem Individuum zugeordnet. [Darstellung beispielsweise bei (Kaufhold 2006, 50f.)]

Die Qualifikation wird auf die Umwelanforderungen, aber nicht auf eine bestimmte Situation bezogen. Sie besteht aus allgemeinen Kenntnissen, Fähigkeiten und Wissen, die wie die Kompetenz - aber nicht die Performanz - offensichtlich in vielfältigen Situationen zum Einsatz kommen können. Dies ist in **Tabelle 1** wiedergegeben.

Kompetenz	Qualifikation
subjektbezogen	anforderungsbezogen
auf Person und gesellschaftliche Werte bezogen	auf generalisierbare tätigkeitsbezogene Kenntnisse, Fertigkeiten und Fähigkeiten bezogen
selbstorganisierter Lernprozess	fremdorganisierter, curriculum-bezogener Lernprozess
Vielfalt der individuellen Handlungsdispositionen	nur zertifizierte Elemente individueller Fähigkeiten
auf die Handlung, die Situation und den Sinn bezogen	auf Tätigkeitsanforderungen, Wissen und Fähigkeiten bezogen

Tabelle 1 Gegenüberstellung von Kompetenz und Qualifikation (Kaufhold 2006, 51)

In einer Gegenüberstellung wie in **Tabelle 1** sieht man, dass die Kompetenz auf die ganze Person, auf das Subjekt, hier handlungsbezogen ausformuliert, bezogen ist. Die Qualifikation bezieht sich auf generalisierte tätigkeitsbezogene Kenntnisse, Fertigkeiten und Tätigkeitsanforderungen in bestimmten Umfeldern, die extern zertifiziert werden können. Hier stehen beispielsweise berufliche Anforderungsprofile im Vordergrund.

Um auch die Abgrenzung der Qualifikation von der Performanz zu erlauben, könnte man die Definition des DQR ergänzen und die Definition der Qualifikation etwa wie folgt um einen Bezug auf allgemeine, noch unbekannte Situationen erweitern.

Definition 3: Qualifikation

Qualifikationen bezeichnen das formale Ergebnis eines Beurteilungs- und Validierungsprozesses in Form von gemessenen Kompetenz- und Fähigkeitsbündeln, bei dem eine dafür zuständige Institution festgestellt hat, dass die individuellen Lernergebnisse vorgegebenen Standards entsprechen und auf noch unbekannte, verallgemeinerte berufliche Situationen übertragbar sind.

3.3.4 Übersicht

Somit ergibt sich folgender Versuch einer Einordnung. Die pädagogische Sicht von Kompetenzen bezieht sich auf den Bildungsprozess. Sie entstammt tendenziell der Vorstellung eines (selbstorganisierten) individuellen Lernprozesses, womit letztlich eine Bildung der Persönlichkeit oder des Individuums auf Basis eigener Motivationen und Interessen gemeint ist. Qualifikation hingegen bezieht sich auf das Erlernen von Sachverhalten, Fertigkeiten und Fähigkeiten wie sie durch einen (fremdorganisierten) oft beruflich orientierten Lernprozess entstehen. Während Kompetenz sich auf die Menge der individuellen Handlungsdispositionen in einer Handlung oder Situation bezieht, soll Qualifikation zertifizierte individuelle Fähigkeiten beschreiben.

Dies sind zwei verschiedene Sichtweisen, die beide Teile der jeweils anderen Sicht beinhalten (Abbildung 5). Geht man von den festgelegten, zertifizierten Abschlüssen und Qualifikationsniveaus aus, so beinhalten diese formal und nicht-formal erworbene Kompetenzen aber auch nicht erlernbare Qualifikationen wie körperliche Dispositionen (beispielsweise die Fähigkeit des Farbsehens, Unterscheiden von Geruchsnuancen, Körpergröße und insbesondere diverse genetisch bedingte Eigenschaften). Geht man umgekehrt vom individuellen Lernprozess aus, so entsteht ein Anteil messbarer und zertifizierbarer Qualifikationen und ein anderer, unbestimmter Anteil individueller nicht beruflich orientierter Kompetenzen, der nicht Bestandteil einer Qualifikation wird.

Inhaltlich werden Qualifikationen auch als „Wissens- und Fähigkeitsbündel“ bezeichnet, während die Kompetenz als der zielgerichtete Einsatz dieser Qualifikationen in entsprechenden Situationen angesehen wird (Gnahs 2007, 21) wobei dies um das Verstehen und die Folgenabschätzung in einer Situation ergänzt werden kann (Kaufhold 2006, 54). Die Betonung der Nutzung in Situationen ist also bei den Kompetenzen stärker, wird aber bei den Qualifikationen implizit gefordert, da deren Nachweis gerade über die Kompetenzen also den Einsatz in Situationen erfolgt.

Es stellt sich außerdem die Frage, ob die Kompetenz des zielgerichteten Einsatzes dieser Wissens- und Fähigkeitsbündel, also erworbener Kompetenzen, selbst Teil dieser Kompetenzen ist oder eine weitere übergeordnete Kompetenz darstellt. Auch die Frage, ob nicht der zielgerichtete Einsatz von Kompetenzen selbst eine zertifizierbare Qualifikation ist, ist nicht unberechtigt.

Für die begriffliche Klarheit scheint es im vorliegenden Zusammenhang sinnvoll, bei der Abgrenzung beider Begriffe auch auf die Erwartungshaltungen an das System Schule wie auch an andere Bildungssysteme zurückzukommen. Die Schulgesetze streben eine umfassende Kompetenzbildung an, die das Individuum befähigen soll, die Situationen seiner Lebensumgebung einschließlich der damit verbundenen Wertmaßstäbe zu bewältigen (Abbildung 5 Bezug a). Dies beinhaltet eine breite Vielfalt von möglichen Lebenssituationen sowohl im privaten wie auch im beruflichen Umfeld. Diese unbekanntes, generalisierten Situationen werden im Folgenden unter dem Begriff der „Anforderungswelt“ zusammengefasst.

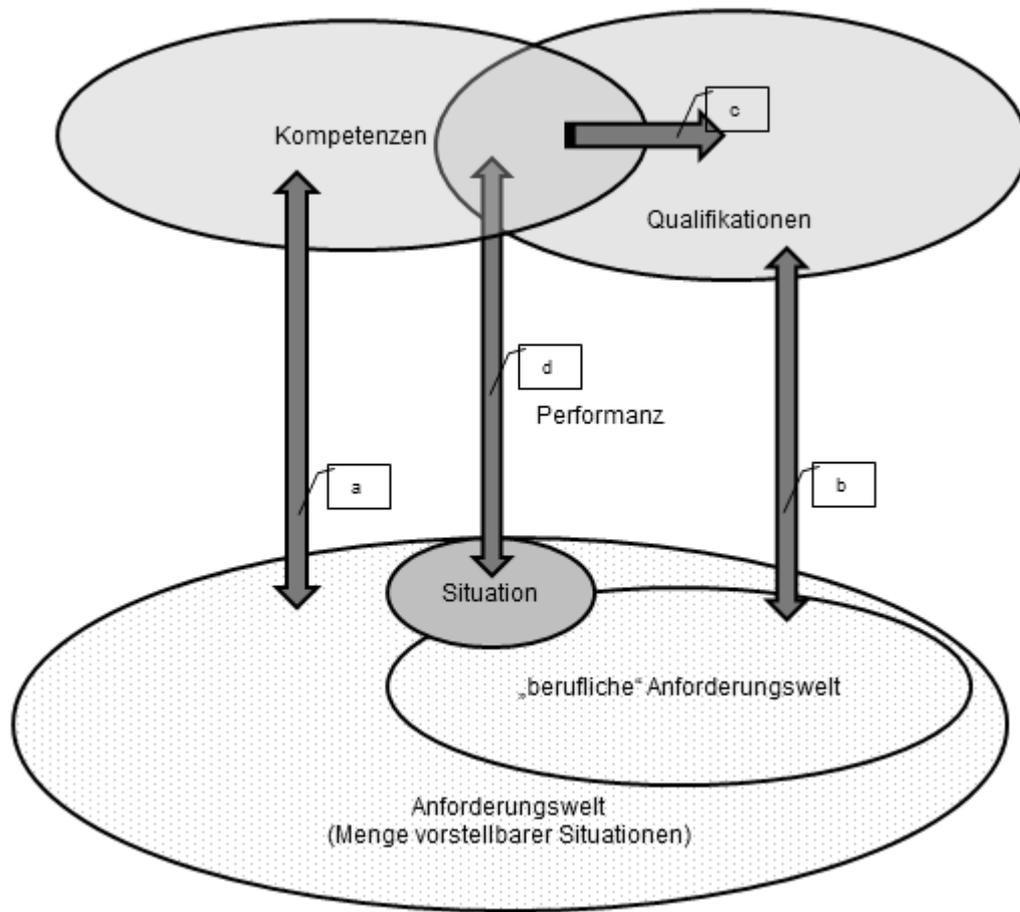


Abbildung 5 Kompetenz - Performanz - Qualifikation

Daneben gibt es die bereits angesprochenen „zubringerfunktions“-orientierten Anspruchsgruppen, die die im weiteren Sinne „beruflichen“ Situationen (einschließlich Lernsituationen in weiterführenden Bildungseinrichtungen) im Blickfeld haben. Bei ihnen stehen die zitierten „Wissens- und Fähigkeitsbündel“ im Vordergrund wie auch die oben erwähnten „objektiv und rechtsfähig zertifizierbaren Fähigkeiten“. Diese Anspruchsgruppen definieren die Qualifikationen, die in beruflichen Situationen erwartet werden. Dies geschieht zumeist zunächst abstrakt, da die tatsächlich beispielsweise im Studium oder in der beruflichen Tätigkeit auftretenden Situationen nur grob und/oder exemplarisch antizipiert werden können. (Bezug b). Allerdings wird der Begriff der Qualifikation aus Sicht der Zubringerfunktion gesehen und die Anforderungssituationen werden auf die „beruflichen“ Situationen reduziert.

Bestimmte Kompetenzen sind beispielsweise in einer beruflichen Welt unmittelbar umsetzbar. So ist eine Berufsausübung ohne Grundrechenarten in vielen Bereichen schwierig, viele Handwerksberufe sind ohne geometrische Grundkenntnisse nicht auszuüben und viele Studiengänge setzen Grundlagen der Differenzialrechnung voraus. Es werden also auch Kompetenzen erworben, die dann bei der Zertifizierung von Qualifikationen beispielsweise in Form des Abiturs nicht mehr ausdrücklich erwähnt werden.

So können sich aus den verschiedenen Perspektiven von Anspruchsgruppen Schwerpunkte in der Zielsetzung eher bei den Kompetenzen oder bei den Qualifikationen ergeben.

In Abbildung 5 ist diese Situation visualisiert. Kompetenzen müssen keine Qualifikationen sein. Sie können aber zu Qualifikationen führen, wenn sie zertifiziert werden. Kompetenzen können immer zu Qualifikationen führen und sind deren Voraussetzung (Bezug c). So kann in der Berufsausbildung die Beherrschung der Prozentrechnung zu einer Qualifikation im Bereich der Kosten-/Leistungsrechnung führen oder in einem Studium die „höhere Mathematik“ der gymnasialen Oberstufe zum Erreichen eines Bachelor- oder Masterabschlusses beitragen. Umgekehrt können Qualifikationen auch Bestandteile haben, die nicht durch Kompetenzen erzeugt worden sind, insbesondere genetisch bedingte, dem Individuum unveränderbar anhaftende Dispositionen.

Die Performanz tritt in einer konkreten Situation auf, die sich auf die berufliche Umwelt, aber auch auf weitere Aspekte der Anforderungswelt beziehen kann. Durch Verallgemeinerung der Ergebnisse dient sie der Messung sowohl von Kompetenzen als auch Qualifikationen (Bezug d). Letztlich misst ein Assessment somit zunächst immer die Performanz.

Aus der Performanz in einer Situation muss versucht werden, sowohl auf die Kompetenzen als auch die Qualifikationen zu verallgemeinern, also auf Aussagen über deren Vorhandensein und Nutzbarkeit in vielen noch unbekanntem Situationen der Anforderungsumwelt.

3.4 Der Kompetenzbegriff

Der Kompetenzbegriff, der bisher nur in der Definition von Weinert vorgestellt wurde, soll jetzt differenziert werden. Wie bereits erwähnt, wurde der Leitbegriff Kompetenz über die Output-Orientierung in der beruflichen Bildung eingeführt (BMBF 2009, 9). Inhaltlich ist das Konzept der Kompetenzorientierung aus der Diskussion um Bildungskonzepte und damit letztlich Bildungsziele entstanden. Sie sollen das „tote“ Wissen durch aktives auf konkrete Situationen übertragbares Wissen und Handlungsfähigkeit ersetzen.

Kompetenzen als aktuelles Bildungskonzept sollen zum lebenslangen selbstständigen Lernen und Anpassen an die beruflichen Herausforderungen dienen (Winther 2010, 6). In dem Satz „Die Modellierung und Messung von Kompetenzen spielt eine entscheidende Rolle bei der Optimierung von Bildungsprozessen sowie der Qualitätssicherung und Qualitätsentwicklung des Bildungswesens.“ (Leutner, Klieme, et al. 2013, 2) wird deren herausragende Bedeutung sichtbar.

3.4.1 Historie

3.4.1.1 Sozialwissenschaftliche Kompetenz

Die Sozialwissenschaften können den Kompetenzbegriff auf verschiedene Entwicklungen zurückführen, so auf die Linguistik Noam Chomskys und die amerikanische funktional-pragmatische Psychologie der späten 1950-er-Jahre (Klieme und Hartig 2007, 14) wie beispielsweise von White (White 1959).

Chomsky (Chomsky 1965) war mit seinem Begriff der „linguistic competence“ weniger an individuellen Messungen als vielmehr an dem Verständnis des generellen zugrundeliegenden Regelsystems interessiert. Es ging also um das „generative, situations-unabhängige kognitive System“ (Klieme und Hartig 2007, 16), dem der Begriff der Anwendung der Sprache in der konkreten Situation als Performanz gegenübergestellt wurde. Demgegenüber interessierte die funktional-pragmatische Psychologie gerade die Fähigkeit einer Person bestimmte situations-abhängige Anforderungen zu meistern.

Generell geht es auch bei sozialwissenschaftlichen Kompetenzen um Anforderungen, die sich auf reale aber variable Situationen beziehen in denen eine Person durch beobachtbare Handlungen vorhandene Dispositionen realisiert.

3.4.1.2 Pädagogische Kompetenz

Verschiedene Fachrichtungen haben ausgehend von diesen grundsätzlichen Überlegungen ihren eigenen, spezialisierten Kompetenzbegriff geprägt. Die Pädagogik hat ihn als einen Terminus zur Lösung der alten Frage nach dem „worthwhile knowledge“ als Ziel der Bildungsprozesse gewählt. Csapo verweist darauf und stellt dies in den Kontext der verschiedenen Blickwinkel oder „Dimensionen“, fachlich, psychologisch und sozial, zwischen denen der Schwerpunkt der Pädagogik sich immer wieder verschiebt. (Csapo 2010, 12)

Der pädagogische Begriff der Kompetenz wird intensiv im Bereich der Erwachsenenbildung und der beruflichen Bildung verwendet. Der Begriff der „beruflichen Handlungskompetenz“ ist eine zentrale Grundlage der Berufspädagogik. (Klieme und Hartig 2007, 22)

Dabei kann an ältere Diskussionen angeknüpft werden (Kaufhold 2006, 61). Erstmals systematisch wird der Begriff in der Pädagogik von Roth (Roth 1971, 180) erwähnt, aufbauend auf dem sozialwissenschaftlichen Begriff (Klieme und Hartig 2007, 19) definiert Roth die drei Begriffe:

- Selbstkompetenz
- Sachkompetenz
- Sozialkompetenz.

Somit gibt die Gliederung von Roth bereits drei Dimensionen der Kompetenz vor. Es hat sich die Tradition entwickelt derartige Gliederungen von Kompetenz in Dimensionen vorzunehmen und es sind in der Folge eine Reihe weiterer Vorschläge entwickelt worden.

Für den Bereich der beruflichen Aus- und Weiterbildung hat es zumindest zeitweilig einen gewissen Konsens dahingehend gegeben, dass zumindest drei Dimensionen zu berücksichtigen sind (Achtenhagen und Baethge 2007, 57) (Gnahn 2007, 26) (Winther 2010, 49). Die Sachkompetenz nach Roth wird dabei in Fach- und Methodenkompetenz aufgespalten. Nach Achtenhagen und Baethge ergibt sich als Struktur (Achtenhagen und Baethge 2007, 57):

- „Eine Dimension muss auf theoretisch/analytische Anforderungen bezogen sein und dabei das Wissen [...] berücksichtigen: ‚Sachkompetenz‘.“
- „Eine zweite Dimension muss auf technische/praktische/funktionelle Anforderungen bezogen sein; [...] ‚Methodenkompetenz‘.“
- „Eine dritte Dimension muss auf zwischenmenschliche Anforderungen bezogen sein: [...] ‚Sozialkompetenz‘.“

Hinzu kommt die Selbstkompetenz. Daneben werden allgemeine Fähigkeiten gestellt, als inhaltliches Basiswissen bezeichnet (Edelmann und Tippelt 2007, 133) und unterteilt in:

- naturwissenschaftliches,
- sozialwissenschaftliches,
- ethisches Basiswissen.

Auch weitere Untergliederungen der Wissensdimension erfolgen [beispielsweise nach (Achtenhagen und Baethge 2007, 57)] in:

- deklaratives Wissen (vernetzt aufzubauen)
- prozedurales Wissen
- strategisches Wissen

Die Kultusministerkonferenz unterscheidet 1999 zwischen Fach-, Personal- und Sozialkompetenz und beschreibt übergreifend die Handlungskompetenz „[...] als die Bereitschaft und die Fähigkeit des einzelnen, sich in beruflichen, gesellschaftlichen und privaten Situationen sachgerecht durchdacht sowie individuell und sozial verantwortlich zu verhalten“ (Kultusministerkonferenz 1999, 9).

Es existiert also eine Reihe von Vorschlägen, wie sich Kompetenz strukturieren lässt. An dieser Stelle soll nicht weiter auf die Unterschiede oder die „richtige“ Gliederung eingegangen werden, vielmehr wird festgestellt, dass verschiedene Möglichkeiten der Strukturierung in Dimensionen und verschiedene Granularitätsstufen von Kompetenz je nach Sichtweise relevant sein können.

3.4.2 Nicht kognitive Fähigkeiten

Ausgehend von den verschiedenen Kompetenzdimensionen werden bestimmte wie die kognitiven Dimensionen oft als relativ leichter zugänglich als andere Dimensionen gesehen. Die Definition von Kompetenz nach Weinert wurde bereits dargestellt:

Kompetenzen sind die „[...] bei Individuen verfügbaren und durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ (Weinert 2001, 27).

Weinert spricht dabei zunächst von kognitiven Fähigkeiten und Fertigkeiten und ergänzt diese durch andere „Bereitschaften und Fähigkeiten“. Andere Autoren, beispielsweise Gnahs, stellen andere Aspekte gleichberechtigt daneben (Tabelle 2).

Komponente	Beschreibung
Fähigkeiten („Wissen“)	Kognitive Fähigkeiten, Regeln und Fakten, prozedurales und deklaratorisches Wissen
Fertigkeiten	auch sensomotorische Fähigkeiten und Nutzung aller Sinne, handwerkliche Fähigkeiten, physisches Sprachvermögen, Fähigkeiten zu singen oder tanzen
Dispositionen	Persönlichkeitseigenschaften, die "Big Five" wie beispielsweise Verträglichkeit oder Gewissenhaftigkeit
Werte	beispielsweise religiöse, politische, kulturelle, familiäre Haltungen und Einstellungen
Motivationen	emotionale Antriebskräfte und Interessen, Abwägungen von Aufwands- und Erfolgserwartungen

Tabelle 2 Kompetenzkomponenten nach Gnahs (Gnahs 2007, 23)

Viele Aspekte dieser Übersicht werden auch von der Definition von Weinert berührt. Offen bleibt an dieser Stelle, ob nicht eine ausdrückliche Ergänzung durch die sensomotorischen Fähigkeiten,

die Sinneswahrnehmung, durch praktische Fähigkeiten des Handwerks, durch physisches Sprachvermögen, Bewegungsfähigkeit, sportliche Fähigkeiten, Singen oder Tanzen zu erwägen wäre, wie sie nach Gnahs in Tabelle 2 aufgeführt sind.

Zumindest einige derartiger Fähigkeiten scheinen gerade auch in bestimmten, beruflichen Umfeldern eine wichtige Rolle spielen zu können. Möglich erscheint zumindest eine getrennte Hervorhebung neben dem kognitiven (deklarativen und prozeduralen) Wissen.

Eine wichtige Rolle spielt zudem der Begriff der Dispositionen als stabile Persönlichkeitseigenschaften, wie sie beispielsweise die „Big Five“ der Psychologie, also Neurotizismus, Extraversion, Offenheit für neue Erfahrungen, Gewissenhaftigkeit und Verträglichkeit darstellen. Sie sind bereits in Tabelle 2 erwähnt. Dabei stellt sich die Frage, welche Eigenschaften eines Individuums quasi genetisch festgelegt und unveränderbar sind und welche durch Bildungsprozesse verändert werden können. In der Wikipedia findet man für Persönlichkeitseigenschaften im weitesten Sinne:

„Persönlichkeitseigenschaften in einem weiten Sinn umfassen alle psychologisch fassbaren individuellen Differenzen des Verhaltens und Befindens sowie ihre biologischen Grundlagen in der psychophysischen Individualität (Konstitution) des Menschen.“ (Wikipedia - Persönlichkeitseigenschaft 2015)

Diese Definition schließt nicht aus, dass es Eigenschaften geben könnte, die keine Kompetenzen sind, weil sie nicht durch Bildungsprozesse beeinflussbar sind. Dies wurde bereits in der Abgrenzung zu Qualifikationen angesprochen. So wird bereits im Bereich körperlicher Dispositionen klar, dass Bildung nicht alle Aspekte eines Individuums erreichen kann. Darüber hinaus ist auch die Frage zu stellen, inwieweit bestimmte Bereiche überhaupt eine Zielvorstellung von Bildung sein sollten. Die Identität eines Individuums ist so wertvoll und gegeben, dass sie nicht Gegenstand des Erziehungsprozesses sein kann.

Verwendet man Disposition im Sinne der biologischen geistigen und körperlichen Basiseigenschaften eines Individuums, so sind diese die Grundlage eines Bildungsprozesses aber von diesem nicht veränderbar. Gleichwohl gehen sie unter Umständen in die Kompetenzmessung mit ein, da sie als Eigenschaft des Individuums nicht von diesem zu trennen sind. Sie sind unveränderlich mit dem Individuum verknüpft und beeinflussen dessen Leistungsfähigkeit. Die Frage ist also, wie diese Dispositionen im Sinne fester Persönlichkeitseigenschaften bei einer Kompetenzmessung erfasst werden.

Der Begriff Disposition wird aber auch anders verwendet. Andere Darstellungen sehen Dispositionen als eine im Bildungsprozess erworbene Eigenschaft an. Klieme und Hartig schreiben anknüpfend an Roth „Kompetenzen sind Dispositionen, die im Verlauf von Bildungs- und Erziehungsprozessen erworben(erlernt) werden [...]“ (Klieme und Hartig 2007, 21) In Übereinstimmung mit der Idee des Erwerbs werden auch durch die Bildungsstandards Kompetenzen als erworbene Dispositionen bezeichnet:

„Kompetenzen beschreiben Dispositionen zur Bewältigung bestimmter Anforderungen. Solche Kompetenzen sind fach-, beziehungsweise lernbereichsspezifisch ausformuliert, da sie an bestimmten Inhalten erworben werden müssen.“ (Bildungsstandards der Kultusministerkonferenz 2005, 16).

Es gibt also eine mehrdeutige Verwendung des Begriffs Disposition. Dispositionen können sowohl invariable als auch im Bildungsprozess erworbene Fähigkeiten und Fertigkeiten umfassen. Allerdings sind wohl auch die erworbenen Dispositionen als langfristig angelegt und zeitlich stabil zu verstehen.

Folgt man dem wie geschildert verbreiteten Gedanken Disposition als Kompetenz, die eine Folge eines Bildungsprozesses ist, aufzufassen, so ist im Grunde eine Verwendung des Begriffes nicht zwingend. Vielmehr geht es dann um eine Ausweitung des Kompetenzbegriffes und es kann auch weiter von Kompetenzmessung bei der Messung derartiger „Dispositionen“ gesprochen werden.

Damit stellt sich die Frage, wie Dispositionen im Sinne stabiler Persönlichkeitseigenschaften, die nicht im Bildungsprozess erworben wurden bei der Kompetenzmessung berücksichtigt werden. Diese Dispositionen werden trotzdem zum Ergebnis einer Leistungsmessung beitragen und gerade unter dem Qualifikationsgedanken erscheint es sogar widersprüchlich diese ausschließen zu wollen, da auch diese Dispositionen Teil einer Qualifikation sind. Es erscheint aber sinnvoll, Dispositionen, die nicht beeinflussbar sind, zumindest nicht gezielt im Rahmen einer Kompetenzmessung messen zu wollen.

Damit ergibt sich die Konsequenz, dass nicht alle bei der Feststellung einer Qualifikation mitgemessenen Aspekte tatsächlich Kompetenzen sind. Zumindest diese feststehenden Dispositionen als Teil des Individuums können Qualifikationen sein, die nicht kompetenzbedingt sind. Somit bestätigt sich die Trennung der Begriffe Qualifikation und Kompetenz. Für die Kompetenz ergibt sich, dass nur durch den Bildungsprozess veränderbare Dispositionen als Kompetenzen aufzufassen sind.

Nun wird es aber auch solche unveränderlichen Dispositionen geben, die nicht in einem Qualifikationsprozess genutzt werden und somit weder Kompetenzen sind noch zu einer bestimmten Qualifikation beitragen. Das bedeutet im Ergebnis, dass ein Individuum über Kompetenzen und Qualifikationen verfügt, aber nicht vollständig durch sie definiert wird. Kompetenzen und Qualifikationen ihrerseits können einander überlagern, haben aber auch jeweils unabhängige Bereiche. Insgesamt gelangt man so zur Darstellung in Abbildung 6.

Kommt man noch einmal auf die Tabelle 2 zurück, so muss man auch die Rolle der Werte in diesem Zusammenhang betrachten. Hier scheint ein wesentlicher Aspekt zu sein, zu klären, ob neben den durch den Bildungsprozess disponiblen Werten wie sie als Komponente von Kompetenzen definiert sind, auch von Werten ausgegangen werden muss, die durch die individuelle Prägung des Individuums beziehungsweise durch andere Faktoren wie das Elternhaus, die Gesellschaft oder religiöse Überzeugungen festgelegt werden. Diese sind im Bereich des Individuums dann ebenfalls außerhalb der Kompetenzen anzusehen und entziehen sich einer Bewertung im Rahmen des Bildungsprozesses jenseits ihrer Verträglichkeit mit dem System als solchem.

Auch diese Überlegungen legen nahe, zu definieren, dass der Kompetenzbegriff nicht das gesamte Individuum erfassen kann.

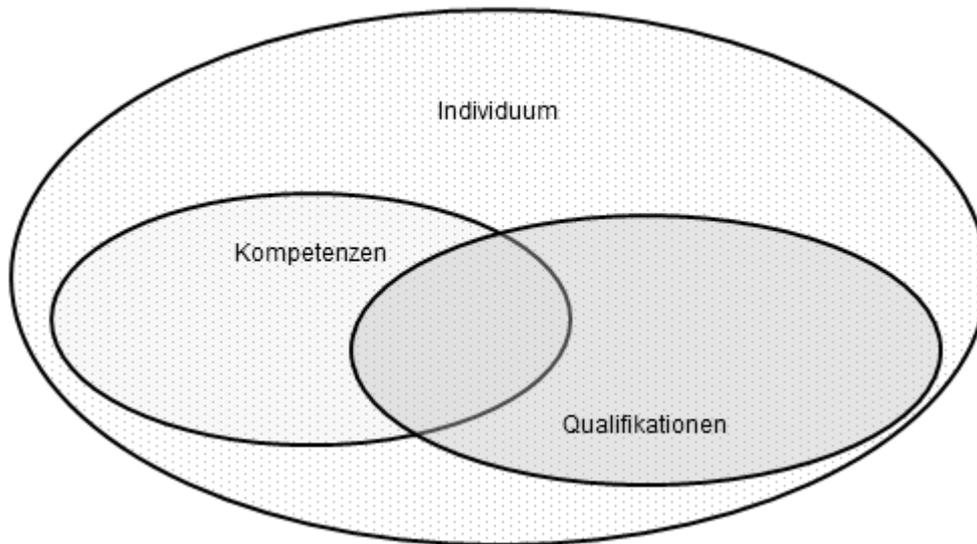


Abbildung 6 Kompetenzen und Qualifikationen als Teil des Individuums

Eher anders verhält es sich mit dem Bereich der Motivation. Dieser ist zwar ebenfalls schwer zu messen, kann aber in der Kompetenzdefinition berücksichtigt werden und wird auch direkt von Weinert in die Definition einbezogen.

3.4.3 Individualkompetenz

Kompetenz im betrachteten Sinn ist also erlernbar, ist außer mit kognitiven auch mit anderen, insbesondere motivationalen, volitionalen aber auch sensomotorischen Fähigkeiten und Fertigkeiten verbunden und kann Eigenschaften wie sie in den „Big Five“ beschrieben sind, oder Werte umfassen soweit diese erlernbar und Teil der bildungspolitischen Ziele sind. Sie ist kontextabhängig und zielgerichtet auf Problemlösungen.

Daher werden Kompetenzen hier in Anlehnung an Weinert etwas modifiziert definiert und in Abgrenzung zur in Abschnitt 3.4.5 definierten „Kompetenz im engeren Sinne“ als dem Individuum als breitestes Spektrum zugeordnete Individualkompetenz bezeichnet.

Definition 4: Individualkompetenz

Individualkompetenzen sind die von Individuen in Bildungsprozessen erworbene oder modifizierte kognitive Fähigkeiten und Fertigkeiten in deklarativer, prozeduraler und anderer Form, sensomotorische Fertigkeiten, damit verbundene motivationale, volitionale, soziale und ethische Bereitschaften und individuelle Eigenschaften, um Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll im Sinne der Bildungsziele erstellen zu können.

3.4.4 Handlungskompetenz

Wenn, wie oben beschrieben, Bildungsziele durch Kompetenzen operationalisiert und stets durch Assessments überprüfbar sein sollen, folgt daraus, dass nur die Kompetenzen eines Lernenden, die durch einen Bildungsprozess gestaltbar sind **und deren Auswirkungen im Handeln beobachtbar sind**, überprüfbare Bildungsziele darstellen können. Die Frage ist also, ob, in welchem Umfang und wie Individualkompetenz messbar und damit überprüfbar ist.

Hierbei geht es nicht nur um den Aspekt des Beobachtens, sondern auch um die nutzbringende, zielgerichtete Anwendung der Kompetenzen beim Handeln in entsprechenden Situationen. Der Begriff der Handlung bestimmt also den Kompetenzbegriff zweifach, einerseits

- als wünschenswerte Umsetzung vorhandener Fertigkeiten und Fähigkeiten in aktives Handeln und andererseits
- als Voraussetzung für die Beobachtung des Vorhandenseins von Kompetenzen.

Dementsprechend ist der Begriff der Handlungskompetenz geprägt worden, wobei verschiedene Aspekte aufgezählt werden (Kaufhold 2006, 62):

1. Kompetenz hat einen „unmittelbaren Handlungsbezug“. Sie kann nur über Handlungen beobachtet werden. Die Handlung umfasst „sowohl situative als auch individuelle Aspekte“. Beide sind bei der Kompetenzerfassung zu beachten.
2. Die Handlungssituation bestimmt die notwendigen (auftretenden und damit messbaren) Kompetenzdimensionen. Sie ist kontextabhängig. (Klieme und Hartig 2007, 17)
3. Die Kompetenzentwicklung ist auch durch „situative Handlungsanforderungen“ bestimmt.
4. Die Kompetenzerfassung kann durch „Lernerfolgskontrollen oder Evaluation“ erfolgen.

Die Forderung, dass Kompetenzen in ihrer Auswirkung im Handeln beobachtbar sein müssen, bietet erst die reale Chance, sie messen zu können. Ohne ein sichtbares, auf eine bestimmte funktional wirksame Kompetenz zurückführbares Ergebnis ist das Vorhandensein einer Kompetenz schwer nachweisbar. Diese Forderung nach einem möglichst eindeutigen Ursache-Wirkungszusammenhang ist pragmatisch nachvollziehbar. Andererseits stellen die damit verbundenen Anforderungen eine Restriktion der Kompetenzinhalte dar. Einfach ausgedrückt heißt dies letztlich, dass nur eine Kompetenz ist, was in seiner funktionalen Wirkung zumindest indirekt messbar ist.

3.4.5 Kompetenz im engeren Sinne

Mit dem Konzept der Handlungskompetenz erfolgt ein Schritt zur Messbarkeit allerdings unter Ausschluss gerade der nicht messbaren Kompetenzen. Messen bedeutet in diesem Zusammenhang in einer Situation beobachten, die Performanz dokumentieren und mit einem Zahlenwert bewerten zu können.

Generell wird bei vielen Kompetenzdefinitionen oft die Messbarkeit vieler Aspekte angezweifelt. Bereiche wie Werte oder Motivationen scheinen sich einer Messbarkeit stark zu widersetzen, da sie eher intrapersonell und schlecht beobachtbar sind. Versuche dies zu überwinden scheinen in ihrer Objektivität stark anzweifelbar (siehe Abschnitt 4.5.1).

Allerdings werden auch Fähigkeiten oder Fertigkeiten im Handeln ohne beispielsweise eine entsprechende Motivation kaum zur Entfaltung kommen. Implizit sind bei jeder Beobachtung immer mehrere, wenn nicht alle Komponenten der Individualkompetenz enthalten.

Gerade das Beispiel der Motivation zeigt diese Kombination. Eine sinnvolle Kompetenzmessung erfordert ein leistungsorientiertes Verhalten. Cronbach unterscheidet zwischen typischem Verhalten und maximaler Anstrengung (Cronbach 1949). Mit leistungsorientiertem Verhalten ist in diesem Sinne ein Verhalten mit maximaler Anstrengung gemeint. Situationen maximaler Anstrengung lassen sich dadurch charakterisieren, dass der Lernende sich in einer Situation befindet, die es ihm

ermöglicht seine bestmögliche Leistung zu erbringen, dass er dies möchte und ihm bewusst ist, dass es sich um eine Prüfungssituation handelt. Die Prüfungssituation muss außerdem für die Messung und Bewertung geeignet sein. (Wilhelm und Nickolaus 2013, 24)

Dies zeigt, dass durch geeignete Rahmenbedingungen bei der Kompetenzmessung Einfluss auf deren Ergebnisse genommen werden kann, indem andere als kognitive Aspekte verändert werden. Hier entsteht ein Konflikt zwischen der Validität im Sinne dessen was tatsächlich gemessen wird einerseits und der Komplexität des Messbaren andererseits. Werden die Anforderungen an die Rahmenbedingungen so hoch gestellt, dass der Einfluss anderer Kompetenzdimensionen auf die zu messende Kompetenzeigenschaft minimiert wird, muss deren Umfang so stark reduziert werden, dass die Situation kaum als realistisch oder komplex eingestuft werden kann. Werden andererseits möglichst reale und komplexe Situationen gewählt, sind stets weitere Kompetenzdimensionen mit ihrem Einfluss beteiligt.

Das ist kein spezielles Problem der Kompetenzmessung. Jede Leistungsbewertung ist in der Praxis Problemen und Fehlern unterworfen (Saldern 2011, 93ff.) Diese sind zudem nur qualitativ bekannt. Die parallele Messung verschiedener Leistungsdimensionen lässt sich nicht vollständig vermeiden. Sie lässt sich aber verringern indem Rahmenbedingungen geschaffen werden, die so kontrolliert wie möglich gestaltet werden, indem Verfahren zur Gestaltung der Leistungsmessungen genutzt werden, die so transparent und valide wie möglich sind und indem eine Beobachtung stattfindet, die so objektiv wie möglich ist.

Diese Anforderungen lassen sich für verschiedene Dimensionen der Individualkompetenz in unterschiedlich ausgeprägter Form gestalten. Unter dem Blickwinkel der Zielvorstellungen

- domänenorientierte, also inhaltsorientierte Kompetenzen,
- Objektivität der Formulierbarkeit und Operationalisierbarkeit und
- Zubringerfunktion für berufliche Qualifikationen

sind die kognitiven Kompetenzaspekte diejenigen, die sich zunächst am besten modellieren lassen. Daher soll hier die ebenfalls gängige Definition von Klieme und Leutner im Rahmen des DFG-Projektes aufgegriffen werden:

Definition 5: Kompetenz (im engeren Sinne)

Pragmatisch funktional können Kompetenzen verstanden werden als „kontextspezifische, kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmen Domänen beziehen“ (Klieme und Leutner 2006, 879).

Damit werden die Kompetenzen im engeren Sinne (Kompetenz i.e.S.) auf kognitive Aspekte ebenso wie auf kontextspezifische Aspekte beschränkt, die in bestimmten Domänen wirksam werden und deren Wirksamkeit sich funktional auf sie zurückführen lässt. Mit dem Hinweis des funktionalen Bezugs wird auch der Messbarkeitsgedanke eingebracht. Dies entspricht der Idee des zurzeit pragmatisch Messbaren. Dem lässt sich entgegenhalten, dass sich die Definition von Bildungszielen nicht generell nur an Messbarkeitsaspekten orientieren darf. Das geschieht aber auch im Regelfall nicht. So wird sowohl von Klieme/Leutner als auch der KMK und den Autoren der Bildungsstandards darauf hingewiesen, dass die Bewertung dieser Kompetenzen keine abschließende Bewertung darstellt. In den Bildungsstandards wird beispielsweise ausdrücklich festgestellt: „Die vorgelegten Standards konzentrieren sich damit auf überprüfbare, fachbezogene

Kompetenzen und vermessen keineswegs das gesamte Spektrum von Bildung und Erziehung.“ (Bildungsstandards der Kultusministerkonferenz 2005, 16).

Im Rahmen dieser Arbeit wird dem insbesondere dadurch Rechnung getragen, dass die Messung und Bewertung nicht die abschließenden Schritte sind. Vielmehr wird durch die Bündelung der Bewertung verschiedener Leistungsmessungen in dem Punkt Leistungsbeurteilung ein abschließender Schritt eingeführt, der eine vollständige Beurteilung außerhalb der Kompetenzmessung unter Einschluss weiterer, insbesondere pädagogischer Faktoren beinhaltet.

Wenn im Folgenden von Kompetenzen gesprochen wird, ist die hier angegebene Kompetenzdefinition als Kompetenzen im engeren Sinne (i.e.S) gemeint. Ist der weitere Kompetenzbegriff gemeint, wird zur Klarheit von Individualkompetenz gesprochen.

Um Missverständnisse leichter vermeiden zu können, wird statt von Kompetenzen im engeren Sinne auch von Fähigkeiten gesprochen, wie es auch Schott und Ghanbari tun (Schott und Ghanbari 2008, 14).

Wohlwissend, dass in jedem Assessment immer mehrere Kompetenzkomponenten wirksam sind und somit gemessen werden, sollen gezielt die am besten messbaren, kognitiven Fähigkeiten operationalisiert und einer Analyse unterworfen werden. Dass dies auch in der schulischen Praxis häufig der Fall ist, stellt einerseits ein Argument für den Praxisbezug dar, soll aber andererseits nicht möglichen Erweiterungen entgegenstehen. Das hier entwickelte Modell erfordert abgesehen von den prinzipiellen Schwierigkeiten einer operationalisierten, konkreten Modellierung keine strukturellen Beschränkungen auf kognitive Fähigkeiten.

3.4.6 Kontextabhängigkeit

Im Rahmen der Kompetenzdiskussion wird auch der Begriff „Schlüsselkompetenz“ oder „Schlüsselqualifikation“ verwendet, der in Deutschland eine wichtige Rolle spielt. Dabei geht es um möglichst universelle, fächer- und inhaltsübergreifende Kompetenzen, wie beispielsweise „Lesen“ oder „Schreiben“.

„Schlüsselqualifikationen sind erwerbbar allgemeine Fähigkeiten, Einstellungen und Wissens-elemente, die bei der Lösung von Problemen und beim Erwerb neuer Kompetenzen in möglichst vielen Inhaltsbereichen von Nutzen sind, so dass eine Handlungsfähigkeit entsteht, die es ermöglicht sowohl individuellen als auch gesellschaftlichen Anforderungen gerecht zu werden.“ (Orth 1999, 107)

Durch die Förderung von Schlüsselkompetenzen soll eine möglichst allgemeine Befähigung erlangt werden, die die Übertragung auf neue inhaltliche Situationen in Form von Anforderungen und Problemstellungen erleichtern oder überhaupt erst ermöglichen soll. Kritisiert wird bei der Definition von Schlüsselkompetenzen häufig, dass Kompetenzen sich nur in einem inhaltlich definierten Umfeld ausdrücken können, was dem Universalitätsanspruch der Schlüsselkompetenzen zuwiderläuft (Winther 2010, 17f.). Ganz deutlich drückt dies unter anderem wiederum die Expertise zu den Bildungsstandards aus: „Aufgrund der zentralen Rolle fachbezogener Fähigkeiten und fachbezogenen Wissens sind Kompetenzen in hohem Maße domänenspezifisch.“ (BMBF, Zur Entwicklung nationaler Bildungsstandards 2003, 75).

Dort wird auch klar dargelegt, dass entgegen der oftmals vorhandenen Betonung fächerübergreifender Schlüsselkompetenzen, die Bedeutung der fachlichen Kompetenzen nicht vernachlässigt werden darf, da sie diese nicht ersetzen können. Es wird auch betont, dass der Stand der Forschung darauf hinweist, dass die fachlichen Kompetenzen die Grundlage für die Entwicklung fachübergreifender Kompetenzen bilden, sich also übergreifende Kompetenzen überhaupt erst auf der Basis fachlicher Kompetenzen bilden können. Daraus wird die Folgerung gezogen, „[...]“, dass konkrete Ausformulierungen und Operationalisierungen des Kompetenzbegriffs zunächst in den Domänen bzw. Fächern zu erfolgen haben. Daraus begründet sich weiterhin die Notwendigkeit, bei der Entwicklung von Kompetenzmodellen auf dem Theorie- und Erkenntnisstand der Fachdidaktiken aufzubauen.“ (BMBF, Zur Entwicklung nationaler Bildungsstandards 2003, 75).

Somit steht die Formulierung von inhaltsbezogenen, oder domänenbezogenen Kompetenzen am Anfang. Dies gilt auch für das im Rahmen der Arbeit entwickelte Modell. Andererseits schließt das eine spätere Ableitung der mit Schlüsselkompetenzen bezeichneten, verallgemeinerten Begrifflichkeiten nicht aus.

3.4.7 Domänen

Betont man den inhaltlich-fachlichen Bezug der Kompetenzen, ergibt sich die Frage wie diese Inhalte zu strukturieren sind und in welchen Granularitäten sie verpackt werden. Hier spielt der Begriff der „Domäne“ eine zentrale Rolle.

Er wird beispielsweise erwähnt in Form

- „[...] des hier vertretenen Kompetenzbegriffs auf Lernbereiche, Fächer bzw. ‚Domänen‘ [...]“ (BMBF, Zur Entwicklung nationaler Bildungsstandards 2003, 75),
- „[...] ihrerseits in bestimmten Lebensbereichen (Domänen) (z.B. öffentlich, privat, beruflich)“ (BMBF, Zur Entwicklung nationaler Bildungsstandards 2003, 149) im sprachlichen Umfeld
- „[...] in einem Gegenstandsbereich (einer ‚Domäne‘, wie Wissenspsychologen sagen, einem Lernbereich oder einem Fach) zu identifizieren...“ (BMBF, Zur Entwicklung nationaler Bildungsstandards 2003, 21f.).

Im Bereich der allgemeinbildenden Gymnasien gibt es beispielsweise die Gliederung in Profile (sprachlich-literaturwissenschaftlich, mathematisch-naturwissenschaftlich, historisch-sozial, ästhetisch-expressiv), die stark an der Fächerorientierung der gymnasialen Oberstufe angelehnt ist, während sich die berufliche Bildung an beruflichen Anforderungen und damit fachübergreifenden Kompetenzen orientieren soll. (Winther 2010, 82f.). Dafür treten im Bereich der beruflichen Bildung Lernfelder oder Lernbereiche an die Stelle der Fächer, die aber ebenfalls eine inhaltliche Gruppierungsfunktion haben.

Der Begriff der Domäne wird im Zusammenhang mit Fächern, Lernbereichen, Lernfeldern, Profilen und beruflichen Anforderungen verwendet. Er umfasst immer mehr als eine einzelne (Handlungs-)Situation, ist manchmal aber auch fach- oder lernfeldübergreifend, was auch mit den verschiedenen Kompetenzniveaus und inhaltlichen Schwerpunkten von Bildungsprozessen sowie den begrifflichen Überlappungen zu tun hat. Insgesamt scheint es sinnvoll und praktikabel zu sein, bei der Definition einer Domäne in der Granularität unterhalb eines Faches aber oberhalb einer Situation zu bleiben.

Achtenhagen und Winther schlagen daher eine näherungsweise Grenzziehung für die Granularität einer Domäne vor, die für diese Arbeit als praktikabel angesehen wird: (Achtenhagen und Winther 2008)

- Sie ist grober als Handeln in einer bestimmten Situation (aufstellen einer Exponentialfunktion zur Beschreibung des Wachstums)
- Sie ist grober als ein curricularer Abschnitt (Lösen von Exponentialgleichungen)
- Sie entspricht maximal dem Inhalt eines Ausbildungsberufes, eines Unterrichtsfaches oder eines Lernfeldes.

Mit dieser Grenze kann die Granularität der Domänen bestimmt werden. Die konkrete Ausgestaltung und Grenzziehung wie auch eine bewusste Überlappung müssen von der Fachdidaktik vorgenommen werden. Das Modell muss an dieser Stelle eine entsprechende Flexibilität aufweisen.

Im Rahmen dieser Arbeit wird bei der inhaltlichen Charakterisierung und Strukturierung einer Domäne vom Domäneninhalt gesprochen.

Definition 6: Domäneninhalt (Content)

Unter dem Domäneninhalt wird die inhaltliche, strukturelle und prozedurale Menge aller Aspekte verstanden, die im Rahmen der Domäne Kompetenzen im engeren Sinne, Qualifikationen und Fähigkeiten beschreiben.

Das hier entwickelte Kompetenzmodell soll den Domäneninhalt mit Zuordnung zu der jeweiligen Domäne beschreiben können. Daraus ergeben sich zwei weitere wesentliche Anforderungen an das Kompetenzmodell.

1. Das Kompetenzmodell muss die Abbildung verschiedener fachdidaktischer Inhalte erlauben, die Gliederung in Domänen und die Aufnahme neuer Domänen erlauben.
2. Das Kompetenzmodell muss die interne Umstrukturierung zulassen, um „modelvariiierende Domänendefinitionen“ (Winther 2010, 83) zu unterstützen.

4 Messprozess

4.1 Assessments

4.1.1 Assessments im Zusammenhang mit Kompetenz und Qualifikation

Die wesentlichen Begrifflichkeiten im Zusammenhang von Kompetenzen (und Qualifikationen) sind, wie bisher entwickelt, in Abbildung 7 zusammengestellt worden. Das Individuum als allumfassende Menge von Kompetenzen, Qualifikationen und anderen Eigenschaften sowie die Zusammenhänge mit den Anforderungswelten sind bereits in Abbildung 5 und Abbildung 6 erläutert worden.

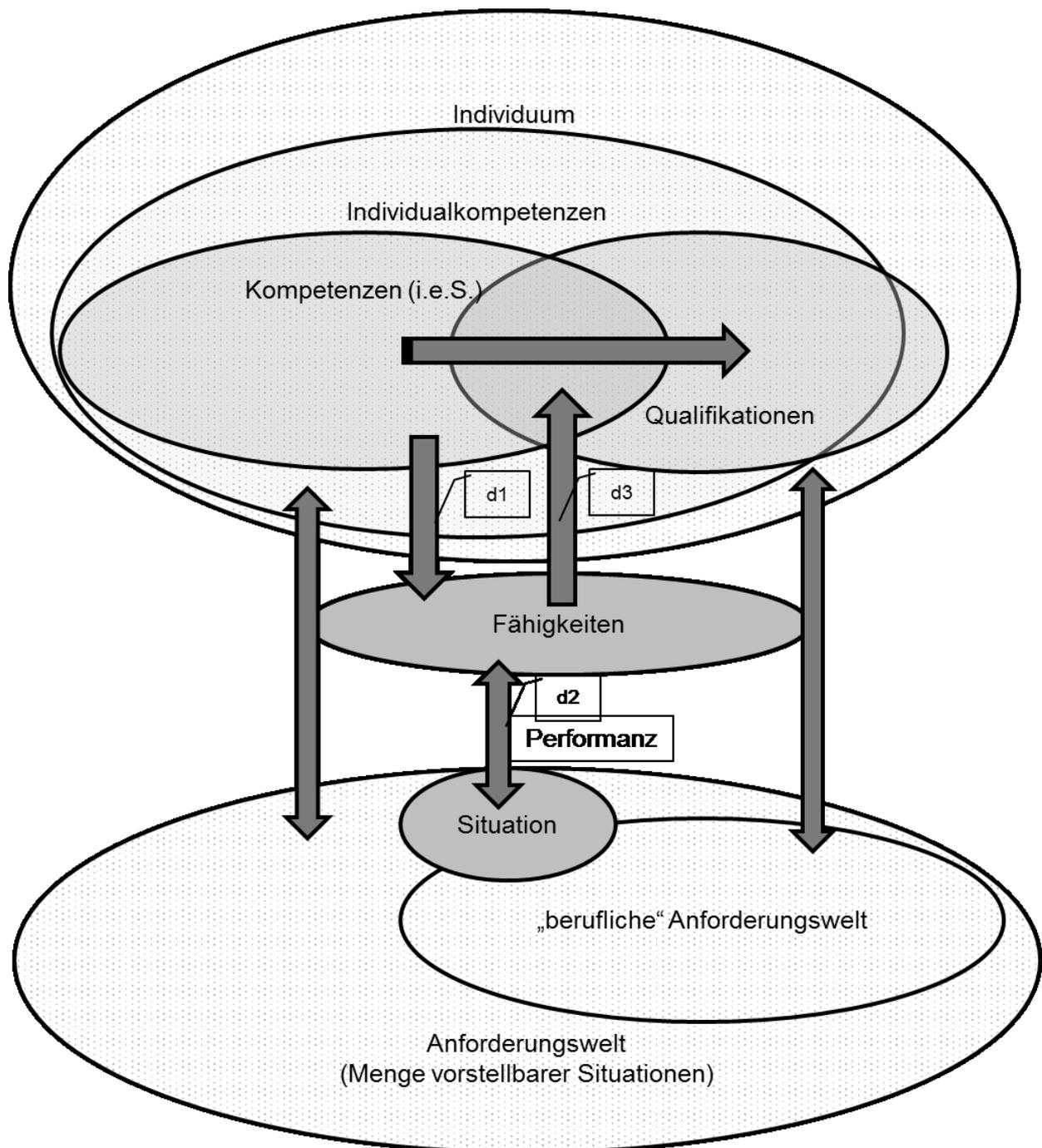


Abbildung 7 Modell der Performanzmessung mittels konkreter Situationen in Assessments

Neben den beiden definierten Kompetenzen ist der Bereich der Qualifikation dargestellt.

Der Bereich der Performanz ist zusätzlich in drei Teile aufgespalten worden, um das Vorgehen bei der Durchführung eines Assessments in diesem Zusammenhang zu visualisieren. Hier geht es um die Operationalisierung des relevanten Ausschnitts des Domäneninhalts bis hin zu Situationen, wie sie in Assessments verwendet werden können.

1. (d1) Zunächst wird der Domäneninhalt modelliert und bis auf die Ebene der zu messenden Fähigkeiten operationalisiert.
2. (d2) Dann können zu den Fähigkeiten konkrete Assessments modelliert werden, die in entsprechende Situationen eingebettet sind.
3. (d3) Schließlich kann durch die Durchführung und Analyse des Assessments die Performanz der einzelnen Personen ermittelt und danach auf die Ebene der Kompetenzen (und Qualifikationen) aus der konkreten Situation abstrahiert werden. Außerdem kann ein individuelles Feedback aus der Analyse abgeleitet werden.

4.1.2 Funktion und Bewertungsform

Assessments können verschiedenen Zielsetzungen dienen. Grundsätzlich ist zwischen Zielsetzungen auf individueller Ebene, also persönlichem Feedback, Versetzungen, Abschlüssen und zwischen Assessments wie PISA zu unterscheiden, die übergeordnete Aussagen über organisatorische, und systemische Aspekte erlauben sollen. (Klieme, Hartig und Rauch 2008, 15). Dabei ist zu beachten, dass die Assessments zweckgebunden unterschiedlich sind: „[...] one size of assessment does not fit all [...]“ (Pellegrino, Chudowsky und Glaser 2001, 222).

Im Folgenden sollen nur Assessments auf individueller Ebene betrachtet werden. Grundsätzlich kann dabei nach dem Zweck zwischen

- summativen und
- formativen

Assessments unterschieden werden. Der Einsatz hängt mit den verschiedenen Funktionen zusammen, die zusammengefasst werden können als (von Saldern 2011, S.84 f.):

- Erziehungsmittel
- Informationserhalt
- Rückmeldung für alle.

Grundsätzlich kann nach der Art der Bewertung unterschieden werden:

1. Ist ein Lernender nahe an den seinen Fähigkeiten entsprechenden möglichen Ergebnissen (ability-referenced)?
2. Ist ein Lernender besser geworden, wenn mit früheren Leistungen derselben Person verglichen wird (growth-referenced)?
3. Ist ein Lernender besser als andere Lernende aus der Vergleichsgruppe, beispielsweise derselben Klasse, anderer Klassen, aller Klassen einer Stufe, anderen Schulen oder noch größeren Vergleichsgruppen (norm-referenced)?
4. Ist ein Lernender nahe an der Zielerreichung bezogen auf ein möglichst objektiv formuliertes Ziel (criterion-referenced)?

Die vier Möglichkeiten werden beispielsweise von van Blerkom erläutert und kommentiert (Blerkom 2009, 17ff.):

1. “With Ability-referenced interpretations the student’s test performance is compared to what we believe the student should be able to do based on his or her ability.”

Die Problematik liegt darin, dass die Fähigkeiten einer Person schwer einzuschätzen sind. Jemand, der nur kurz Mathematik gelernt hat, wird schwächer abschneiden als es seinen eigentlichen Fähigkeiten entspricht. Jemand, der nicht mit Deutsch als Muttersprache aufgewachsen ist, wird in einem Aufsatz oder Diktat normalerweise schwächer abschneiden als es seinen eigentlichen Fähigkeiten entspricht.

2. „With growth-referenced interpretations we compare the student’s test score after instruction with the score of a similar test given prior to instruction.“

Es geht bei dieser Bewertungsart darum, den Lernzuwachs oder Zuwachs an Kompetenz² zu bewerten.

Dabei ergeben sich Probleme hinsichtlich vorgegebener Kompetenzniveaus, insbesondere aber im Bereich der Qualifikationen. Beides erfordert objektive Maßstäbe, die sich nicht aus dem Zuwachs bestimmen lassen. Ein weiteres Problem ist, dass jedes Assessment einem Messfehler unterliegt. Da hier zwei Assessments auszuführen sind, können sich die Fehler stark vergrößern.

Drittens ist die anzunehmende Lernkurve zumeist nicht linear, sondern S-förmig. Dies bedeutet, dass zwei Lernende, die an verschiedenen Punkten starten, unterschiedliche Steigungen haben können und somit unterschiedliche Ergebnisse hinsichtlich des Lernzuwachses zeigen. Der Erfolg hängt dann sehr vom Vorwissen und dem Zeitpunkt der Tests ab.

3. „With norm-referenced interpretations we compare the score that a student received on a test with the scores from some norm group.“

Hier wird das Ergebnis mit dem Ergebnis der anderen Mitglieder der Lerngruppe oder anderer Lerngruppen verglichen. Das allgemeine Ergebnis kann aus einer Gesamterhebung oder aus einer repräsentativen Stichprobe anderer Assessment-Teilnehmer gewonnen werden. Oft werden die Ergebnisse der Assessment-Teilnehmer nur in eine auf- bzw. absteigende Reihenfolge gebracht und dann werden prozentual Noten verteilt.

Der Vorteil dieser Art der Messung ist, dass niemand für schlechten Unterricht bestraft oder für guten Unterricht belohnt wird, zumindest innerhalb der Gruppe. Ein Nachteil ist dagegen, dass Ergebnisse schwer mit anderen Lerngruppen vergleichbar sind. Abhängig von der gewählten Vergleichsgruppe kann das Ergebnis sehr unterschiedlich ausfallen. Damit ergibt sich also kaum eine Aussagefähigkeit über die tatsächlichen Fähigkeiten hinsichtlich externer Vorgaben.

4. „With criterion-referenced interpretations a student’s test score is compared to some pre-set standard or criterion.“

Dies entspricht der Idee einer Output-Orientierung im Sinne eines an externen Kriterien gemessenen Kompetenzstandes. Hier werden Prozentzahlen hinsichtlich der Erreichung von Zielen vorgegeben. Es müssen allerdings die Ziele klar vorgegeben werden. Wenn jemand 75% der Punkte erreicht hat, sollte er nicht 75% der Fragen richtig beantwortet haben, sondern 75% der erwarteten Kompetenzen gezeigt haben. Jedes Ziel sollte zumindest an einer Stelle gemessen werden können.

² Das entspräche der Idee der Kompetenzentwicklungsmodelle (Abschnitt 4.2.2.3).

Ein solches Vorgehen entspricht dem output-orientierten Ansatz und dem Domäneninhalt als objektiver Basis. Daher erscheint diese Bewertungsart an dieser Stelle am sinnvollsten. Andererseits sind viele Situationen denkbar, in denen auf eine andere Methode zumindest partiell zurückgegriffen werden soll.

4.1.3 Messung und Messbarkeit

Jede Bewertung beruht auf einer Messung. Eine Messung bedeutet die Ermittlung eines quantitativen Wertes, der an einer Messskala orientiert sein muss. Diese Skala hat einen Messbereich und eine Dimension, die durch die Art der zu messenden Größe und das Ziel der Messung bestimmt ist. Jede Messung muss unabhängig von der Art der Variablen zumindest vier Punkte berücksichtigen. Neben der Messskala sind dies:

- „Objektivität/Intersubjektivität
- Reliabilität/Zuverlässigkeit
- Validität/Gültigkeit“. (Gnahn 2007, 52)

4.1.3.1 Skala

Eine Skala ordnet den Merkmalsausprägungen die entsprechenden Werte zu. Skalen unterscheiden sich nach dem Messniveau, wobei im Wesentlichen fünf Niveaus zu unterscheiden sind (Schwarze 2001, 33ff.):

- Nominalskala
- Ordinalskala
- Intervallskala
- Verhältnisskala
- Absolutskala

Die Nominalskala erlaubt nur eine Zuordnung zu Merkmalswerten, die miteinander weder verrechnet noch in eine "Größer-kleiner-Relation" gebracht werden können.

Eine Messung auf der Ordinalskala erlaubt zumindest einen "Größer-kleiner"-Vergleich der Merkmalswerte. Dabei ist nur ein Vergleich möglich, die Größe des Unterschieds ist nicht berechenbar. Die Ordinalskala spielt gerade in Kompetenzniveau-modellen, wie sie in Abschnitt 4.2.2.2 thematisiert werden, eine Rolle, da Kompetenzniveaus zumeist auf einer Ordinalskala festgelegt werden.

Auf allen anderen Skalen ist eine weitere Rechnung mit den Messwerten möglich. Sie werden auch unter dem Begriff der Kardinalskala zusammengefasst. Die Unterschiede liegen in der Art der möglichen Berechnungen. Die Intervallskala erlaubt eine Berechnung des Unterschieds zwischen zwei Werten verfügt aber über keinen natürlichen Nullpunkt und keine natürliche Einheit. Die Verhältnisskala erlaubt auch die Bestimmung von Verhältnissen zwischen den Werten, besitzt also einen absoluten Nullpunkt. Die Absolutskala schließlich weist neben dem absoluten Nullpunkt auch eine absolute Skala auf.

4.1.3.2 Objektivität

Die drei Gütekriterien, die für jeden Test zu berücksichtigen sind, sind Objektivität, Reliabilität und Validität. Die Objektivität ist grundlegend für die beiden anderen Kriterien. Zunächst verlangt ein Assessment also Objektivität und somit, dass die Messung unabhängig vom Messenden ist (Intersubjektivität). Es darf nicht relevant sein, wer das Assessment durchführt. Wenn man einen Stein wiegen möchte, muss es egal sein, wer den Stein auf die Waage legt (Saldern 2011, 91).

Es können die Objektivität bei der Durchführung (Assessment-Teilnehmer strengt sich bei der Beobachtung maximal an) und die Objektivität bei der Datenerhebung (keine Sympathie/Antipathie des Bewertenden) unterschieden werden. Im Beispiel mit der Waage ist die Objektivität der Durchführung unmittelbar gewährleistet.

In der Praxis wird die Objektivität leicht verletzt. Dies kann beispielsweise dadurch geschehen, dass in unterschiedlichen Prüfungsgruppen

- vor einem Assessment unterschiedliche Ansagen zu dessen Bedeutung gemacht werden,
- unterschiedliche Bearbeitungszeiten zur Verfügung stehen,
- Fragen erlaubt oder nicht erlaubt werden,
- die Technik unterschiedlich eingeübt wurde oder
- andere gravierende Unterschiede vorliegen.

So dürfte die Objektivität der Durchführung in der Praxis häufig verletzt sein. Anzustreben sind demzufolge Standards für die Durchführung des Assessments. Psychologen sprechen im Zusammenhang mit der gleichartigen Durchführung von einer Standardisierung (Gerrig und Zimbardo 2008, 330f.).

Die Objektivität der Datenerhebung kann beispielsweise durch eine Anonymisierung der Arbeiten erreicht werden, was aber in der täglichen Praxis eher die Ausnahme darstellt. Wesentlich ist daher in jedem Fall eine Dokumentation der Rahmenbedingungen und der Vorgehensweise vor, während und nach der Durchführung eines Assessments.

4.1.3.3 Reliabilität

Die Reliabilität beschreibt die Zuverlässigkeit eines Assessments, also des Messinstrumentes selbst. Insbesondere sollte die Durchführung eines Assessments bei konstantem Wert der zu messenden Eigenschaft zu demselben Ergebnis führen.

Ein Assessment muss also unter denselben Bedingungen³ wiederholbar sein. Voraussetzung ist außerdem das Vorliegen von Objektivität, da andernfalls bereits deren Verletzung jeden Versuch die Reliabilität des Assessments sicherzustellen unmöglich macht. (Rost 2013) (Saldern 2011, 91f.)

Im obigen Beispiel bedeutet dies, dass eine Waage immer dann, wenn derselbe Stein darauf gelegt wird, dasselbe Gewicht anzeigen soll. Auch eine andere geeichte Waage sollte zu demselben Ergebnis kommen. Dies gilt unter gleichen oder unter näherungsweise gleichen Bedingungen.

Reliabilität könnte beispielsweise durch einen Re-Test mit denselben Personen, durch ein vergleichbares Parallel-Assessment oder durch die Messung der internen Konsistenz (beispielsweise split-half-Methode) zumindest näherungsweise bestimmt werden (Gerrig und Zimbardo 2008, 328f.). Statistisch ließe sie sich beispielsweise durch das Varianzenverhältnis, also das Verhältnis der Varianz der tatsächlichen Werte, sofern bekannt, zur Varianz der beobachteten Werte bestimmen (Rost 2013).

4.1.3.4 Validität

Wesentlich ist schließlich die Validität eines Tests, also die Fähigkeit des Messinstrumentes tatsächlich das zu messen, was es messen soll. Die Voraussetzungen für Validität sind zumindest

³ Dieselben Bedingungen können streng genommen nie erreicht werden. Es geht also um die bestmöglich gleichen Bedingungen, soweit diese nicht - erkennbar - zu stark divergieren.

Objektivität und Reliabilität (Rost 2013) (Saldern 2011, 92). Vereinfacht gesagt, wenn sichergestellt werden kann, dass richtig gemessen wird, kommt es darauf an, das Richtige zu messen. So wäre beispielsweise das Wiegen des Steins kaum geeignet, dessen Umfang zu bestimmen.

Bei einer Messung beobachtbarer Eigenschaften einer Person, beispielsweise der Körpergröße, der Haarfarbe oder Schuhgröße ist dies einfacher als bei einer Messung von Empathie, Verantwortungsbewusstsein oder auch Intelligenz. Im ersten Fall handelt es sich um manifeste Variablen, also direkt beobachtbare Größen, im zweiten Fall um latente, nur indirekt bestimmbare Variablen.

Da die letzteren Eigenschaften nicht direkt beobachtbar sind, kann nur aus der Beobachtung der Handlung einer Person und der Messung der dabei beobachteten manifesten Variablen darauf geschlossen werden. Das beobachtete Verhalten dient dann als Indikator für die Eigenschaft, hier die Kompetenz. Diese besondere Schwierigkeit ist letztlich bei jeder Kompetenzmessung zu berücksichtigen. Beobachtet und gemessen werden können bei der Performanz in einer Situation letztlich als manifeste Variable nur Indikatoren, die dann dem Rückschluss auf die vorhandenen Kompetenzen als latente Variable dienen. (Gnahs 2007, 50). Ein solcher Rückschluss ist überhaupt nur möglich wegen der Forderung, dass sich Kompetenzen im engeren Sinne funktional auf Situationen und Anforderungen beziehen.

Hinsichtlich der Validität unterscheidet man zwischen: (Gerrig und Zimbardo 2008, 329ff.)

- Inhaltlicher Validität (Augenscheinvalidität)
- Kriteriumsvalidität
- Konstruktvalidität

Die inhaltliche Validität besagt, dass im Assessment tatsächlich nach den Inhalten gefragt wird, an denen man interessiert ist. Das ist solange nicht mit besonderen Schwierigkeiten behaftet, solange man davon ausgehen kann, dass der Assessment-Teilnehmer Interesse an einer möglichst richtigen Antwort hat. Die Psychologie kennt aber auch Beispiele, wo der Befragte Interesse hatte, die vermeintlich angestrebte Zielrichtung der Fragen zu erkennen und sie in dem vom ihm intendierten Sinne zu beantworten, um das gewünschte Ergebnis zu erhalten. Inhaltliche Validität kann also problematisch sein.

Daher kann man Assessments bevorzugen, die auf die Kriteriumsvalidität abzielen, wenn ein Kriterium bekannt ist, das mit dem Abschneiden in einem Assessment in bekanntem Zusammenhang steht. Ist beispielsweise bekannt, dass das erfolgreiche Abschneiden von Oberschülern in der gymnasialen Oberstufe mit den Abschlussklausuren in Fächern wie Mathematik, Deutsch oder Englisch in der Sekundarstufe I korreliert, so können diese Klausuren als valide Kriterien bezüglich der Aussichten bei einem Wechsel in die gymnasiale Oberstufe angesehen werden.

Verfügt man über kein sinnvolles Kriterium zur Messung der gewünschten Eigenschaften, kann dieses als Konstrukt aufgebaut werden. Wenn dieses Konstrukt aus verschiedenen Merkmalen besteht und ein Assessment hochgradig mit diesen Merkmalen korreliert während es das bei anderen Merkmalen nicht tut, so kann man von Konstruktvalidität hinsichtlich des angegebenen Konstruktes sprechen.

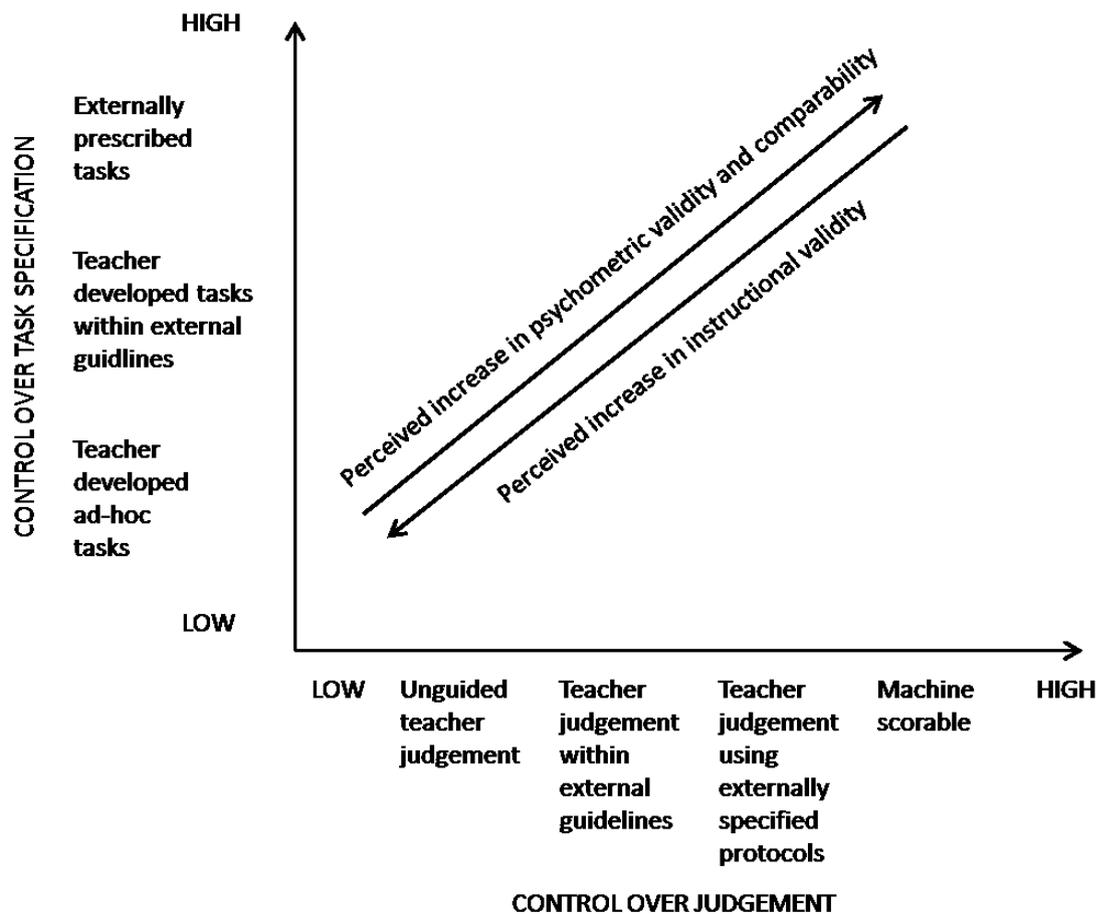


Abbildung 8 Validität nach (Wilson 2004, 10)

Ein anderer Aspekt von Validität, der sich gerade im pädagogischen Umfeld ergibt, ist der Konflikt zwischen der Maxime, dass sich Assessments am Unterricht orientieren sollen und der übergreifenden Vergleichbarkeit (Klasse, Schule, Bundesland, Deutschland, Welt) dieser Assessments. Wilson bezeichnet dies als „instructional validity“ respektive als „psychometric validity“ (oder „comparability“) und visualisiert den grundsätzlichen Widerspruch wie in Abbildung 8 dargestellt. Während man im hier gewählten pädagogischen Umfeld hinsichtlich der inhaltlichen Validität selten Probleme erwarten kann, kann dieser Widerspruch regelmäßig auftreten. Hilfreich kann die getrennte Nutzung von summativen und formativen Assessments sein, die auf denselben Kompetenzmodellen beruhen, die das Spannungsfeld aber letztlich auch nicht komplett auflösen können (Klieme, Bürgermeister, et al. 2010, 66).

4.1.3.5 e-Assessments

Schließlich muss der Aufwand für ein Assessment berücksichtigt werden. Der Aufwand, der in realen Situationen betrieben werden kann, ist stets beschränkt. Soll ein Verfahren zur Kompetenzmessung sinnvoll anwendbar sein, muss es stets auch mit den vorhandenen Ressourcen durchführbar sein.

Eine besondere Bedeutung kommt in diesem Zusammenhang e-Assessments zu, also elektronisch durchgeführten und zumindest teilweise automatisch auswertbaren Assessments. Auch in dieser Arbeit wird beispielhaft ein e-Assessment verwendet. Diese Art der Assessments wird einerseits bei vielen Großbefragungen verwendet, andererseits immer wieder kritisiert. Insgesamt gibt es um e-Assessments kontroverse Diskussionen, wo immer sie angewendet werden. So findet man im

universitären Bereich beispielsweise an der Universität Mainz in der Praxis derartige Auseinandersetzungen. Letztlich kommt es jedoch auf die Art der Aufgabenstellung an. Was immer für derartige Assessments spricht, ist das Argument der Zeitersparnis und somit von Ressourcen, die man für Unterricht, Forschung oder Lehre verwenden kann. (Gartz 2014)

Gerade die geschlossenen Fragestellungen vieler e-Assessments erscheinen manchmal problematisch und offenere Fragestellungen werden als wünschenswert angesehen, auch weil vorgegebene Lösungen manche Fehler nicht zulassen (Büchter und Leuders 2011, 169).

Andererseits sind auch klassische paper-and-pencil-Assessments fehlerbehaftet, beispielsweise durch Halo-Effekte, was im Wesentlichen auf die handelnde Lehrkraft zurückzuführen ist. Gerade die Objektivität ist keineswegs immer gewährleistet. Daher wird beispielsweise vorgeschlagen, Beurteilungsfehler der Lehrkraft durch vielfältigere Ergebnisse zu verringern (Saldern 2011, 96f.).

Generell lassen ein geringerer Einsatz von Ressourcen, ein schnelleres Generieren der Ergebnisse, die Möglichkeit neuer Testformate sowie Vorteile insbesondere hinsichtlich Objektivität und Reliabilität e-Assessments attraktiv erscheinen. Die Bestimmung der Korrelation zwischen einem e-Assessment und einem klassischen paper-and-pencil-Assessment hat zwar unterschiedliche Ergebnisse gezeigt, scheint aber im Allgemeinen vorhanden zu sein, wenn die Assessments inhaltlich vergleichbar sind. (Jurecka 2008, 199f.)

Die Probleme aller Arten von Assessments gelten auch für den schulischen Bereich. Kann öfter und mit geringerem Aufwand ein Assessment durchgeführt werden, das außerdem ein gezieltes Feedback für Lehrkräfte und Lernende und die Gestaltung individueller Hilfen für die Schülerinnen und Schüler in der Schule beinhaltet, kann dies dem Gesamtprozess letztlich nur förderlich sein, wie beispielsweise auch Hattie zeigt (Hattie 2013, 212).

Insofern werden im Rahmen dieser Arbeit e-Assessments präferiert, aber andere Assessments nicht ausgeschlossen.

4.2 Kompetenzdiagnostik

Bereits in der Zielsetzung wurde die zentrale Rolle der Kompetenzmessung angesprochen. Es sei noch einmal mit Leutner und Klieme auf die Bedeutung hingewiesen: „Die Modellierung und Messung von Kompetenzen spielt eine entscheidende Rolle bei der Optimierung von Bildungsprozessen sowie der Qualitätssicherung und Qualitätsentwicklung des Bildungswesens“ (Leutner, Klieme, et al. 2013, 2).

4.2.1 Forschungstraditionen

Die klassischen Forschungstraditionen zur Modellierung und Messung von Kompetenz lassen sich nach Winther (Winther 2010, 17ff.) einteilen in

- 1) Die behaviouristische Linie, die stark in den USA vertreten wird und auf Tests auf der Ebene der Performanz abzielt. Sowohl die Situation als auch durch sie ausgelöstes Verhalten sind a priori möglichst exakt zu bestimmen. Dementsprechend muss die zu messende Fähigkeit klar definiert werden. Die Messung ist subjektbezogen, also auf die Person ausgerichtet. Gemessen werden Wissen (knowledge) und Können (ability). Die wesentliche Kritik an der behaviouristischen Linie ist, dass sich Kompetenz erst in

komplexen Situationen in der Fähigkeit zeigt, das Wissen und Können zielgerichtet anwenden zu können.

- 2) Die Konzentration auf „generic skills“, wie sie in Großbritannien, Australien aber auch Deutschland unter dem Stichwort „Schlüsselqualifikationen“ erfolgt ist, möchte möglichst unabhängig von bestimmten Inhalten objektive Fähigkeits- und Einstellungsbündel definieren und zum Ziel des Lernprozesses machen. Der Hauptkritikpunkt ist, dass der Universalitätsanspruch dieser Schlüsselqualifikationen nicht aufrecht zu erhalten ist, weil sich allgemeine Fähigkeiten weder inhaltsfrei vermitteln noch messen lassen.
- 3) Die kognitivistische Auffassung findet sich erstmalig bei Chomsky und wird vor allem von der OECD vertreten. Auch wenn die Definitionen sehr vielfältig sind (Winther 2010, 21), geht es bei der kognitivistischen Auffassung von Kompetenz um Strukturen und Regeln, die Fähigkeiten hervorbringen, während Performanz die beobachtbare Fähigkeit in einer Situation darstellt.

Unter dem Blickwinkel der in dieser Arbeit verfolgten Ziele erscheint es wichtig, sich an konkreten Inhalten zu orientieren, da nur sie auch ein konkretes Feedback ermöglichen können.

Berücksichtigt man zudem die Betonung des breiten situativen Ansatzes und die Definition von Fähigkeiten beziehungsweise Kompetenzen im engeren Sinne erscheint der kognitivistische Ansatz für das hier verfolgte Ziel am geeignetsten.

4.2.2 Kompetenzmodellarten

Damit stellt sich die Frage, wie Strukturen und Regeln für die Kompetenz in Modellen dargestellt werden können. Die Expertise „Zur Entwicklung nationaler Bildungsstandards“ sagt „Kompetenzmodelle stellen damit die Grundlage für Operationalisierungen von Bildungszielen dar, die den Output des Bildungssystems über das Erstellen von Testverfahren [...] empirisch zu überprüfen erlauben. Die Rolle von Kompetenzmodellen besteht darin, zwischen abstrakten Bildungszielen und konkreten Aufgabensammlungen zu vermitteln.“ (BMBF 2003, 71).

Das Ziel dieser Modelle ist eine Abstraktion von einer konkreten Situation (Lernsituation, Leistungsüberprüfungssituation) vorzunehmen und allgemeinere Aussagen zu treffen, „[...] weil man auf Fähigkeiten zielt, die nicht allein aufgaben- und prozessgebunden erworben werden, sondern ablösbar von der Ursprungssituation, zukunftsfähig und problemoffen sind. Man erkennt damit unschwer, dass auch Kompetenzmodelle, theoretische Beschreibungen der Struktur dieser spezifischen Fähigkeiten und der Stufen ihres Erwerbs, als Antwort auf typische moderne Problemlagen konzipiert sind.“ (BMBF 2003, 65).

Aus unterrichtlicher Sicht übernehmen Kompetenzmodelle die Rolle, die zuvor Lehrziele hatten. Beide beschreiben die Fähigkeiten, die die Lernenden im Bildungsprozess erwerben sollen (Schott und Ghanbari 2008, 14).

Schaut man sich die Ansätze dieser Modelle hinsichtlich der Operationalisierung näher an, so lassen sich grundsätzlich drei verschiedene Modelltypen unterscheiden: (Winther 2010, 37ff.)

- Kompetenzstrukturmodelle,
- Kompetenzniveaumodelle und
- Kompetenzentwicklungsmodelle.

4.2.2.1 Kompetenzstrukturmodelle

Kompetenzstrukturmodelle beinhalten eine systematische, strukturierte Beschreibung des Inhaltes einer Domäne aus Sicht der Fachwissenschaft und Fachdidaktik. Die essentiellen Prinzipien und Ideen der Domäne werden zur Strukturierung der verschiedenen Kompetenzen verwendet, die dann in Situationen eingebunden werden können. (Winter 2014, 38)

Die Struktur der Kompetenzen scheint sich eher über eine Modellierung als über statistisch empirische Modelle ermitteln zu lassen. Bei der Frage, ob empirische Modelle eine einzige oder tatsächlich unterschiedliche Kompetenzbereiche messen, ergibt sich ein differenziertes Bild. Während Untersuchungen etwa bei Pisa 2003 zeigen, dass dort eher ein einziges domänenübergreifendes Dimensionsmerkmal als Ursache der gezeigten Leistungen gemessen wird, geben andere Studien eher Grund zu der Annahme der Eignung mehrdimensionaler Modelle. (Winther 2010, 39)

Die Annahme eines mehrdimensionalen Modells als Erklärung der gezeigten Kompetenzen erscheint also sinnvoll. Bei der Gestaltung des Assessments werden den einzelnen Kompetenzbereichen oder Kompetenzdimensionen zumeist direkt Items zugeordnet. Teilweise werden auch a-priori Annahmen über das Kompetenzniveau einzelner Items oder Antwortalternativen getroffen. (Bayrhuber, et al. 2010) (Reiss, Heinze und Pekrun 2007)

Bei mehrdimensionalen Modellen lassen sich Kompetenzen auf mehrere, nicht voneinander isolierte Teile zurückführen. Auch lassen sich themenbezogene und auf breitere Kompetenzen angelegte Aufgaben miteinander kombinieren. (Klieme, Bürgermeister, et al. 2010, 69) Die Beschränkung auf eine Verfeinerungshierarchie erscheint nicht zwingend und schränkt die Flexibilität der Modellierung ein.

Das bedeutet für das in dieser Arbeit entwickelte Model, dass eine Abbildung verschiedener Dimensionen als Erklärungsmerkmale sinnvoll ist und möglich sein muss. Die Einstufigkeit einer direkten Zuordnung von Kompetenzen zu Aufgaben erscheint nicht unmittelbar zwingend. Sie wird durch optional mehrstufige Gliederungen innerhalb der Dimensionen ersetzt.

4.2.2.2 Kompetenzniveaumodelle

Diese Modelle ordnen Kompetenzniveaus Anforderungen in verschiedenen Aufgaben so zu, dass das jeweilige Niveau gerade für deren Lösung ausreicht und versuchen die Kompetenz eines Subjekts zu messen, indem in konkreten Testsituationen gemessen wird, welche Anforderungen bewältigt werden. Die Kompetenz wird auf einer kontinuierlichen Skala gemessen, die durch Stufung in verschiedene Niveaus unterteilt werden kann. Die Anforderungen der einzelnen Testitems und das Kompetenzniveau werden in den zugehörigen statistischen Modellen simultan geschätzt. Das einfachste Modell dieser Art ist das dichotome Rasch-Modell. Grundsätzlich lassen sich so Aussagen über den eigentlichen Test hinaus über die Kompetenz des Individuums treffen. (Winther 2010, 40ff.)

Die inhaltliche Charakterisierung von Kompetenzniveaus durch Anforderungsmerkmale der Aufgaben erscheint zumindest teilweise möglich. Allerdings ist der Zusammenhang zwischen Aufgabenmerkmalen und Schwierigkeitsgrad nicht unbedingt eindeutig und wird beispielsweise vom Aufgabentyp beeinflusst (Neubrand, et al. 2002, 117).

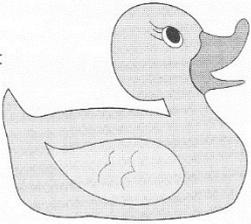
Zudem sind in der Praxis die Lehrkräfte eher an wenigen, komplexen, fachlich interessanten Aufgaben orientiert, während bei der Ermittlung von Kompetenzniveaus eher viele auf bestimmte Kompetenzen abzielende und hierarchisch gestufte Aufgaben relevant sind und die Basis für psychometrische Analysen bilden (Leutner, Fleischer, et al. 2007, 153).

Hier wird beispielsweise mit Kompetenzniveaus (Ebenen) gearbeitet, die durch Beispiele oder abstrakte Zuordnungsregeln definiert werden (Reiss, Heinze und Pekrun 2007, 115). Dies findet man auch direkt in Form von Kompetenzskalen (Leutner, Fleischer, et al. 2007, 156).

Zusätzlich zu der Zuordnung der entsprechenden Niveaus durch psychometrische Modelle gibt es Versuche, bereits bei der Konstruktion der Assessment-Items bei der Vorgabe von Antwortalternativen eine Zuordnung dieser Alternativen zu Kompetenzniveaus zu treffen. Ein Beispiel bei dem unter anderem auf fachdidaktische und entwicklungspsychologische Forschung zurückgegriffen wurde ist in Abbildung 9 dargestellt. Inwiefern diese Zuordnung im Konsens getroffen werden kann, erscheint allerdings im realen Kontext zumindest fraglich.

Gummiente

Instruktion des Testleiters:
 Ich habe hier eine Gummiente. Mit so einer Gummiente kann man in der Badewanne spielen. Wie ihr seht, schwimmt die Gummiente oben auf dem Wasser (Demonstration!).



Woran liegt es, dass die Gummiente schwimmt?

Kreuze nach jeder Antwort ‚Richtig‘ oder ‚Falsch‘ an!

	Richtig	Falsch
1. Die Gummiente schwimmt, weil sie innen hohl ist.	<input type="checkbox"/>	<input type="checkbox"/>
2. Die Gummiente schwimmt, weil das Wasser sie nach oben drückt.	<input type="checkbox"/>	<input type="checkbox"/>
3. Die Gummiente schwimmt, weil sie sehr leicht ist.	<input type="checkbox"/>	<input type="checkbox"/>

Was ist die beste Antwort? Nr. _____

Abbildung 9 Beispiel eines Items mit Antworten auf verschiedenen Niveaus (Hardy, et al. 2010, 119)

Kompetenzniveaus werden zumeist in Taxonomien organisiert. Taxonomien bilden Systeme von Kompetenzniveaus. Die bekannteste Taxonomie für Lernziele stammt von Bloom (Bloom 1976) nach (Bloom 1956). Dabei handelt es sich um verallgemeinerte, nicht domänengebundene Kompetenzniveaus.

Im kognitiven Bereich sind als Niveaus die Stufen

1. Evaluation
2. Synthese
3. Analyse
4. Anwenden
5. Verstehen
6. Wissen

definiert worden. In einer revidierten Fassung wurde Wissen durch „Erinnern“ ersetzt und die „Analyse“, „Synthese“ und „Evaluation“ durch die drei Begriffe „analysieren“, „evaluieren“ und „erschaffen“ (Anderson und Krathwohl 2001, 313ff.).

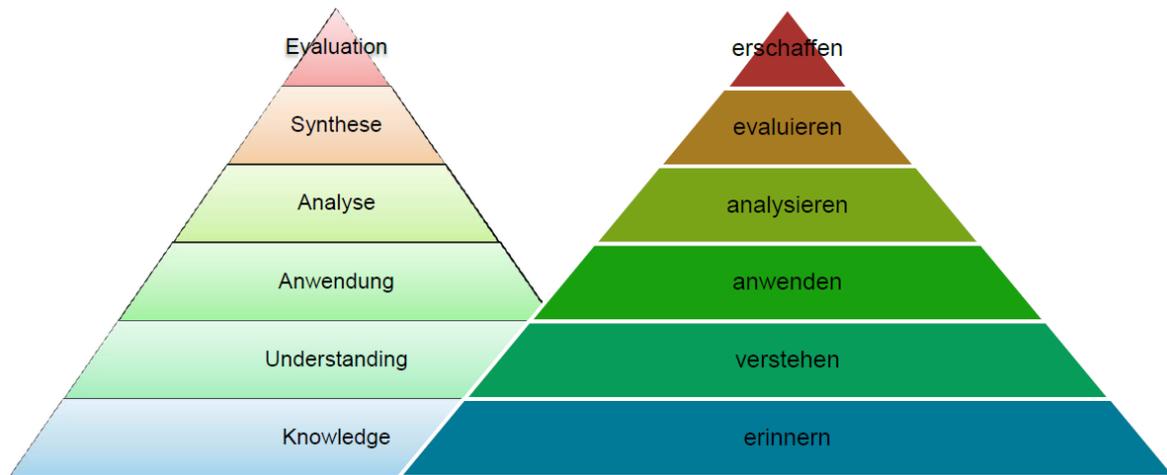


Abbildung 10 Ursprüngliche und von Anderson/Krathwohl revidierte Bloom'sche Taxonomie (ETH - Zürich nach Bloom ... 2015)

Ein anderes prominentes Beispiel ist die SOLO-Taxonomie (SOLO = Structure of the Observed Learning Outcome) nach Biggs (Biggs und Collis 1982) und (Biggs 2003):

1. Extended abstract
2. Relational
3. Multi-structural
4. Uni-structural
5. Unstructured

Vielfach werden den einzelnen Kompetenzstufen Verben zugeordnet, die im Sinne von Operatoren diese Stufen anzeigen. Dementsprechend findet man Operatorenlisten beispielsweise bei Bloom direkt (Bloom 1976) oder für SOLO bei Poloczek (Poloczek 2007) sowie in vielen Curricula.

Neben den betrachteten eindimensionalen Taxonomien gibt es auch mehrdimensionale Taxonomien. Mehrdimensionale Taxonomien findet man beispielsweise bei Baumgartner (Baumgartner 2014) oder in einer Übersicht des Instituts für Erziehungswissenschaft der Universität Tübingen (siehe Abbildung 11).

Dimension	Ausprägungen			
1. Wissensart	Fakten	Prozeduren	Konzepte	Metakognition
2. Kognitiver Prozess	Reproduktion	Naher Transfer	Weiter Transfer	Problemlösen
3. Wissensseinheiten	Eine WE	Bis zu 4 WE		Mehr als 4 WE
4. Offenheit	Definiert/konvergent	Definiert/divergent	Ungenau/divergent	
5. Lebensweltbezug	Kein	Konstruiert	Authentisch	Real
6. Sprachlogische Kompl.	Niedrig	Mittel		Hoch
7. Repräsentationsformen	Eine	Integration	Transformation	

Abbildung 11 Kompetenzniveaudimensionen (Universität Tübingen 2012, 2)

Schließlich darf in der Praxis auch die weit verbreitete Unterscheidung der Anforderungsbereiche I, II und III nicht vernachlässigt werden, wie sie zum Beispiel in den Bildungsstandards für das Abitur im Fach Mathematik wieder vorgenommen wurde (Kultusministerkonferenz 2012, 11).

Daneben existieren in der Praxis verschiedene Kompetenzstufendefinitionen, deren Stufung in vielen Fällen aber kritisch zu hinterfragen ist. (Saldern 2011, 162ff.)

Generell ist das Konzept der Kompetenzniveaus akzeptiert und es muss im hier entwickelten Modell eine Möglichkeit der flexiblen Definition der verschiedenen Stufen wie auch der Operatoren geschaffen werden.

Das Modell soll daher die Möglichkeit der Abbildung verschiedener Niveaustufen basierend auf einer kontinuierlichen Skala ermöglichen und sowohl die Mehrdimensionalität als auch die Verwendung von Operatoren einbeziehen können.

4.2.2.3 Kompetenzentwicklungsmodelle

Kompetenzentwicklungsmodelle orientieren sich am Entwicklungsprozess der Lernenden. Kompetenzen werden nicht nur aus den Anforderungen, sondern auch entwicklungspsychologisch erklärt. Vorbild sind in vielen Fällen die Arbeiten Piagets. Jede Entwicklungsstufe setzt das differenzierte Verständnis der vorhergehenden (einfacheren) Entwicklungsstufen voraus (Winther 2010, 42). Tatsächlich lassen sich viele Erfolge durch deren Berücksichtigung erklären (Hattie 2013, 52).

Einwände sind, dass durch die Modelle die Entwicklungsstufen und deren Zusammenhang nicht plausibel gemacht werden (Winther 2010, 42). Sie können aber die Probleme des Auffindens der Kompetenzstufen deutlich machen (ebd). Aus Sicht des hier zu entwickelnden Modells stellt dieser Ansatz einen Vorher-Nachher-Vergleich dar, also die Modellierung und den Vergleich verschiedener Zustände. Dies stellt technisch die Anforderung an die Speicherung und den Vergleich von Zuständen, aber keine an eine eigene Typisierung.

Das Modell kann also Kompetenzentwicklungsmodelle als Hilfestellung in den Modellierungsprozess und die Analyse beispielsweise durch Vergleich mehrerer Assessments für einen Assessment-Teilnehmer einbinden, muss selbst aber keinen Entwicklungsprozess modellieren.

4.2.3 Quantitative Kompetenzdiagnostik

Ein Kompetenzmodell bildet die Basis, den Kern für die Kompetenzdiagnostik, muss aber um die eigentliche Messung erweitert werden. Auf Grund der Bedeutung des Themas wurde 2006 das Schwerpunktprogramm 1293 (Kompetenzdiagnostik) der Deutschen Forschungsgemeinschaft

(DFG) gegründet. Die dort verfolgten eher quantitativen Ansätze haben zu einer Strukturierung des Bereiches geführt wie in Abbildung 12 dargestellt.

Die Kultusministerkonferenz (KMK) hat deren Entwicklung vorangetrieben und das Institut zur Qualitätssicherung im Bildungswesen (IQB) zur wissenschaftlichen Unterstützung gegründet.

Die Kompetenzmodelle des Kerns (1) beschreiben die strukturelle Gliederung der Kompetenzen sowie die Definition von Kompetenzniveaus. Hier sieht man die Anknüpfung an die Kompetenzstrukturmodelle und die Kompetenzniveaumodelle. Die psychometrischen Modelle (2) folgen der Frage, wie sich Kompetenzmodelle einer Erfassung zugänglich machen lassen. Danach geht es im Rahmen der Messkonzepte und Messverfahren (3) um die konkrete Gestaltung dieser Modelle in empirischen Verfahren. Und schließlich beschreibt die vierte Ebene (4) die Nutzung von so gewonnenen Informationen für verschiedene Akteure des Bildungswesens (Fleischer, et al. 2013, 8ff.).

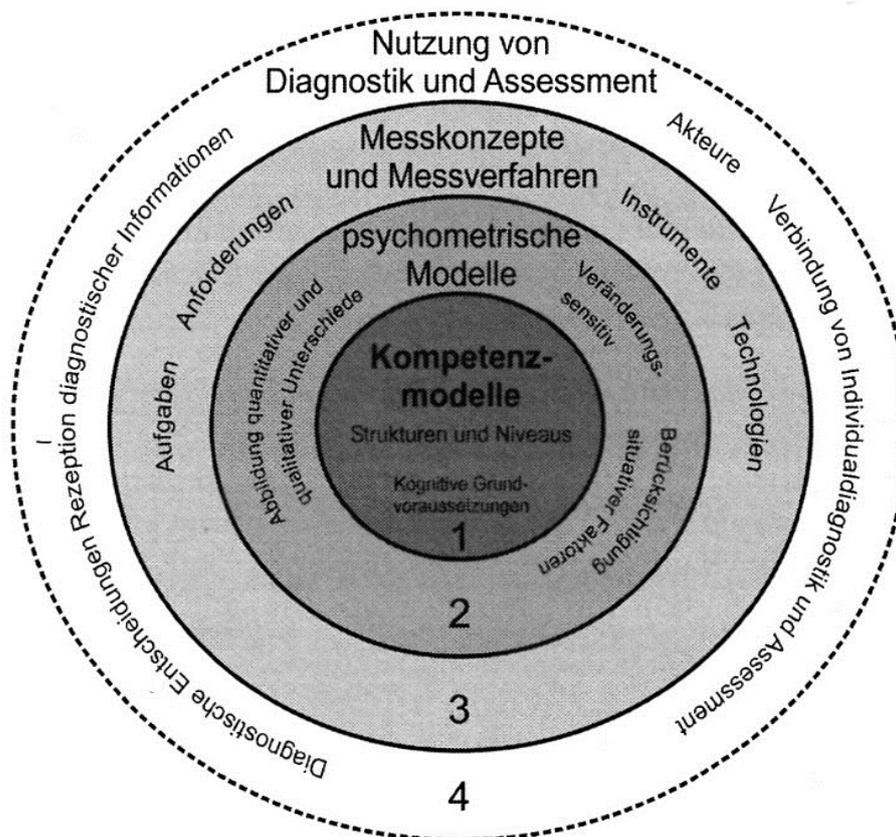


Abbildung 12 Struktur des DFG Schwerpunktprogramms 1293 (Fleischer, et al. 2013, 7)

Man sieht, dass die Modellierung der Kompetenzen den Kern jeder Messung und jeder Nutzung darstellt. Aufbauend auf definierten Kompetenzstrukturen können statistische Modelle gebildet werden, die mittels geeigneter Instrumente den am Bildungsprozess beteiligten Stakeholdern Informationen liefern können.

4.2.4 Individualfeedback

Im Rahmen des DFG-Schwerpunktprogramms Kompetenzdiagnostik standen zunächst Large-Scale-Assessments wie PISA oder TIMSS im Vordergrund, weil die Systeme als zu komplex für die

Individualdiagnostik angesehen werden. Das schließt den Nutzen dieser Systeme für die Individualdiagnostik nicht prinzipiell aus, sondern wirft die Frage nach einer solchen Nutzung auf und nach der Möglichkeit daraus Feedback für die Assessment-Teilnehmer wie auch die Lehrenden zu erzeugen.

„In deutschen Schulen kommt Feedback häufig zu spät.“ (Saldern 2011, 86). Der Satz zeigt ein Grundproblem der täglichen Unterrichtspraxis auf. Ein erfolgreicher Lernprozess erfordert, dass der Lernende ein Feedback über seine erreichten Erfolge und die noch verbesserungsfähigen Bereiche bekommt. So selbstgesteuert und selbstbestimmt Lernen auch erfolgen soll, so stehen doch auch die Bildungsziele im rechtverstandenen Interesse des Lernenden. Lernen muss auch in Fähigkeiten münden, die als Ziele eines Lernprozesses definiert sind und deren Erreichen ist zu überprüfen und das Ergebnis dem Lernenden zurückzumelden. Gleichzeitig muss dem Lehrenden ein detailliertes Feedback zur Verfügung stehen, um durch geeignete Steuermechanismen den Bildungsprozess zu optimieren.

Jede solche Überprüfung kann einen den Lernprozess lenkenden, formativen Charakter oder einen bewertenden, summativen Charakter haben. Häufig mischen sich in der Praxis beide Aspekte. Gerade im Fall des formativen Charakters steht durch das Aufzeigen von Zielabweichungen eine Steuerung des weiteren Lernprozesses im Mittelpunkt. Je häufiger und zielgenauer ein Feedback durch eine solche Überprüfung erfolgt, desto genauer kann der individuelle Lernprozess gesteuert werden. So wird Feedback als zentrales Element eines formativen Assessments betont und kann mit weiterführenden Instrumenten gekoppelt werden (Klieme, Bürgermeister, et al. 2010, 66).

Der Komplexitätsaspekt ist also zentral für die Individualdiagnostik, schließt aber deren Sinnhaftigkeit nicht aus (Pant 2013, 77). Insbesondere gilt das auch für das Feedback für die Lehrenden, die eine möglichst genaue Rückmeldung über die Lerngruppe benötigen (Hattie 2013, 206). Obwohl also das Feedback im Rahmen der Individualdiagnostik als komplex angesehen wird (Winther 2010, 263), sind somit Modelle für Assessments auch im Bereich der Individualdiagnostik wünschenswert.

4.2.5 Psychometrische Modelle

Die zweite Ebene des DFG-Schwerpunktprogramms (siehe Abbildung 12) weist die psychometrischen Modelle zur statistischen Analyse aus. Diese Modelle werden auch unter dem Begriff der Item Response Theory (IRT) zusammengefasst. Entsprechende Modelle sind beispielsweise auch als Erweiterungspakete der Statistiksoftware R implementiert.

4.2.5.1 Grundidee des RASCH-Modells

Die Idee der IRT stammt aus der Psychologie und wird in vielen Situationen für psychologische Tests verwendet. Die zu messenden Eigenschaften sind dabei latente Variablen, also Variablen, die nicht unmittelbar beobachtet werden können, sondern aus anderen manifesten Variablen abgeleitet werden müssen. Die manifesten Variablen entsprechen im Wesentlichen den Items, während die latenten Variablen der Bestimmung von Kompetenzen dienen. Es wird dann der Wert für die Kompetenz, also die latente Variable, für eine Person in Form eines Traits (Kompetenz) geschätzt. Dabei wird derjenige Wert für den Trait einer Person gewählt, für den die Ergebnisse dieser Person in den manifesten Variablen (Items) am wahrscheinlichsten sind. Das kann man mit einer Diagnosesituation vergleichen, bei der die Diagnose gewählt wird, die die beobachteten Symptome am wahrscheinlichsten erklärt (Embretson und Reise 2000, 54).

Viele dieser Modelle erfordern relativ große Datenmengen insbesondere, wenn viele Parameter zu schätzen sind. Das RASCH-Modell ist eine Variante der IRT, die eine relativ einfache Gleichung besitzt und auch für kleinere Datenmengen valide Ergebnisse liefern kann (Hartig und Frey 2013, 50). Auch wenn es ursprünglich getrennt von anderen IRT-Modellen entwickelt wurde, entspricht es einem einparametrischen Modell (1PL) der IRT, so dass von einem IRT-Modell gesprochen werden kann (Demars 2010, 15).

Das RASCH-Modell lässt nur dichotome Items zu, unterscheidet also nur zwischen „richtig“ und „falsch“. Die dem Modell zugrundeliegende Idee ist, dass die Wahrscheinlichkeit, dass eine Person ein Item richtig beantworten kann von zwei Faktoren abhängt:

- Je höher die Kompetenz einer Person hinsichtlich eines Items ist, desto höher ist die Wahrscheinlichkeit, dass die Person dieses Item richtig beantwortet.
- Je geringer die Schwierigkeit eines Items ist, desto höher ist die Wahrscheinlichkeit, dass eine Person dieses Item richtig beantwortet.

Dementsprechend berücksichtigt das RASCH-Modell in seiner Modellfunktion diese beiden Eigenschaften in Form von Parametern. Die Modellfunktion ist definiert als

$$P(U_{ij} = 1 | \theta_i, \beta_j) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}$$

und wie bei allen Wahrscheinlichkeiten umfasst der Wertebereich die reellen Zahlen im Intervall $[0,1]$.

Dabei gilt:

- θ_i beschreibt die Kompetenz der Person i . Je höher der Wert ist, desto höher die Kompetenz der Person i bezogen auf die gemessene Variable.
- β_j beschreibt die Schwierigkeit des Items j . Je höher der Wert ist, desto schwieriger ist das Item j bezogen auf die gemessene Variable.

Damit beschreibt die Modellfunktion die Wahrscheinlichkeit, dass eine Person mit der Kompetenz θ_i ein Item mit der Schwierigkeit β_j richtig lösen kann. An der Differenz $\theta_i - \beta_j$ erkennt man, dass diese umso größer ist, je größer der Unterschied zwischen individueller Kompetenz und Schwierigkeitsgrad des Items ist. Die Funktionsgleichung beschreibt eine logistische Funktion, wie sie beispielhaft in Abbildung 13 dargestellt ist.

Diese Funktion stellt die Wahrscheinlichkeit für die Lösung eines Items in Abhängigkeit der individuellen Kompetenz dar. Die Kurve wird zumeist als ICC (Item Characteristic Curve) bezeichnet, da sie vom Schwierigkeitsgrad eines einzelnen Items abhängig ist. Der itemspezifische Schwierigkeitsgrad β_j bestimmt die Lage der Funktion. Man sieht in Abbildung 13, dass die Lösungswahrscheinlichkeit für ein Item genau dann 0,5 beträgt, wenn eine Person eine Kompetenz θ hat, die gerade mit dem Schwierigkeitsgrad β des Items übereinstimmt. Je schwieriger ein Item ist, desto weiter wird seine ICC nach rechts verschoben und umgekehrt. Die Steigung der ICC wird als deren Trennschärfe bezeichnet. Je steiler die Kurve, desto trennschärfer das Item, da die Wahrscheinlichkeit, ein Item zu lösen, bereits bei kleinen Kompetenzschwankungen stark zu- oder abnimmt. Form und Steigung aller ICC sind im RASCH-Modell (im Gegensatz zu komplexeren

IRT-Modellen) gleich, da die Modellgleichung keine Parameter enthält, die diese weiter beeinflussen könnten.

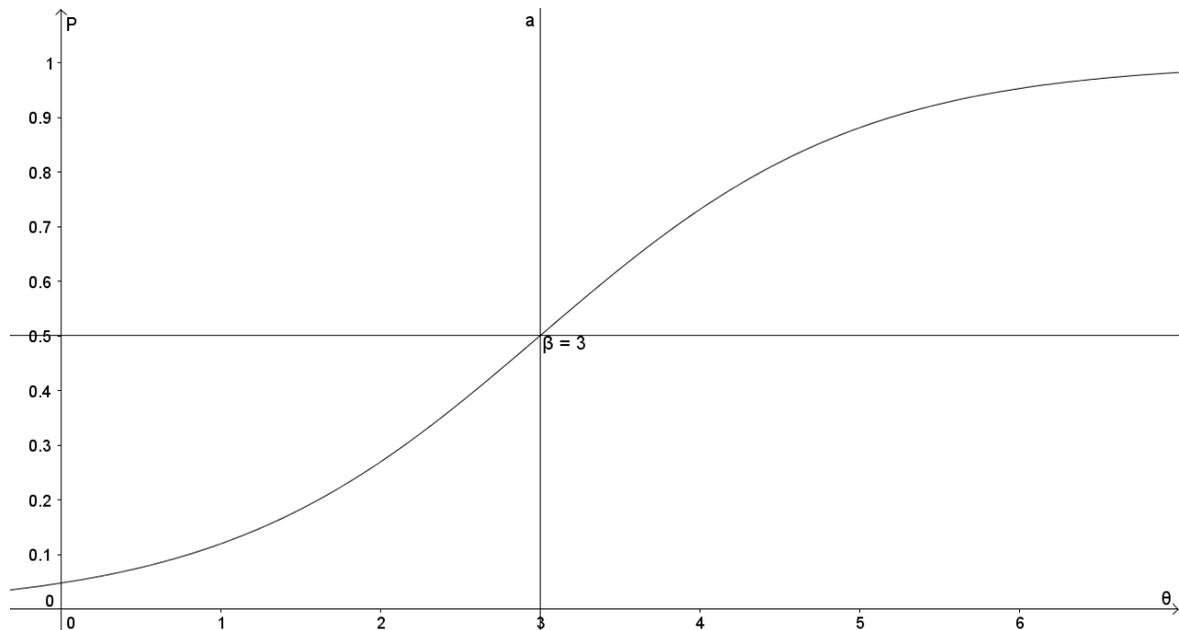


Abbildung 13 ICC-Beispiel

4.2.5.2 Voraussetzungen

Wie jedes statistische Modell setzt das RASCH-Modell die Erfüllung bestimmter Grundannahmen voraus (Strobl 2010, 14ff.):

1. Stochastische Unabhängigkeit der Items: Die Wahrscheinlichkeit, dass eine Person ein Item lösen kann, darf nicht von der Lösung einer anderen Aufgabe abhängen. Das bedeutet insbesondere, dass Aufgaben nicht aufeinander aufbauen dürfen.
2. Stochastische Unabhängigkeit der Personen: Die Wahrscheinlichkeit, dass eine Person ein Item lösen kann, darf nicht davon abhängen, dass eine andere Person dieses Item lösen kann. Das bedeutet insbesondere, dass die Personen nicht voneinander „abschreiben“ dürfen.
3. Spezifische Objektivität für Aufgaben und Personen: Man vergleicht für jeweils zwei Personen und zwei Items die Wahrscheinlichkeit, diese richtig zu lösen. Dann muss für alle Items gelten, dass wenn eine der Personen ein Item mit größerer Wahrscheinlichkeit lösen kann, gilt dies auch für das andere Item. So darf die Lösungswahrscheinlichkeit eines Items nicht für eine Person A höher als für eine Person B sein, wenn dies für ein anderes Item umgekehrt gilt. Durch die parallelen ICCs mit gleicher Steigung ist die spezifische Objektivität im RASCH-Modell im Gegensatz zu einigen anderen IRT-Modellen grundsätzlich gewährleistet. Das unter Umständen dahinter stehende Differential Item Functioning (DIF) weist allerdings auf ein Problem hin, das immer zu beachten ist. Enthalten Items spezifische Eigenschaften, für die andere Kompetenzen benötigt werden (beispielsweise Fremdwörter im Mathematiktest) kann dies zu einer Verzerrung der Messergebnisse führen.
4. Eindimensionalität: Es wird nur eine Kompetenz gemessen. Items und Kompetenzen liegen auf einer Achse mit derselben Skala und werden auch direkt voneinander subtrahiert. Sie müssen daher auch auf derselben Kardinalskala messbar sein. Die

genauere Modellherleitung zeigt allerdings, dass nur die Intervallskala hier gesichert ist. Damit sind beispielsweise Mittelwertberechnungen erlaubt.

Ein komplett „kalibrierter“ psychologischer Test erfordert eine sorgfältige Konstruktion der Items, um eine Verzerrung der Messungen zu verhindern. So sollten die entsprechenden Kompetenzen gemessen und die Messung anderer Kompetenzen vermieden werden, um keine Verzerrung der Ergebnisse zu erhalten. Dementsprechend sollte die Einhaltung der Modellvoraussetzungen getestet werden (Strobl 2010, 2f.).

4.2.5.3 Einhaltung der Modellvoraussetzungen

4.2.5.3.1 Konstruktive Schritte

Die Einhaltung der Modellvoraussetzungen des RASCH-Modells erfordert im ersten Schritt Sorgfalt bei der Konstruktion des Modells, beispielsweise durch die Vermeidung von Abhängigkeiten zwischen Items. Das bedeutet insbesondere, dass keine aufeinander aufbauenden Items verwendet werden dürfen. Die Einhaltung dieser Voraussetzung stellt beispielsweise dann ein Problem dar, wenn Prozesse mit verschiedenen zwangsläufig aufeinander aufbauenden Schritten überprüft werden sollen. Mögliche Auswege bieten dann entweder die Prüfung einer Schrittreihenfolge in einem Item oder die Wahl eines anderen statistischen Modells. Hier bestünde Bedarf einer tiefergehenden Analyse. Probeweise wurden bei dem im Rahmen dieser Arbeit durchgeführten Assessment zwei Items zur Beschreibung der Schritte für die Berechnung lokaler Extrempunkte außerhalb des eigentlichen Tests hinzugefügt. Der augenscheinliche Aussagegehalt der Ergebnisse ist relativ hoch, erfordert allerdings sinnvollerweise eine nicht dichotome Bewertung, was komplexere statistische Modelle voraussetzt. Daher wurde auf die Berücksichtigung beider Items beim eigentlichen Assessment verzichtet.

Die zweite Notwendigkeit ist die Sicherstellung der Voraussetzungen während der Durchführung. Insbesondere muss verhindert werden, dass die Assessment-Teilnehmer anderen Assessment-Teilnehmern irgendwelche Hinweise zu den Lösungen der Items zukommen lassen können.

Der dritte Schritt zur Sicherstellung der Einhaltung der Modellvoraussetzungen besteht schließlich in der Überprüfung der Ergebnisse durch entsprechende statistische Tests.

4.2.5.3.2 Graphischer Test

Ein beliebter Test ist der sogenannte graphische Test, bei dem die Personen in zwei Gruppen geteilt (gesplittet) werden. Das Splitkriterium für die Bildung dieser zwei Gruppen kann frei gewählt werden. So kann ein statistischer Parameter wie der Mittelwert oder der Median ebenso verwendet werden, wie die Ausprägung eines anderen Merkmals, beispielsweise das Geschlecht der Assessment-Teilnehmer.

Nimmt man als Splitkriterium den Mittelwert, werden zwei Gruppen gebildet, deren Ergebnisse über beziehungsweise unter dem Mittelwert liegen. Anschließend werden die Aufgabenparameter (Schwierigkeitsgrade) in beiden Gruppen getrennt geschätzt und in einem Koordinatensystem, dessen Achsen den Parameterschätzungen in den beiden Gruppen entsprechen, gegeneinander aufgetragen. Sind die Voraussetzungen des Modells für ein Item erfüllt, so wird man erwarten können, dass der Parameter in beiden Gruppen gleich geschätzt wird. Er würde somit (mit steigender Schwierigkeit) immer weiter oben auf einer Winkelhalbierenden durch den Ursprung eingetragen werden. Gibt es dagegen Abweichungen, wird der entstehende Punkt mit beiden Mittelwertschätzungen des Items von dieser Linie abweichen. Nun sind immer statistische

Schwankungen zu erwarten, so dass die Winkelhalbierende nur selten getroffen wird. Als signifikant wird die Abweichung eines Items angesehen, wenn dessen parallel geschätztes zweidimensionales Konfidenzintervall, die Konfidenzumgebung, die als Ellipse um die Schätzung eingetragen wird, die Winkelhalbierende nicht berührt. Durch Entfernen solcher Items können Items eliminiert werden, die die Modellvoraussetzungen verletzen. In Abbildung 14 sind die beispielhaften Testergebnisse für zwei Splitkriterien eingetragen. Einmal wurde der Gesamtmittelwert des Assessments⁴, einmal das Geschlecht der Teilnehmer verwendet, um die Mittelwerte der einzelnen Items aufzutragen. Man sieht beispielsweise im rechten Bild an dem weit nach links verschobenen Item, dass dieses von den Frauen als schwierig, von den Männern dagegen als einfach empfunden wurde. Dies ist ein Beispiel für Differential Item Functioning.

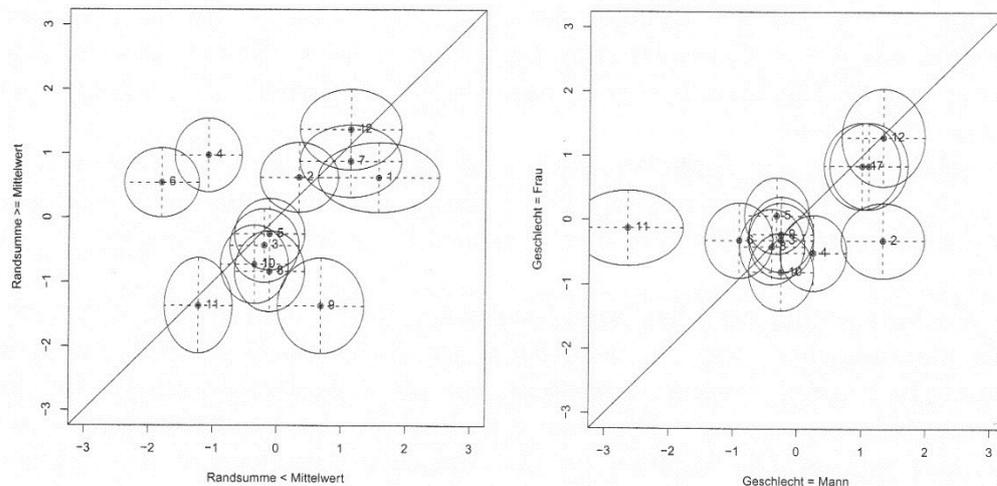


Abbildung 14 Graphischer Test für zwei Splitkriterien (Strobl 2010, 41)

4.2.5.3.3 Andersen LR-Test

Ein anderes bekanntes Verfahren, das ebenfalls auf der Aufteilung in Gruppen beruht ist der Likelihood-Quotiententest nach Andersen. Dabei werden die Personen wiederum in theoretisch beliebig viele Gruppen eingeteilt, die außerdem verschiedenen Kriterien entsprechen können. So könnten Gruppen nach Geschlecht, nach Altersklassen oder anderen Merkmalen gebildet werden. Am gängigsten ist die Einteilung nach den erreichten Ergebnissen, also ähnlich dem beschriebenen Vorgehen im graphischen Test. Anschließend werden die Schwierigkeitsgrade der Aufgaben getrennt nach den Gruppen berechnet. Die Ergebnisse werden in die Likelihood-Funktionen eingesetzt und es wird ein Quotient gebildet, der LQ-Quotient aus der Likelihood für alle Personen geteilt durch das Produkt der Likelihoods der Einzelwahrscheinlichkeiten. Im Idealfall der Unabhängigkeit sind Nenner und Zähler gleich und der Quotient 1, bei Verletzung der Annahmen des RASCH-Modells, also insbesondere der Unabhängigkeit ist er kleiner als 1.

Durch eine Transformation des Quotienten erhält man die Teststatistik $T = -2 \ln(LQ)$. Diese lässt sich mit der χ^2 -Verteilung testen. Die Nullhypothese lautet „Das Modell ist richtig, die Voraussetzungen sind erfüllt“. Sie kann nur für kleine p-Werte auf dem entsprechenden

⁴ Die Fähigkeit einer Person, ein Item zu lösen, hängt laut Modellvoraussetzung nicht von ihrer Fähigkeit ab, andere Items zu lösen.

Signifikanzniveau widerlegt werden, also beispielsweise für $p=0,05$ auf einem Signifikanzniveau von 95%.

4.2.5.3.4 Wald-Test

Der Wald-Test basiert ebenfalls auf der Bildung verschiedener Personengruppen und der Annahme, dass die geschätzten Item-Parameter sich für verschiedene Gruppen nicht signifikant unterscheiden dürfen, wenn die Modellannahmen des RASCH-Modells gelten. Als Teststatistik wird

$$W_j = \frac{(\hat{\beta}_{j,1} - \hat{\beta}_{j,2})^2}{\hat{\sigma}_{j,1}^2 + \hat{\sigma}_{j,2}^2}$$

gewählt, wobei der Zähler die quadrierte Differenz der für die beiden Gruppen geschätzten Schwierigkeitsgrade des Items j enthält. Die Quadrierung erfolgt wie bei anderen Streuungsmaßen außer zur Beseitigung des Vorzeichens auch zur besonderen Betonung großer Differenzen. Der Nenner stellt die Normierung der Abweichungen mit den jeweils wieder in beiden Gruppen geschätzten Varianzen desselben Items dar. Über eine weitere Transformation kann der Test mittels Standardnormalverteilung durchgeführt werden und sowohl sehr kleine als auch sehr große Testwerte (auf dem entsprechenden Signifikanzniveau) führen zur Ablehnung des Items im Test. (Strobl 2010, 44f.)

4.2.5.3.5 Weitere Tests

Weitere Tests sind der globale Wald-Test, der Martin-Löf-Test (auch Test auf Eindimensionalität) oder der χ^2 -Anpassungstest. Dies sind nur einige Beispiele, die durch die Stochastik bereitgestellten Testverfahren. Im Rahmen dieser Arbeit sind einige beispielhaft für das geschätzte Modell verwendet worden, auch unter dem Aspekt der Verfügbarkeit in dem R-Paket, um eine möglichst konsistente Auswertung zu ermöglichen.

4.3 Assessment-Modelle

Nachdem wichtige Teilaspekte der Kompetenzmessung auf Basis der Struktur des DFG-Projekts erläutert wurden, sollen jetzt grundlegende Aspekte und Gedanken anhand von existierenden Ansätzen für Assessment-Modelle betrachtet werden. Die Auswahl und Gliederung folgt dabei Winther, die ihre Auswahl damit begründet, dass „[...] die Instrumentenentwicklung auf elaborierten Theorien basiert und nicht allein auf die Konstruktion von Items ausgerichtet ist.“ (Winther 2010, 59). Dies korreliert mit dem hier verfolgten, in der Aufgabenstellung formulierten Ziel, ein Modell für die Kompetenzbeschreibung zu entwickeln, das mittels geeigneter Instrumente eine Messung erlaubt und eine Erklärungskomponente besitzt. Der Fokus liegt nicht auf der Entwicklung eines singulären Assessments, sondern auf der Basis für die Konstruktion und Einbettung von Assessments in ein domänenorientiertes Kompetenzmodell.

4.3.1 Assessment Triangle

Letztlich stellt jedes Assessment ein Messmodell dar, das Messkonstrukt, Messinstrument und Messvariablen wie beschrieben beinhaltet. Was man misst und wie dies zu bewerten ist, lässt sich wie in Abbildung 15 darstellen.

Das Bild des „Assessment Triangle“ wurde im Rahmen des NAC-Report „Knowing what students know“ (Pellegrino, Chudowsky und Glaser 2001) von Pellegrino und anderen entwickelt. Dabei beschreibt „cognition“ das Modell, das über das Wissen des Lernenden zugrunde gelegt wird. Es

werden also die Domäneninhalte aus der Sicht des Assessment-Designers beschrieben. „Observation“ beschreibt die Methoden, die zur Messung verwendet werden, Aufgaben, Gespräche, Arbeitsergebnisse, also die verschiedenen Messinstrumente in der Assessment-Situation.

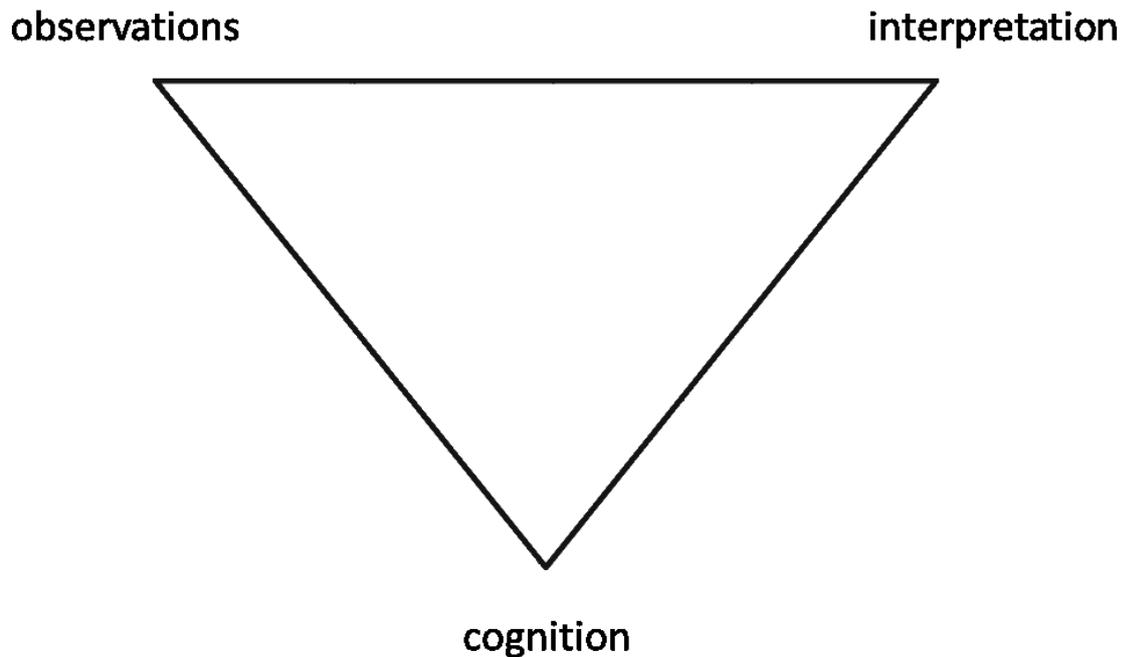


Abbildung 15 Assessment Triangle (Pellegrino, Chudowsky und Glaser 2001)

„Interpretation“ schließlich beschreibt wie die Beobachtungen auf das Modell („cognition“) zu beziehen sind. Das „Assessment Triangle“ lässt sich gut mit der Gliederung in Abschnitt 4.1.1 beziehungsweise der Darstellung in Abbildung 7 in Übereinstimmung bringen. Der Bereich „cognition“ entspricht d1, „observation“ d2 und „interpretation“ entspricht d3, also der Idee die Ergebnisse im Sinne einer Ursachenanalyse zurückzuverfolgen und geeignet zu aggregieren und zu analysieren.

4.3.2 Evidence-Centered Assessment Design (ECD)

Das ECD (Evidence-Centered Assessment Design) ist ein Assessment-Modell, das das Ergebnis eines Assessments als Basis einer „Beweisführung“ darüber ansieht, was die Ausführenden kennen oder erreicht haben (Mislevy und Haertel 2006, 7).

Das Modell geht von mehreren Stufen aus, wie in Tabelle 3 dargestellt. Sie werden in Übersetzung und in Anlehnung an Mislevy und Riconscente (Mislevy und Riconscente 2005) von Winther angegeben als (Winther 2010, 65):

- 1) Domänenanalyse
- 2) Domänenmodellierung
- 3) Konzeption der Assessment-Rahmenbedingungen
- 4) Entwicklung des Assessments
- 5) Einsatz des Assessments

ECD-Schicht	Interpretation	Merkmale	Repräsentationsformen (des Outputs, beispielhaft)
Domänenanalyse (domain analysis)	Informationssammlung der Domänenspezifika, einschließlich der Informationen über das Erlernen und Vermitteln der Domänenmerkmale	Konzepte, Terminologien, Methoden und spezifische Repräsentationsformen	Formeln, Symbole, Mindmaps, Lehrpläne, technische Modelle
Domänenmodellierung (domain modeling)	Narrative Struktur der Domänenspezifika (zentrale Ideen, typische Arbeits- und Lernprozesse)	Wissen, Fertigkeiten, Fähigkeiten, mögliche Arbeitsprodukte oder Beobachtungsergebnisse	Modell der Domäne (Beziehungen, Argumentationen, Bezugspunkte)
Konzeption des Assessment-Modells (conceptual assessment framework)	Modellformulierung zur Konstruktion von Assessments (Blue Print)	Latente Variablen, mögliche Messmodelle sowie Spezifikationen des Testaufbaus	Mathematische und/oder graphische Repräsentation des Testmodells
Entwicklung des Assessments (assessment implementation)	Erstellung des Assessments (Itemkonstruktion)	Aufgabenmaterial (einschließlich aller notwendigen Informationen zur Bewältigung)	Itembeschreibungen, Präsentationsregeln, Kodierungen, Scoringverfahren
Einsatz des Assessments (assessment delivery)	Operativer Assessmentdurchlauf und Reporting	Aufgaben, Arbeitsprodukte und das entsprechende Antwortverhalten, Testscores	Numerische und graphische Zusammenfassung des Assessments, Testgüte, Reports auf unterschiedlichen Analyseebenen

Tabelle 3 Stufen des ECD-Modells (nach (Winther 2010, 65))

4.3.2.1 Domänenanalyse

Bei der Domänenanalyse im Sinne des ECD geht es um das Herauskristallisieren der domänen-spezifischen Anforderungen. Zu beachten ist dabei, dass diese im Hinblick auf

- 1) den psychologisch determinierten Lernzugang und
- 2) den fachlich determinierten Lernzugang.

voneinander unterscheiden können (Winther 2010, 66f.).

So ist psychologisch beispielsweise aus kognitiver Perspektive die Frage relevant, wie mathematisches Wissen erworben werden kann und wie auf Wissensrepräsentationen zugegriffen wird, während aus soziokultureller Perspektive von Interesse ist, wie beispielsweise mathematische Fähigkeiten zur aktiven gesellschaftlichen Teilnahme in Abhängigkeit der jeweiligen Gruppe beitragen können.

Fachlich ist beispielsweise die Handelskalkulation aus Sicht eines Mathematiklehrers eine ökonomische Anwendung der Prozentrechnung mit vermehrtem und vermindertem Grundwert, während sie aus Sicht des Ökonomelehrers eine Methode zur Preisbestimmung ist, die sich des Mittels der Prozentrechnung bedient.

4.3.2.2 Domänenmodellierung

Die Domänenmodellierung umfasst die inhaltliche Strukturierung der in der Analyse gefundenen Informationen. In Anlehnung an Toulmin (Toulmin 1958) werden von Winther drei Bausteine genannt (Winther 2010, 67f.):

Baustein 1: Anspruch des Assessments (claim)

Der Anspruch drückt sich aus in der Formulierung

- des inhaltlichen Wissens, der Fähigkeiten und Fertigkeiten einer Domäne und
- des für die Bewältigung des Assessments relevanten Wissens, der relevanten Fähigkeiten und Fertigkeiten (beispielsweise Heuristiken zur Problemlösung).

Baustein 2: Output des Assessments (ground oder data)

Die Daten ergeben sich beispielsweise aus den Lösungen der Aufgaben, aus Beobachtungen während des Assessments oder anderen spezifischen Informationen während der Durchführung. Diese Daten können das Ergebnis manifester oder latenter Variablen sein.

Baustein 3: Modell über den Zusammenhang (warrant)

Der Bezug zwischen dem Anspruch des Assessments und den ermittelten Daten beruht auf theoretischen Vorstellungen, einem Modell. Man geht davon aus, dass der Zusammenhang zwischen der konstruierten Situation des Assessments und den erzeugten Daten analysierbar ist und so Rückschlüsse von den Daten auf die Erfüllung des Anspruchs gezogen werden können.

4.3.2.3 Konzeption des Assessment-Modells

Das Design des eigentlichen Assessments erfolgt durch drei Modelle, das „Student Model“, das „Task Model“ und das „Evidence Model“, das die beiden erstgenannten Modelle verbindet. Das „Student Model“ dient der Operationalisierung des Assessment-Anspruchs (claim). Dabei müssen nicht nur die entsprechenden Elemente definiert werden, sondern auch die Beziehung zwischen der theoretischen Kompetenz θ und den gemessenen realen Werte $\theta \pm x$, wobei x andere Komponenten des Assessments beschreibt, beispielsweise Missverständnisse oder falsches Verständnis der Aufgaben, Ablenkungen im Prüfungsverfahren, Hinweise des Lehrers, Übernahme fremder Gedanken, was über die Schätzfehlerschätzungen der entsprechenden Verfahren tendenziell realisiert werden kann.

Das „Task Model“ beschreibt die Rahmenbedingungen des Assessments. Idealerweise geht man in der beruflichen Bildung von möglichst authentischen Assessments als Arrangements, die beruflichen Prozessen ähneln, aus. Dies hat in der Praxis allerdings Grenzen, so dass das Material und die Vorgehensweise angenähert werden müssen. In jedem Fall sollte versucht werden Handlungssituationen vorzugeben, um die Handlung „beobachten“ zu können. In theoretischen eher fachbezogenen Unterrichtsformen wird man auf anwendungsorientierte Aufgaben zurückgreifen müssen. Zum „Task Model“ gehören auch Angaben beispielsweise zur Bearbeitungszeit, zu den Hilfsmitteln wie dem Taschenrechner oder weiteren Materialien.

Das „Evidence Model“ verbindet die beiden genannten Modelle und besteht selbst aus zwei Modellen, dem „Statistic Model“ und den „Evidence Rules“. Im ersten Fall handelt es sich um ein konkretes statistisches Modell. Bei ihm werden zumeist psychometrische Modelle (IRT) oder

seltener Modelle der klassischen Testtheorie (CTT) verwendet. Bei den „Evidence Rules“ geht es um das Regelsystem, das der Evaluation der Assessmentergebnisse dient. Dabei betrachtet man die beobachtbaren (manifesten) Variablen und deren Zuordnung zu den Anforderungen insbesondere in Form der Kodierungs- und Scoringregeln.

4.3.2.4 *Entwicklung des Assessments (Assessment implementation)*

Hier ist beim ECD die Implementierung der bisher getroffenen theoretischen Vorbereitungen vorgesehen. Dies betrifft insbesondere die Item-Entwicklung.

4.3.2.5 *Einsatz des Assessments (Assessment delivery)*

Der Einsatz umfasst die Durchführung des Assessments und die Ermittlung der Antworten für die Items. Die Ergebnisse werden mittels des Scoring-Modells ermittelt und dann mittels des statistischen Modells akkumuliert und den Anforderungen gegenübergestellt.

4.3.3 Die Four Building Blocks

Einen anderen Ansatz für die Entwicklung von Assessments findet man bei Wilson, der die „Four Building Blocks“ einführt (Wilson 2004b). Wilson postuliert, dass die Entwicklung der Assessment-Items nicht willkürlich erfolgen darf, sondern sich an möglichst fundierten Theorien ausrichten muss. Bei der Auswahl soll der Fokus nicht allein auf Ansätze ausgerichtet werden, die sich nur auf die Konstruktion von Items beziehen, sondern einem breiteren theoretisch fundierten Fundament folgen. Wilson beschreibt seinen breiteren Assessment-Ansatz selbst mit den Worten: „In this volume, the word instrument is defined as a technique of relating something we observe in the real world (sometimes called manifest or observed) to something we are measuring that only exists as part of a theory (sometimes called latent or unobserved). This is somewhat broader than the typical usage, which focuses on the most concrete manifestation of the instrument – the items or questions.“ (Wilson 2004b, 3f.).

Man erkennt aus den vorangegangenen Darstellungen, dass diese Beschreibung letztlich die verschiedenen Anforderungen und verschiedenen Vorstellungen von sinnvollen Assessments und sinnvollen Beurteilungen grundsätzlich umfasst. „Letztlich ist in ein Assessmentmodell eine Vorstellung darüber zu implementieren, wie sich die über die Testformate gesammelten Daten nutzen lassen.“ (Winther 2010, 63). Dies entspricht auch dem Interpretationsgedanken des „Assessment Triangles“. Der Zusammenhang zwischen beobachtbarer manifester Variable und nicht direkt beobachtbarer latenter Variable entspricht dem Funktionalitätsgedanken der Kompetenzen im engeren Sinne. Die „Four Building Blocks“ sind in Abbildung 16 wiedergegeben

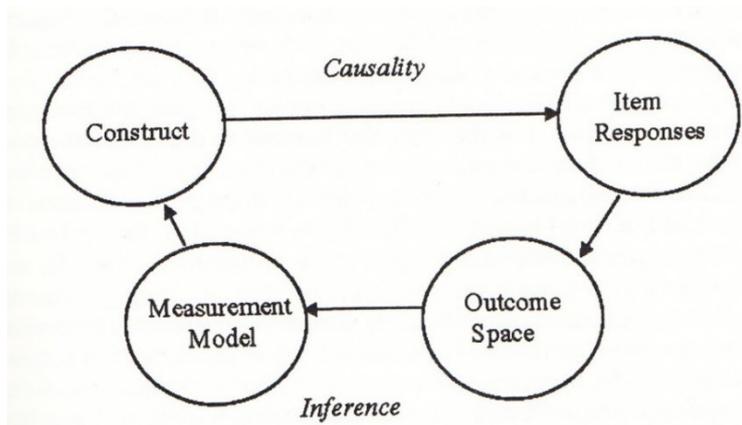


Abbildung 16 Four Building Blocks (Wilson 2004b, 17)

Wesentlich ist, dass jeder Block zunächst einen Bedarf repräsentiert und in einem bestimmten Kontext steht. Er soll eine Aufgabe, einen Zweck erfüllen. Erst wenn dieser klar ist, kann das Instrument zur Erreichung dieses Zwecks definiert werden, das letztlich das Ergebnis den Aufgaben des Blocks entsprechend liefert. Jeder der vier Blöcke wird also mit bestimmten Instrumenten genutzt und produziert ein bestimmtes Ergebnis.

Der erste Block „Construct“ dient der Beschreibung des Messkonstrukts. Dabei geht es um das Verständnis einer Person von einer bestimmten Domäne oder ihrer Haltung zu einer solchen. Ein Beispiel aus der Chemie ist: „The focus is on the domain knowledge they have acquired during instructional interactions in terms of how the students are able to think and reason with chemistry concepts.“ (Wilson 2004b, 7). Das Messinstrument ist bei Wilson die „Construct Map“.

Der zweite Block „Item Responses“ dient dem Item Design, also dem Entwurf der Items. Zur Gewinnung der Beobachtungen benötigt man eine konkrete Vorgehensweise. Dabei steht im Vordergrund, dass basierend auf dem Messkonstrukt Annahmen getroffen werden, dass bestimmte Antworten auf Items erst auf bestimmten Kompetenzniveaus möglich sind, wie es bereits im Rahmen der Kompetenzniveaumodelle in Abschnitt 4.2.2.2 beschrieben wurde. Es wird also wieder ein Kompetenzniveaumodell eingeführt. Es werden Annahmen über kausale Beziehungen (causality, cause) zwischen den Antworten und dem zu messenden Konstrukt getroffen. Später werden aufgrund der tatsächlichen Antworten Schlussfolgerungen (inference, infer) bezüglich des Konstrukts getroffen. Die Entwicklung der Items muss die kausalen Beziehungen und die Schlussfolgerungen aus den Antworten so nahe wie möglich zueinander bringen. Das Ergebnis des Blocks sind konkrete Items und zugehörige Antwortalternativen.

Der dritte Block ist der „Outcome Space“, der eine strukturierte und gemessene (scored) Darstellung der tatsächlichen Beobachtungen ist, also eine Bewertung. Hier kann es gegenüber dem geplanten Ablauf und den geplanten Antworten Abweichungen geben.

Der vierte Block schließlich ist das „Measurement Model“. „The measurement model must help us to understand and evaluate the scores [...]“ (Wilson 2004b, 16), es dient also der Auswertung und Beurteilung der Ergebnisse.

4.4 Einordnung in den Bildungsprozess

Die bisherigen Beschreibungen sollen genutzt werden, um die in dieser Arbeit bereits zu Beginn gegebene schematische Darstellung des Bildungsprozesses (Abbildung 1) um eine detailliertere Sicht aus dem Blickwinkel von Assessment und Feedback zu ergänzen. Diese detailliertere Sichtweise ist in Abbildung 17 dargestellt und gibt die Vorstellung von Beobachtung, Bewertung und Beurteilung wieder.

Assessments haben wie beschrieben die Aufgabe einer Beobachtung und letztlich auch Bewertung der einzelnen Lernenden. Dabei können wertvolle Informationen für alle Beteiligten gewonnen werden, für die Lernenden, die Lehrenden und andere Stakeholder des Prozesses.

Je detaillierter diese Informationen sind und je besser deren Zustandekommen durch das Assessment selbst und dessen modelltechnische Grundlagen erklärt werden kann, umso besser können diese Informationen nicht nur zu einer Bewertung, sondern auch zu einer besseren Justierung der verfügbaren Kompetenzen führen.

Daher ist es wichtig, eine Erklärungskomponente in das Modell einzubeziehen, wie es besonders das ECD betont, um Ursachenanalysen zu ermöglichen und allen Beteiligten durch gezieltes Feedback die Möglichkeit einer gezielten Nacharbeit und einer Schließung der Kompetenzlücken zu geben.

- 1) Grundlage ist der Bildungsprozess, bei dem sich die Lernenden in einem Lernprozess mit dem Domäneninhalt vertraut machen und der Erfolg in Assessments ermittelt wird.
- 2) Wesentliche Einflussfaktoren sind unter anderem beispielsweise die Lehrerqualifikation und die organisatorischen Rahmenbedingungen.
- 3) Viele Beurteilungsprozesse erfolgen auf Basis von Assessments und sonstigen Beobachtungen und Prozessen. Die Transparenz leidet aber, wenn die Ziele und Domäneninhalte nur implizit vorhanden sind. Es bleibt ein großer Spielraum in der Beurteilung des Lernerfolgs insbesondere im Hinblick auf die objektiv erwünschten Ziele.
- 4) Demgegenüber soll hier die gesamte Beurteilung des Erfolgs des Bildungsprozesses ausgehend von den anzustrebenden Kompetenzen (und Qualifikationen) erfolgen.
- 5) Basis für die Definition dieser anzustrebenden Ziele sind Schulgesetze aber auch nachgeordnete Vorgaben wie Bildungsstandards sowie gesellschaftliche Erwartungen.
- 6) Aus den Zielen des Bildungsprozesses wird die Gestaltung des Assessments durch Operationalisierung angestrebt.
- 7) Die Gestaltung des Assessments bedeutet die detaillierte Beschreibung des Domäneninhalts und die Ableitung des Instrumentes und der zu messenden Variablen aus dem Domäneninhalt.
- 8) Auf dieser Basis erfolgt die Assessmentdurchführung.
- 9) Das Ergebnis der Durchführung wird zunächst in Form einer unbewerteten Beobachtung festgehalten. Dabei kann eine klare Zuordnung zu den definierten Variablen und somit dem strukturierten Domäneninhalt erfolgen(a).
- 10) Die Beobachtungen können in einem weiteren Schritt durch Zuordnung von quantitativen Werten in eine Bewertung umgesetzt werden.
- 11) Die Bewertung kann durch Ergänzung mit stochastischen Verfahren zu einer Kompetenzmessung erweitert werden.
- 12) Analog zu den Kompetenzen kann dies für Qualifikationen geschehen. Insgesamt kann durch die Orientierung an den aus den Zielen abgeleiteten Variablen so eine Auswertung erfolgen, die sich an den Zielen orientiert (b).
- 13) Schließlich kann (außerhalb des Modells) aus den so ermittelten Ergebnissen und zusätzlichen auch klassisch gewonnenen Erkenntnissen eine Gesamtbeurteilung erstellt werden. Diese ist in erheblichem Maß wiederum an den Zielen orientiert (c).
- 14) Zusätzlich können die Beobachtungen ohne Bewertung für ein unmittelbar an dem Domäneninhalt orientiertes Feedback genutzt werden.
- 15) Dies kann basierend auf den Ergebnissen individuell für den einzelnen Lernenden erfolgen, wobei individuelle Stärken und Schwächen unmittelbar auf den Domäneninhalt bezogen werden können.
- 16) Durch Kopplung mit den Inhalten des Bildungsprozesses und ergänzenden Hilfen kann den Lernenden Unterstützung für den weiteren Bildungsprozess gegeben werden.
- 17) Der andere Aspekt des Feedbacks ist der des Gruppenfeedbacks bei dem die Gesamtsituation und das Gesamtergebnis des Bildungsprozesses dem Lehrenden wesentliche Hinweise auf die Gestaltung des Bildungsprozesses geben können.

Ausgegangen wird also von den Zielen (Kompetenzen, Qualifikationen, berufliche Situation, Leitidee Output-Orientierung) des Bildungsprozesses. Dies bedeutet, dass die relevanten

Bildungsstandards mit ihren Zielvorgaben auch in Form von Kompetenzen dokumentiert werden. Sie müssen dafür analysiert werden und daraus die in einer Domäne zu ermittelnden Kompetenzen mit ihren Dimensionen beschrieben werden. Dadurch entsteht durch Operationalisierung eine Fähigkeitsbeschreibung, das Kompetenzmodell, das aus dem Blickwinkel eines Assessments das Messkonstrukt beschreibt. Dabei handelt es sich um ein Kompetenzstruktur- und Kompetenzniveaumodell.

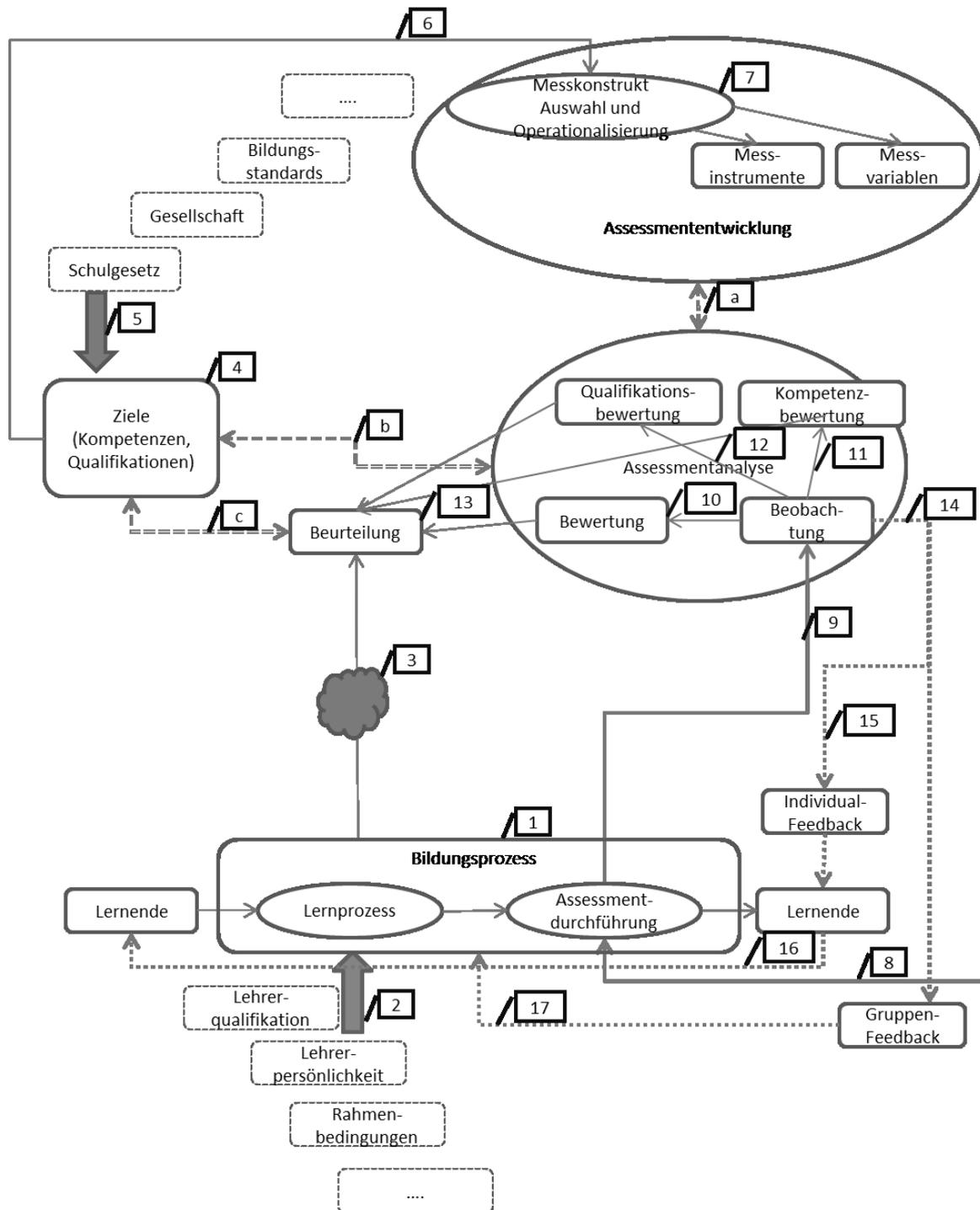


Abbildung 17 Beobachtungs-, Bewertungs-, und Beurteilungsprozess

Dieses Messkonstrukt generiert (causes) wie bei Wilson beschrieben die Messvariablen, sowohl manifeste als auch latente Variablen. Die Messinstrumente bestehen beispielsweise aus klassischen paper-and-pencil-Klausuren oder e-Assessments mit deterministischen Bewertungen (Scoring) oder statistischen Modellen. Diese Inhalte können bis zu Anforderungen an konkrete Assessment-Items operationalisiert werden. Durch die Hinzunahme von Operatoren zur Beschreibung der Fähigkeiten können diese beobachtbar gemacht und so als die Basis für eine Leistungsmessung instrumentalisiert werden.

Beim Messinstrument können durchaus unterschiedliche Vorgehensweisen genutzt werden, die von den formulierten Zielen abhängig sind. So zeigt Winther, dass die Akzentuierung der Modellierung für allgemeinbildende Fächer und für die berufliche Pädagogik unterschiedlich sein kann. Im ersten Fall steht der Inhalt am Anfang und dann folgt die Einbettung in Aufgabenstellungen und die Konstruktion von Situationen. Im beruflichen Umfeld kann dies anders sein. Hier kann die Situation, der Geschäftsvorfall oder die übergreifende Idee eines Lernbereiches am Anfang stehen, aus denen dann die Aufgaben abgeleitet werden, was letztlich zum Inhalt führt (Winther 2010, 51).

Definition 7: Messinstrument

Das Messinstrument ist die Form des Assessments, beispielsweise eine klassische Klausur oder wie hier bei der Modellvalidierung verwendet eine e-Klausur (e-Assessment).

Definition 8: Messvariable

Die Messvariablen stellen die Größen dar, die für die klassische Bewertung und Kompetenzbestimmung gemessen werden. Dabei wird zwischen manifesten und latenten Variablen unterschieden. Während sich manifeste Variablen direkt beobachten und bestimmen lassen, handelt es sich bei latenten Variablen um solche, die nicht direkt beobachtet werden können, die aber dem Modell zugrunde gelegt sind und für die die beobachteten Variablen nur als Indikatoren dienen. Die Messvariablen werden im Rahmen der Kompetenzbewertung analysiert.

Bei einem klassischen Assessment können die manifesten Variablen Punktzahlen von Aufgaben, Teilaufgaben oder der Gesamtklausur sein. Sollen Einzelaussagen über bestimmte Aspekte des Messkonstrukts getroffen werden, können dies im Rahmen der Assessment-Analyse die Bewertungen für eine bestimmte Aufgabe, für die Sprachbeherrschung, die Rechtschreibung oder andere Größen sein.

Im Rahmen der Assessment-Durchführung erfolgt durch Beobachtung der Personen und Dokumentation der Ergebnisse die Dokumentation des „Outcoming Space“ in Form von Antworttypen, die die Basis für die Ermittlung der individuellen Messwerte sind (Beobachtung).

Die beobachteten Ergebnisse können unter vier verschiedenen Aspekten genutzt werden.

- Der Ermittlung der Zielerreichungsgrade bezogen auf die manifesten Variablen und damit letztlich auf die Qualifikation. (Bewertung, Scoring);
- einer weitergehende statistische Analyse auf Basis der manifesten Variablen zur Kompetenzbewertung der latenten Variablen(Kompetenzbewertung);
- einer Qualifikationsbewertung auf Basis der so ermittelten Ergebnisse;

- der Erstellung eines individuellen und detaillierten Feedbacks bezüglich des Modellkonstruktes für die Assessment-Teilnehmer. (Individualfeedback).

Die Ergebnisse der Bewertung lassen sich letztlich auch in ähnlicher Weise wie die Bewertung der klassischen Assessments zur Beurteilung der Leistung heranziehen. Sie können aber auch in Bezug zu den gesamten modellierten Kompetenzen des Modellkonstrukts oder eines Qualifikationsstandards verwendet werden. Damit sind auch andere Anspruchsgruppen im Sinne der Zubringerfunktion direkt eingebunden.

Mit der Beurteilung wird das Kompetenzmodell verlassen. Die Beurteilung des Lernenden erfolgt auf Basis der Assessments sowie anderer, den gesamten Bildungsprozess umfassenden, Beobachtungen der für die Beurteilung verantwortlichen Person. Dabei können andere individuelle Eigenschaften und Entwicklungen unter Würdigung pädagogischer Aspekte eingebracht werden.

Die detaillierten Feedbacks müssen nicht mit einer Bewertung oder gar Beurteilung verbunden sein. Hier kommt es entscheidend auf den Detaillierungsgrad des Feedbacks in Form der Beobachtung und der daraus ableitbaren Maßnahmen an. Dies ist für formative Assessments sinnvoll. Mit den so gewonnenen Informationen können die Lernenden gegebenenfalls Teile des Bildungsprozesses mit spezifischen Materialien ergänzen oder nacharbeiten. Für die Lehrenden ergibt sich ein Feedback, das Defizite im Prozess für die Gruppe aufzeigt und ebenfalls zielgerichtete Nachjustierungen des Bildungsprozesses erlaubt. Man sieht, dass die Ergebnisse sowohl für summative als auch formative Zwecke nutzbar sind.

4.5 Pädagogische Probleme

Im Folgenden soll auf mögliche andere Perspektiven als bisher beschrieben eingegangen werden.

4.5.1 Bildungsideal und qualitative Ansätze

Im Zusammenhang mit dem Kompetenzbegriff stellt sich auch die Frage nach der Beziehung zum klassischen Bildungsideal. Die wesentliche Abgrenzung besteht darin, dass unter Bildung – oder besser unter dem humanistischen Bildungsideal – zumeist die gezielte Wertevermittlung, insbesondere Menschlichkeit, Humanität, allseitig entfaltete Persönlichkeit und kritische Selbstständigkeit verstanden wird, während Kompetenz sich hinsichtlich der Wertung neutral verhält (Gnahs 2007, 22f.).

Dies bedeutet nicht, dass Kompetenz nicht auch Werte enthalten kann. Sie werden aber, wie beschrieben, als Teil der Kompetenz aufgefasst, die neben Wissen, Fertigkeiten, Motivationen und sensomotorischen Grundlagen stehen oder als außerhalb davon stehende grundlegende Eigenschaften. Die Werte werden als Komponente oder Dimension der Kompetenz berücksichtigt, es wird aber nicht festgelegt, welche konkreten Ausprägungen dieser Werte erwünscht sind (ebd.). Kompetenz verhält sich hier also positiv ausgedrückt neutral, negativ ausgedrückt gleichgültig.

Diese Neutralität führt zu Kritik. Kompetenzen im Sinne der drei von Roth genannten Aspekte Sach-, Selbst- und sozialer Kompetenz sowie verwandter Strukturierungen decken einen breiten Raum der Diskussion von Bildungsprozessen ab. Trotz der weiten Verbreitung gab und gibt es die Kritik eines Fehlens des normativen „Bedeutungsüberschusses“ im Sinne des akademischen Bildungsbegriffs einerseits und der „Nutzung“ von Kompetenz im Sinne einer persönlichen Bewertung von Personen andererseits. Zudem gibt es Ansätze, die Entwicklung der Persönlichkeit

und ihrer Identität vom Abschneiden in Leistungsmessungen abzukoppeln (Klieme und Hartig 2007, 21).

Letztlich zielen viele Kritikpunkte gerade auf den Unterschied zwischen dem umfassenderen Begriff der Individualkompetenz und dem auf Messbarkeit ausgerichteten Kompetenzbegriff im engeren Sinne. So greift Beyen aus Sicht eines Praktikers in Anlehnung an Cranach/Bangerter statt auf den Kompetenzbegriff auf einen allgemeineren Handlungsbegriff zurück, auf ein „[...] motiviertes, gezieltes, geplantes, gewolltes, kontrolliertes und bewertetes Verhalten.“ (Beyen 2013, 283) und lehnt eine Operationalisierung ab. Stattdessen werden andere Konzepte als Zielvorstellung, zum Beispiel das Identitätskonzept nach Hausser, als Orientierungsraster im Bildungsprozess vorgeschlagen. (Beyen 2013, 285)

Das Identitätskonzept von Hausser (Hausser 1995) geht insofern einen anderen Weg, als dass es die Identität des Menschen in den Mittelpunkt stellt. Die Identität wird als Innenperspektive eines Menschen aufgefasst, die nur von ihm selbst erforscht werden kann. Die Domänen des Lernens sind bevorzugte Bereiche dieser Identitätsbildung, da sie dem Individuum Gelegenheit zum Ausdruck seiner Persönlichkeit bieten.

Generell wird bei derartigen Ansätzen oftmals die Messbarkeit der dann als Kompetenzen aufgefassten Eigenschaften angezweifelt. Bereiche wie Werte oder Motivationen scheinen sich in einem derartigen Verständnis einer Messbarkeit stark zu widersetzen, da sie eher intrapersonell und daher schlecht beobachtbar sind. Dabei darf aber nicht übersehen werden, dass intrapersonelle Aspekte auch ein „In-sich-selbst-Hineinsehen“, eine Introspektion und entsprechende interpersonelle Bewertung zumindest grundsätzliche Messbarkeitsmöglichkeiten eröffnen (Beyen 2013, 285). Diese Methoden sind gut zur Gewinnung individueller Aussagen geeignet, jedoch nicht zur Verallgemeinerung und Erhebung allgemeiner Fakten (Oldenburg 2004). Bei einem derartigen Ansatz stehen letztlich neben der Person auch der Prozess und dessen Gestaltung im Vordergrund.

Im Sinne der hier verfolgten Gliederung der Zielsetzungen von Bildungsprozessen, wie sie in Abbildung 7 dargestellt ist, werden die Individualkompetenzen oder sogar das gesamte Individuum in diese Zielsetzung einbezogen.

Grundsätzlich wird auch bei der Beurteilung des Erfolgs von Bildungsprozessen zwischen qualitativen und quantitativen Methoden unterschieden. Eine grundlegende Kritik qualitativer Methoden im Umfeld von Kompetenzmessungen besteht darin, „[...] dass auch und gerade qualitativ arbeitende Sozialwissenschaftler ihre Daten weder als Messung noch als Diagnose von Dispositionsausprägungen interpretieren würden“ (Klieme und Hartig 2007, 23). Stattdessen werden idealisierende Vorstellungen von vollständigen beruflichen Handlungen und selbstorganisierten Prozessen oder eher betriebswirtschaftlich motivierte „Kompetenzmodelle“ als Leitbild für die Gestaltung des Unterrichtsprozesses und die Einführung empirisch nicht überprüfbarer Kompetenzdimensionen postuliert. Dies hängt wieder eng mit dem betrachteten Kompetenzbegriff zusammen. Man kann durchaus der Meinung sein, dass dieser als Kompetenz im engeren Sinne zu stark auf die Messbarkeit und kognitive Fähigkeiten reduziert und in die Nähe von Qualifikationen gerückt wird.

So stehen Ansätzen, die die zu messenden Aspekte stark einschränken, aber statistisch nachvollziehbar sind, quasi „am anderen Ende“ pädagogische Ansätze gegenüber, die mit eher qualitativen Methoden sich an klassischen Bildungs-, Sinn- und Identitätsmotiven orientieren, die

aber den Untersuchungsbereich so umfangreich definieren, dass sie am Ende qualitative, aber unsystematische und empirisch nicht überprüfbare Ergebnisse erbringen (Jahn 2014, 4). Damit gerät der Bildungsprozess zumindest in die Gefahr des Prinzips „Hoffnung“.

Wenn auch in dieser Arbeit im Rahmen der Assessments daher mehr quantitativen Methoden gefolgt und im Sinne der obigen Forderung von Klieme und Hartig der Schwerpunkt auf quantitative Kompetenzmodelle gelegt wird, so stellt dies nicht unbedingt einen Widerspruch zum qualitativen Ansatz dar. Letztlich ergänzen sich beide Ansätze immer im Sinne einer Rückkopplung von kreativen Ansätzen und deren Überprüfung. Qualitative und quantitative Beobachtungen ergeben unterschiedliche Perspektiven, die kombiniert werden müssen (Perels 2010, 270).

4.5.2 Input- und Prozessorientierung

Daneben gibt es andere praxisorientierte Ansätze, die traditionell eher input- oder prozessorientiert sind. Auch wenn diese im Rahmen der Output-Orientierung zumindest im curricularen Überbau etwas in den Hintergrund treten, heißt dies nicht, dass sie komplett verschwinden oder künftig nicht auch gegenläufige Bewegungen denkbar sind.

So werden immer wieder neue Ansätze wie beispielsweise das „Design Based Research“ entstehen. Interessant ist hier in diesem Zusammenhang die Aussage: „Aber auch die große Menge an Daten, die im Design-Based-Ansatz gesammelt werden, stellt neben der genannten Interpretation schon in der Organisation und Auswertung hohe Anforderungen.“ (Jahn 2014, 13). Auch hier könnte also eine strukturierte Ablage und datentechnische Kopplung zwischen Evaluation und Prozess möglicherweise unterstützend wirken. Somit stellt sich sicher die Frage nach dem leistbaren Zeitaufwand, gleichwohl soll dieser nicht allein im Vordergrund stehen. Vielmehr dürfen Prozess-Orientierung und Output-Orientierung nicht als Widerspruch gesehen werden. Eine gute Assessment-Methodik ersetzt keinen guten Bildungsprozess, ein guter Bildungsprozess kann aber niemals ohne gute Assessment-Technik evaluiert werden.

Es geht hier nicht nur um eine ausschließlich output-orientierte, möglichst automatisierte Testtechnik, sondern generell um die strukturierte Dokumentation aller Zusammenhänge sowohl der Inhalte der Domäne als auch der damit zusammenhängenden Assessments und der Ergebnisse der einzelnen Personen. Die Nachvollziehbarkeit und Analysemöglichkeit für Transparenz und Feedback sind sowohl für den Prozess als auch für dessen Ergebnis wesentlich.

4.5.3 Realitätsnähe

Ein anderer Kritikpunkt an einem an Messbarkeit orientierten Vorgehen ergibt sich aus der Realitätsnähe der Assessments. Quantitative Erfassungen können in der beruflichen Bildung grundsätzlich auf einer externen Messung im Sinne der Performanz bei der Arbeit mit möglichst realen Situationen oder unter internen Bedingungen, also in Form von Prüfungen (Tests) innerhalb des Bildungssystems mit simulierten Situationen, erfolgen.

Eine externe Erfassung bringt die Problematik mit sich, dass ein valider, reliabler und objektiver Test bestenfalls für eine kleine Anzahl von Personen und zeitlich begrenzt möglich ist. Hinzu kommt das Problem der Einordnung der Bedeutung der beobachteten Arbeitsschritte. Letztlich lassen sich im Wesentlichen nur wenige Kompetenzen beobachten. Somit ist ein solches Verfahren problematisch (Achtenhagen und Baethge 2007, 61f.).

Es bleibt also die Möglichkeit der Messung unter internen Bedingungen in Form von Prüfungen, deren Aufgabenstellung entsprechend komplex zu gestalten ist, um aus den Ergebnissen entsprechende Schlüsse auf die Kompetenzen ziehen zu können. (ebd., 62) Daran kann wiederum bemängelt werden, dass die für eine quantitative Messung notwendige Standardisierung und die Isolierung einzelner Variablen in derartigen Assessments eher zu künstlich-konstruierten Laborbedingungen führen, die den in der Praxis anzutreffenden von vielfältigen variablen Einflüssen geprägten Situationen nicht gerecht werden (Jahn 2014, 4).

Auch wenn sich dieser Widerspruch nicht vollständig auflösen lässt, bedarf es doch einer Methodik, die einen Zusammenhang zwischen den intendierten Kompetenzen und dem Assessment-Design dokumentiert und inhaltliche Erklärungshilfen zu den ermittelten quantitativen Größen liefert, um eine möglichst gute Kompensation der Nachteile von Assessments unter internen Bedingungen zu erreichen.

4.5.4 Ansätze der Leistungsbewertung

Neben den klassischen Klausuren (paper-and-pencil-Assessments) oder elektronischen Klausuren (e-Assessments) werden in der pädagogischen Literatur eine ganze Reihe weiterer Messinstrumente besonders im Hinblick auf die Kompetenzermittlung genannt. Beispielhaft sei hier auf die Aufzählung von Meyer (Meyer 2004, 119) verwiesen:

- Lernentwicklungsbericht: Information für den Schüler über seinen aktuellen Stand, seine individuellen Stärken, seine Defizite sowie über die erzielten Fortschritte,
- Bewertungsgespräche: Information für den Schüler und die Eltern über den aktuellen Stand und Absprache der Ziele sowie der nächsten Schritte,
- Beobachtungsbogen: Information über die Entwicklung mit den Kompetenzen und Defiziten des einzelnen Schülers,
- Portfolio: Zusammenfassende Information über die Leistung eines Schülers (beispielsweise mit Assessmentergebnissen, Texten, Bildern oder anderen Handlungsprodukten).

Hinzu kommen die im Zusammenhang mit Begriffen wie „Neue Lernkultur“ genannten Ansätze. Hier geht es in den Grundzügen darum, dass „[...] die Leistungsbewertung in ein engeres Wechselverhältnis zum Lernprozess [...]“ (Winter 2014, 3) gebracht werden soll. Es geht also darum, Lernprozess und Leistungsbewertung miteinander zu verzahnen. Die weitere Diskussion dieser Ansätze soll nicht Bestandteil dieser Arbeit sein. Grundsätzlich wird hier der Ansatz verfolgt, die Ergebnisse des Unterrichtsprozesses in Form gewonnener Kompetenzen als Grundlage der Assessments zu verwenden. Daher erscheint im Rahmen der Zielsetzung sinnvoll, eine an objektiven Kriterien der Lernenden festgemachte Leistungsbeurteilung zu ermöglichen, um gerade auch Qualifikationen beurteilen zu können. Eine Beurteilung eines Prozesses dagegen ist per se immer von der Gestaltung des Prozesses abhängig. Da gerade dieser einer großen pädagogischen Freiheit unterliegt, wird man letzten Endes schwer vergleichbare Ergebnisse erhalten.

Allerdings muss ein Messverfahren, das mit vertretbarem Aufwand und einer „gewissen“ Objektivität allgemeine Ergebnisse liefern soll, letztlich auf empirischen, beobachtbaren und messbaren Ergebnissen beruhen. Interviews, prozessbegleitende Beobachtungen, subjektive Bewertungen und ähnliche Methoden der qualitativen Bewertung können ergänzende Informationen oder wertvolle Hypothesen und individuelle Einblicke liefern. Sie werden hier aber nicht

weiter betrachtet. Außerdem soll aus den geschilderten Gründen konsequent das Prinzip der Trennung von Prozess und Messung verfolgt werden.

Soll auf der anderen Seite der Skala die Kompetenzmessung nicht zu einem reinen Zählen richtiger und falscher Antworten werden und sollen Kompetenzen empirisch bestimmt werden können, so muss eine Fülle empirischer Einzelbeobachtungen zu einer Kompetenzabschätzung zusammengefasst werden (Klieme und Hartig 2007, 24)

5 Ansatz

Nach den inhaltlichen Begriffsklärungen kann jetzt eine Formulierung des gewählten Ansatzes vorgenommen werden. Diese Arbeit verfolgt wie eingangs geschildert, die Absicht, die Ziele des Bildungsprozesses inhaltlich so zu beschreiben, dass sie sich als Grundlage für Assessments eignen, die mit statistischen und deterministischen Methoden ausgewertet werden können und eine Basis für ein inhaltliches Feedback bilden.

5.1 Architektur

Für die Formulierung des Ansatzes wurde in Abschnitt 4.4 aufgezeigt, wie sich das Modell in den Bildungsprozess einfügen soll. Bezogen auf die Architektur des DFG-Schwerpunktprogramms 1293 wie in Abschnitt 4.2.3 beschrieben und in Abbildung 12 dargestellt sollen alle Ebenen adressiert werden, wobei der Fokus auf der Ebene 1 also dem Kernmodell liegt.

Klieme und Hartig weisen in ihrem Überblick über Kompetenzkonzepte in den Sozialwissenschaften mit Blick auf die intensiven Arbeiten an psychometrischen Modellen darauf hin „[...] dass eine Reihe psychometrischer Modelle verfügbar ist. Eine größere Herausforderung als die rein technische Entwicklung statistischer Modelle stellt zweifellos die Formulierung fundierter theoretischer Kompetenzmodelle und deren Umsetzung in empirisch prüfbare Modelle und Messverfahren dar. Sie erfordert die Anwendung von pädagogisch-psychologischer und fachdidaktischer Expertise in der jeweiligen Domäne.“ (Klieme und Hartig 2007, 26).

Die Modellierung muss auf dieser Ebene eine Operationalisierung der Domäneninhalte in Form von Kompetenzstruktur und Kompetenzniveaus unter Zuhilfenahme fachdidaktischer Expertise leisten, um insbesondere auch die Überprüfung und Messung des Erfolgs mittels entsprechender Assessments ermöglichen zu können. Interessant ist beispielsweise, dass Feedback und generell formative Assessments getrennte Kompetenzdimensionen für die Analyse erfordern (Csapo 2010, 24). Also nur wenn die Kompetenzen detailliert modelliert werden, ist eine sinnvolle Individualdiagnostik realisierbar. Hier scheinen in den letzten Jahren viele punktuelle Ansätze verfolgt worden zu sein, die aber einer einheitlichen Beschreibungslogik bedürfen. Das Modell ist also im Sinne der Struktur, wie sie im DFG-Schwerpunktprogramm entwickelt wurde (Abbildung 12) im Kern des Modells zu verorten. Gleichzeitig soll exemplarisch unter Nutzung verschiedener Instrumentarien eine Verbindung bis zur äußeren vierten Ebene hergestellt werden.

5.2 Anforderungen

5.2.1 Domäneninhalt

Zunächst geht es auf Ebene 1 darum, den Domäneninhalt mit Zuordnung zu der jeweiligen Domäne beschreiben zu können. Dabei ist zwischen der Kompetenz und der Qualifikation sowie den Eigenschaften eines Individuums, die von beiden Begriffen nicht erfasst werden, zu unterscheiden (siehe Abschnitte 3.3 und 3.4.2). Als wesentlich sollen hier zunächst die Kompetenzen angesehen werden, Qualifikationen, wie sie in Abschnitt 3.3.3 in Anlehnung an den DQR definiert sind, werden nur im Ausblick in Abschnitt 10.2.1 betrachtet.

Der Kern des Referenzrahmens des DFG-Schwerpunktprogramms enthält Kompetenzen in Form von Strukturen und Niveaus.

Bereits bei dem kurzen Abriss zur pädagogischen Kompetenz in Abschnitt 3.4.1.2 hat sich gezeigt, wie vielfältig die Vorschläge sind, Kompetenz zu strukturieren. Die Strukturierung ließe sich sowohl hinsichtlich der Dimensionen als auch der Granularitätsstufen weiter verfeinern. Das bedeutet für das hier entwickelte Modell, dass die Möglichkeit der Abbildung verschiedener Dimensionen als Erklärungsmerkmale notwendig ist.

Entsprechendes gilt für die Modellierung der Niveaustufen, die ebenfalls eine Mehrdimensionalität und die Verwendung von Operatoren berücksichtigen muss (Abschnitt 4.2.2.2).

Die Wahl eines kombinierten Kompetenzstruktur- und Kompetenzniveaumodells, die daraus resultiert, wurde in den Abschnitten 4.2.2.1 und 4.2.2.2 erläutert. Das Modell kann darüber hinaus Kompetenzentwicklungsmodelle beispielsweise durch Vergleich mehrerer Assessments für einen Assessment-Teilnehmer unterstützen, muss dies aber nicht zwingend in die Modellierung einbeziehen (Abschnitt 4.2.2.3).

Wie in Abschnitt 3.4.6 dargestellt, ist die Messung von Kompetenzen zunächst immer an Inhalte gebunden. Somit steht die Formulierung von inhaltsbezogenen, oder domänenbezogenen Kompetenzen am Anfang. Dies muss entsprechend auch für das hier entwickelte Modell gelten. In Abschnitt 4.1.1 wird bereits eine dreistufige Vorgehensweise abgeleitet, die auch mit der Beschreibung im „Assessment triangle“ in Abschnitt 4.3.1 korrespondiert. Grundsätzlich geht es dabei um

1. Modellierung des Domäneninhalts
2. Modellierung der konkreten Assessments in konkreten Situationen
3. Durchführung des Assessments, Beobachtung und Messung der Performanz und Ableitung der Aussagen für die Kompetenz und das Feedback.

Gleichzeitig soll die Abbildbarkeit verschiedener Domänen berücksichtigt werden, um „modellvariierte Domänendefinitionen“ zuzulassen, wie sie in Abschnitt 3.4.7 beschrieben wurden. Die Modellierung wird dadurch erschwert, dass die inhaltlichen Begrifflichkeiten in vielen Domänen nicht klar gegeneinander abgegrenzt sind und viele der angestrebten Kompetenzen nicht direkt beobachtbar sind, sondern nur in Handlungssituationen indirekt sichtbar werden. Da es nicht realistisch erscheint eine allgemein akzeptierte und darüber hinaus domänenübergreifende inhaltliche Modellierung zu implementieren, muss eine flexible Methodik angestrebt werden, die verschiedene Ansprüche erfüllen kann. Diese Methodik muss

1. eine Anpassbarkeit an verschiedene Begrifflichkeiten und Strukturen hinsichtlich der Kompetenzen,
2. eine Nutzbarkeit über verschiedene Domänen hinweg,
3. eine flexible Integration verschiedener Assessments und
4. eine Analyse mit verschiedenen Scoring-Modellen und statistischen Modellen

zur Verfügung stellen.

5.2.2 Assessment

Aus dem Domäneninhalt soll wie beschrieben das Assessment abgeleitet werden. Dabei wird, wie in Abschnitt 3.1 beschrieben, die Output- oder „Outcome-Orientierung“ (Bildungsstandards der Kultusministerkonferenz 2005, 6) in den Mittelpunkt gerückt. Dies geschieht in der Annahme, dass es der breiten gesellschaftlichen Erwartung an die Ziele von Schule ebenso wie der Intention der Bildungsstandards entspricht.

Daraus folgt, dass die Messbarkeit das Verständnis von Kompetenzen als Kompetenzen im engeren Sinne erfordert, wie sie in Abschnitt 3.4.5 definiert wurden. Der weitere Begriff der Individualkompetenzen aus Abschnitt 3.4.3 soll im Allgemeinen nicht verwendet werden.

Hierzu passt auch der kognitivistische Ansatz wie in Abschnitt 4.2.1 als sinnvoll beschrieben, der den breiten situativen Ansatz und die konkrete Definition der Inhalte beziehungsweise Kompetenzen im engeren Sinne beinhaltet.

Dies legt weiter die kriterienorientierte Bewertung („criterion-based“) nahe, die das Modell unterstützen soll und die in Abschnitt 4.1.2 bereits als sinnvolle Bewertungsart dargestellt wurde. Das hier zu entwickelnde Modell sollte aber auch die anderen Bewertungsarten erlauben.

In jedem Fall soll bei der Dokumentation der Assessment-Ergebnisse zwischen der formativen, rein beschreibenden Beobachtung und der Bewertung dieser Beobachtung unterschieden werden (siehe auch Abschnitt 3.2). Allerdings sollte im Interesse der Validität für summative und formative Assessments auf dasselbe Kompetenzmodell zugegriffen werden (Abschnitt 4.1.3.4).

Wesentlich im Sinne der weiteren Gütekriterien ist in jedem Fall eine Dokumentation der Rahmenbedingungen und Vorgehensweise vor, während und nach der Durchführung eines Assessments (siehe Abschnitt 4.1.3.2 und Abschnitt 4.1.3.3).

Weiterhin ist zu beachten, dass es sich zunächst um die Messung der Performanz im Assessment wie in Abschnitt 3.3.2 definiert handelt. Die Ableitung von Aussagen zur Kompetenz erfolgt getrennt. Dafür wird im Sinne der Ebene 2 des Referenzmodells ein IRT-Modell genutzt. Bei ihm handelt es sich um das in Abschnitt 4.2.5 beschriebene Rasch-Modell.

Diese Nutzung sowie alle weiteren Schritte im Rahmen der oberen Ebenen des Modells sollen nur exemplarisch durchgeführt werden, um die Nutzbarkeit des hier entwickelten Modells zu zeigen.

5.2.3 Evidenz

Das Modell soll über eine Erklärungskomponente verfügen, die von Lernenden und Lehrenden genutzt werden kann. Daher sind eine inhaltliche Zuordnung von Domäneninhalt und Assessment sowie eine Erklärungskomponente für eine differenzierte Zuordnung der Semantik der einzelnen Lösungen zu ergänzen.

Dies entspricht der Idee des ECD, wie in Abschnitt 4.3.2 beschrieben, wo das „Evidence Model“ eine Brückenfunktion einnimmt. Wesentlich ist dabei die Zuordnung der beobachtbaren (manifesten) Variablen zu den latenten Variablen des Domäneninhalts.

Ähnliches findet man, wie in Abschnitt 4.3.3 beschrieben, bei den „Four Building Blocks“ von Wilson, wo Items kausal aus dem Domäneninhalt abgeleitet (cause) und umgekehrt die Ergebnisse als Basis einer Schlussfolgerung (infer) genutzt werden.

5.2.4 Ressourcen

Eine Modellierung ist wie beschrieben die Grundlage einer Messung. Ist das zu Messende nicht definiert, kann auch keinerlei Messung erfolgen. Allerdings gibt es in der Praxis auch den Aspekt der verfügbaren Ressourcen. In der Praxis sind Assessments, die dem kompetenzorientierten Ansatz folgen und entsprechende Lösungen mit vertretbarem Zeitaufwand erlauben, eine wesentliche Anforderung. So beschreibt Winther als eine Perspektive die Entwicklung von Assessmentsystemen aus dem Pool der beruflichen Anforderungssituationen und stellt besonders

die „[...] Konstruktion von Testitems sowie deren Präsentation auf Basis eines validen Kompetenzmodells [...]“ (Winther 2010, 263) als Herausforderung für künftige Entwicklungen heraus. Besonders positiv werden in diesem Zusammenhang Ansätze erwähnt, die die Assessment-Entwicklung eng an die Testvalidierung knüpfen.

Eine Ersparnis ist in der Praxis insbesondere im Bereich der Ergebnisfeststellung zu erwarten, die einen erheblichen Zeitbedarf hat. Wie in Abschnitt 4.1.3.5 dargelegt, werden e-Assessments präferiert, andere Assessments jedoch nicht ausgeschlossen.

5.3 Einordnung in eine Metamodell-Architektur

Für den hier verfolgten Ansatz soll auf die Modellierungstechniken der Informatik zurückgegriffen werden. Modellierung ist ein grundlegendes Mittel der Informatik. So sind Modellierungstechniken und Sprachen beispielsweise für die Systementwicklung, die Unternehmens- oder Wissensmodellierung entwickelt worden. Modelle stellen bei einem konstruktionsorientierten Verständnis die Frage nach dem „Wofür“ (Thomas 2006). Dabei wird im Sinne der Abdeckung aller vier Ebenen des DFG-Modells der Gesamtkomplex als Objekt eines Informationssystems aufgefasst - eines Informationssystems, das die Abbildung der Domäne, zugehöriger Assessments und deren Ausführung umfasst.

In diesem Sinne geht es um die konzeptuelle Modellierung dieses Informationssystems. Grundlegende Prinzipien sind hierbei Hierarchisierung, Strukturierung, Abstraktion, Verbalisierung, Modularisierung, Lokalität und Geheimnisprinzip, (Balzert 1998, 581) die den entsprechenden Modellen und Methoden zugrunde liegen. Hier soll kein komplettes System entwickelt werden, sondern zwei wesentliche Grundlagen, die Datensicht in Form eines kompletten Modells zur Beschreibung der Informationsinhalte und ein Vorgehensmodell zur Befüllung dieses Modells sollen realisiert werden. Vorgehensmodelle werden entworfen, um die Durchführung der Entwicklungsaktivitäten strukturieren zu können (Eicker 2015).

Das Modell ist auf der Metaebene 1 gemäß Meta Object Facility angeordnet und in der Syntax der UML2 Klassendiagramme beschrieben. Die UML-Syntax erleichtert das Verständnis, wird allerdings nur in begrenztem Umfang eingesetzt und bei Bedarf durch semiformale Konstrukte ergänzt, um die Verständlichkeit zu verbessern. Das Modell stellt das Ergebnis einer längeren Entwicklung und Verfeinerung dar. Ein Vorgängermodell wurde bereits 2011 veröffentlicht (Wieken 2011).

Die Entwicklung einer Datensicht in Form eines Informationsmodells entspricht dem Ansatz zunächst ein Modell zur Abbildung der Domäneninhalte und deren prototypische Umsetzung in ein Assessment anzustreben. In diesem Sinne wird der Ansatz des Core-Situation-Personal-Modells (CSP) entwickelt.

Die Grundidee des CSP ist eine Externalisierung des Domäneninhalts in Form einer strukturierten, integrierten und analysierbaren Dokumentation. Die Struktur soll sowohl den inneren Aufbau eines Systems beschreiben als auch die Beziehungen zwischen den Objekten zweckorientiert widerspiegeln. Da alle hier zu strukturierenden Bereiche, insbesondere aber die Kompetenzaspekte, einer hohen Variabilität und Dynamik unterliegen, erscheint es angebracht eine angemessen flexible Form der Modellierung zu präferieren.

Es wurde daher das Konstrukt eines in eine Metaebenenarchitektur eingebetteten Informationsmodells gewählt. Dies erlaubt die strukturierte Modellierung der wesentlichen Aspekte sowie eine flexible Erweiterung oder Anpassung.

Metamodelle sind Modelle, bei denen jedes Element einem Element der übergeordneten Metaebene zugeordnet wird. Es besitzt also ein Typobjekt, eine Klasse oder einen Classifier und einen Link zu diesem Classifier. Der Classifier enthält die Beschreibung des Elementes, seine Eigenschaften und Aktionen. Die einzelnen Elemente sind Instanziierungen des Classifiers. Eine formale Beschreibung, als MOF 2 bezeichnet, findet sich in der Definition der Meta Object Facility in der aktuellen Fassung von 2014 (Object Management Group (OMG) 2014). MOF 2 weist in vielen Bereichen Ähnlichkeiten zu der Unified Modeling Language 2 (UML 2) auf und ist auch in UML beschrieben. UML ist seinerseits eine Anwendung von MOF 2 mit vier Metaebenen.

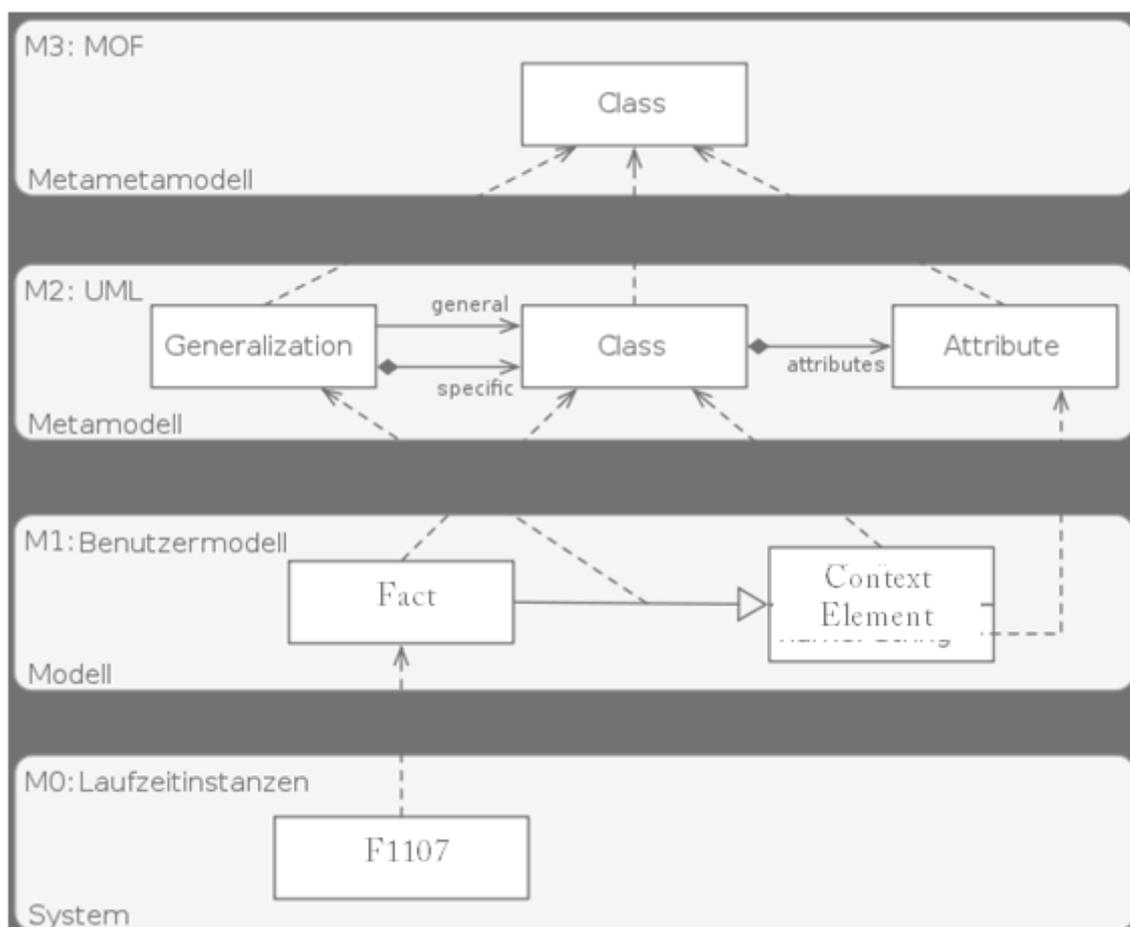


Abbildung 18 MOF Metaebenen für CSP-Informationsmodelle

Die meisten Anwendungen der MOF 2 beschränken sich auf maximal vier Metaebenen, auch wenn theoretisch beliebig viele Ebenen möglich sind. Mindestens werden zwei Ebenen benötigt, da sonst kein Link zwischen Metaebenen möglich ist. Relationale Datenbanken benutzen beispielsweise drei Ebenen (Systemtabellen, Tabellen und Zeilen). (Object Management Group (OMG) 2014, 6f.)

UML ist Bestandteil der Metaebene M2 in der MOF. Damit ergibt sich, dass jedes konzeptionelle Modell, das in UML modelliert wird, zur Metaebene M1 gehört und Metamodell für die konkreten Instanzen ist.

Dies ist neben der weiten Verbreitung und Verfügbarkeit von UML ein wichtiges Argument dafür, dass bei der Modellierung des Informationsmodells UML verwendet wird. So ist das hier vorgestellte CSP-Informationsmodell als UML-Klassenmodell beschrieben, wenn auch viele Möglichkeiten der Modellierung nicht genutzt werden. Es gehört also zur Metaebene M1. In Abbildung 18 sind die Typen **FACT** und **CONTEXT ELEMENT** als Beispiele aus dem CSP-Modell eingefügt. Der konkrete Fact F1107 „Bedeutung der Parameter beim exponentiellen Wachstum“ wird später erläutert. Bei der Umsetzung beispielsweise in eine relationale Datenbank wird die UML-Modellierung des Core-Situation-Personal-Modells zum Schema in der Datenbank, mithin zur Metaebene M1 der Datenbankarchitektur. Der eigentliche Domäneninhalt wird in der Datenbank beschrieben und gehört somit zur Metaebene M0.

Ein Argument für die Einbettung in eine Metaarchitektur ist, dass Flexibilität eine große Rolle für eine Anwendbarkeit eines solchen Modells spielt. Die Verwendung eines Metamodells erlaubt eine Flexibilität in der Umgestaltung und Beschreibung des Domäneninhaltes sowie bei der Zuordnung von Inhalten zu bestimmten Domänen.

Nun kann man nicht davon ausgehen, dass ein solches Modell alle Aspekte beinhaltet und in allen Bereichen richtig ist. Für ein vollständiges Informationssystem wird entsprechende Software benötigt, die auf dem Modell arbeitet. Hier erscheint es zumindest als Gedanke gerechtfertigt, diese Software nicht fest mit der Struktur des Modells zu verbinden, sondern für künftige Änderungen des Modells zu flexibilisieren. Daher erscheint eine zweite Metaebene gerechtfertigt, die das Modell und steuernde Funktionen dieses Modells für die Software beinhaltet. Ein solches die Software steuerndes Modell wäre ebenfalls auf der Metaebene M2 anzusiedeln.

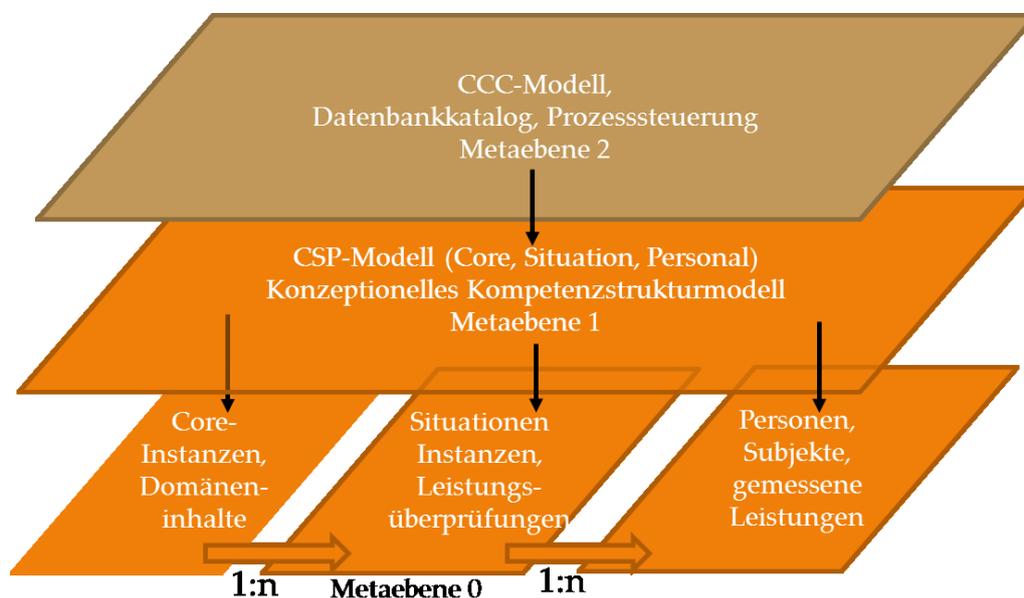


Abbildung 19 Metamodellebenen des CSP-Modells

Die Grundarchitektur, die sich aus diesem Gedanken ergibt, ist in Abbildung 19 dargestellt. Die senkrechten Pfeile stehen für Instanziierungsbeziehungen, in umgekehrter Richtung verläuft jeweils der Classifier-Link. Das CSP-Modell beinhaltet verschiedene Bereiche, von denen die wichtigsten der eigentliche Domäneninhalt (CORE), die diesem Inhalt zugeordneten Situationen für Assessments (SITUATION) und schließlich die persönlichen Ergebnisse der Lernenden

(PERSONAL) sind. Die Beziehungen zwischen diesen Komponenten sind keine Beziehungen zwischen Metaebenen, sondern Beziehungen innerhalb einer Ebene.

Das CSP-Modell wird hier als UML-Modell beschrieben. Die UML-Syntax ist auf der Metaebene M2 angesiedelt. Die Software müsste somit die Modellierungskomponenten selbst implementieren, um die benötigten Details ergänzen. Ein solches Modell, hier als CONTENT CONTROL CATALOG (CCC) bezeichnet, ist nicht Bestandteil dieser Arbeit.

5.4 Idee des Core-Situation-Personal-Modells (CSP)

Entsprechend der Problemstellung wird hier zunächst die grundlegende Architektur skizziert. Das Modell besteht aus einem Informationsmodell zur Dokumentation aller relevanten Inhalte und einem Vorgehensmodell, das den Umgang mit den Informationen beschreibt. Zunächst wird in Kapitel 6 auf das Informationsmodell eingegangen. Das Vorgehensmodell wird in Kapitel 7 beschrieben und beide Modelle in Kapitel 8 angewendet.

Das CSP-Modell ist hierarchisch dreigeteilt in die Bereiche Kernmodell (CORE), Situationsmodell (SITUATION) und Personenmodell (PERSONAL). Die entsprechenden Bereiche für das Informationsmodell sind in Abbildung 20 als Pyramide dargestellt.

Die Basis der Pyramide bildet das Kernmodell (CORE), das der strukturierten Beschreibung der Domäneninhalte dient. Es ist die Basis für alle anderen Modelle und zugleich das inhaltlich komplexeste Modell. Das Kernmodell soll es erlauben, die relevante Sicht auf die zu vermittelnden Domäneninhalte inhaltlich strukturiert zu beschreiben.

Das Situationsmodell (SITUATION) erlaubt die inhaltliche Beschreibung eines Assessments, die immer als situationsbezogen aufgefasst wird. Zu jeder Domäne sollen mehrere Assessments beschrieben und ein Pool von Aufgabenstellungen verwaltet werden können.

Das Personenmodell (PERSONAL) dient der Dokumentation der persönlichen Ergebnisse in einem oder mehreren Assessments. Es bündelt die Ergebnisse für eine Person.

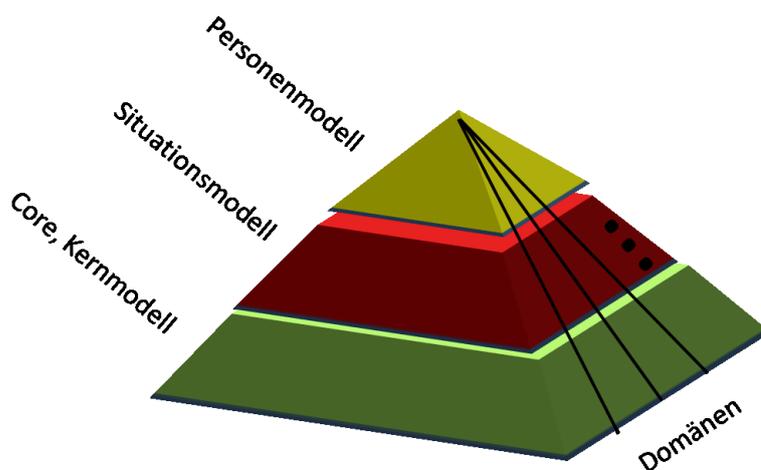


Abbildung 20 Die drei Ebenen des CSP-Modells

5.5 Vorgehen

Bei der Entwicklung des Informationsmodells werden die einzelnen Bereiche des Modells konkret dargestellt und in einem gemeinsamen Modell und den jeweiligen Teilmodellen abgebildet. Dabei werden das Core-Modell, das Situation-Modell und das Personal-Modell nacheinander entwickelt. Insbesondere im Core-Modell zeigt sich, dass die Untergliederung in Teilmodelle im Interesse einer Modularisierung sinnvoll ist. Im Bereich der inhaltlichen Beschreibung und der Kompetenzaggregation wird eine grafische Darstellung entwickelt. Dies entspricht einem Kompetenzstruktur- und einem Kompetenzniveaumodell. Im Sinne des DFG-Modells (siehe Abbildung 12) entspricht das dem Kernmodell (Bereich 1).

Anschließend wird das Situation-Modell zur Beschreibung eines konkreten Assessments entwickelt und ebenfalls in Teilmodelle modularisiert. Insbesondere wird zwischen dem Assessment selbst und seiner Analyse in Form direkt messbarer manifester Variablen und statistisch messbarer latenter Variablen (DFG-Modell, Bereiche 2 und 3) unterschieden.

Schließlich werden die durch die Ausführung des Assessments tatsächlich beobachtbaren und bewertbaren Ergebnisse der einzelnen Assessment-Teilnehmer dokumentiert. Dies geschieht im Personal-Modell. Darauf aufbauend kann mittels der im Situation-Modell dokumentierten unterschiedlichen Möglichkeiten eine Bewertung und Analyse aus verschiedenen Blickwinkeln erfolgen. Diese können eine klassische Bewertung, eine psychometrische Kompetenzschätzung und ein individuelles Feedback beinhalten (DFG-Modell, Bereiche 3 und 4).

Im Anschluss daran wird ein Vorgehensmodell entwickelt, das von der Analyse der externen (curricularen) Ziele und Vorgaben bis zur Analyse führt (siehe Abbildung 17). Bis auf den bewusst nicht berücksichtigten Beurteilungsprozess ist damit die Einbettung in den Bildungsprozess beschrieben.

Dabei werden neben der allgemeinen Vorgehensweise und der Beschreibung der Messbarkeitsanforderungen insbesondere zwei Ansätze, das Evidence Centered Design (ECD) und die „Four Building Blocks“ aufgegriffen aber zusätzlich auf die Wiederverwendbarkeit sowohl der Domäneninhalte als auch der Assessments besonderer Wert gelegt.

Als Testbeispiel schließt sich ein praktischer Teil mit der konkreten Anwendung des CSP-Modells in einer Domäne an. Dabei geht es um lineares und exponentielles Wachstum, ein Thema im Mathematikunterricht in einer 11. Klasse eines beruflichen Gymnasiums. Die Ergebnisse werden abschließend dargestellt. Hierfür wird auf ein e-Assessment zurückgegriffen, auch wenn dies vom Modell nicht zwingend gefordert wird, um den Aufwand bei der Durchführung zu reduzieren. Das Problem der e-Assessments ist ihre Akzeptanz. Klassische Klausuren in all ihren Schwächen (Saldern 2011, 89) sind praxistauglich. Jedes andere Verfahren muss zunächst Akzeptanz schaffen. Diese Akzeptanz kann geschaffen werden, indem

1. vergleichbare Ergebnisse mit klassischen Assessments nachgewiesen werden,
2. nachvollziehbare Ergebnisse mit Erklärungsursachen für einzelne Leistungen verknüpft und so Zusatznutzen in Form individuellen Feedbacks geschaffen wird.

Daher soll bei der Validierung exemplarisch ein Abgleich mit parallel geschriebenen klassischen Leistungsüberprüfungen (Klausuren) erfolgen. Ziel ist es, eine Korrelation beider Messungen im „Groben“ also in der Zielerreichung in Form einer Notenbewertung zu untersuchen.

Außerdem sollen die Ergebnisse detailliert für einzelne Kompetenzen ermittelt werden. Die ermittelten Scores sollen mittels psychometrischer Analyse mit dem RASCH-Modell im Hinblick auf eine Messung der Kompetenz als Trait untersucht und so, wenn möglich, abgesichert werden. Damit wären die Ergebnisse der Kompetenzerreichung über die konkrete Leistungsmessung hinaus nutzbar.

Danach erfolgt exemplarisch eine individuelle Analyse zur Erstellung eines Feedbacks für die Assessment-Teilnehmer und die Lehrenden.

Schließlich soll der Aufwand für das ganze Verfahren anhand der Durchführung der konkreten Leistungsmessung beurteilt werden.

6 CSP - Informationsmodell

6.1 Architektur des CSP-Modells

6.1.1 Strukturelle Modellierungsentscheidungen

Zunächst werden einige Designentscheidungen dokumentiert, die aus den bisher diskutierten Anforderungen an das CSP-Informationsmodell resultieren.

Die ersten Anforderungen ergeben sich aus der Einordnung in die Metaebenenstruktur.

Metaebene 1

Das Kompetenzmodell wird als Metaebene 1 aufgefasst, die beschriebenen Instanzen als Metaebene 0. Bei der Umsetzung in eine Datenbank entspricht das dem Schema und den Instanzen.

Metaebene 2

Das Modell wird mit der UML2 als Bestandteil der Metaebene M2 beschrieben. Für die Unterstützung des Modells über Software ist die Umsetzung einer solchen Metaebene 2 zur Steuerung des Ablaufes wünschenswert.

Innerhalb des Modells ist auf Transparenz aller Beschreibungen zu achten. Nur wenn die Domäneninhalte klar dokumentiert sind, kann ein valides Feedback erfolgen.

Dokumentierte Vorgehensweise

Das Modell muss deutlich machen, welche Inhalte zu einer Domäne gehören, welche überprüft werden sollen, wie die Leistungsüberprüfung konstruiert wurde, wie die Leistung erbracht wurde und wie die Ergebnisse bearbeitet und evaluiert wurden.

Ein Feedback kann nur dann hinsichtlich eines Ziel erfolgen, wenn es auf einem in dieser Hinsicht validen Assessment beruht. Folglich muss der Zusammenhang zwischen dem intendierten Ausschnitt des Domäneninhaltes, dem Assessment und der Verfügbarkeit der entsprechenden Fähigkeiten beim Lernenden auswertbar sein. Das bedeutet, dass Konstruktion, Evaluation und Feedback dieselbe dokumentierte, vernetzte und auswertbare Grundlage besitzen sollten.

Traceability

Das Modell muss die Konstruktion und Evaluation für ein zielgerichtetes Feedback dokumentieren. Um Entscheidungen in einem strukturierten Modell treffen zu können, müssen auswertbare Pfade zwischen den relevanten Informationen bestehen (traceability).

Aggregation

Die Granularität der Modellbeschreibung bestimmt die Granularität der möglichen Evaluation und des Feedbacks. Neben der gewünschten feinsten Granularität muss eine Verdichtung (Aggregation) der Ergebnisse zur Gewinnung von übergreifenden Aussagen möglich sein.

Der Entwurf eines Assessments wird durch die Domäneninhalte und intendierten Kompetenzniveaus bestimmt, vor allem aber durch fachdidaktische Erfahrung. Bereits bei der

Beschreibung und Definition der Kompetenzen hat sich gezeigt, dass sich diese nur indirekt aus der Beobachtung des Verhaltens in Situationen, also bei Anwendung des Assessments ableiten lassen. Die Eignung des konstruierten Assessments muss also bei der Beurteilung der Kompetenz der Assessment-Teilnehmer simultan mit deren Kompetenz überprüft werden. Dies ist unter anderem das Grundprinzip der IRT-Statistikverfahren. Daher muss auch diese Überprüfung im Modell dokumentierbar sein.

Assessmentbewertung

Das Modell muss neben den individuellen Bewertungen der Assessment-Teilnehmer eine Möglichkeit der Beschreibung und Evaluation des Assessments selbst ermöglichen.

6.1.2 Die CSP-Ebenen

Wie bereits beschrieben ist das CSP-Modell (vgl. Abbildung 20) dreigeteilt in die Bereiche Kernmodell (CORE), Situationsmodell (SITUATION) und Personenmodell (PERSONAL). Die drei Bereiche befinden sich im Sinne des Metamodells auf derselben Metaebene M1.

Die Basis der Pyramide bildet das Kernmodell, das der strukturierten Beschreibung der Domäneninhalte dient. Es ist die Basis für alle anderen Modelle und zugleich das inhaltlich komplexeste Modell. Dieses Modell soll es erlauben, die relevante Sicht auf die zu vermittelnden Domäneninhalte inhaltlich strukturiert zu beschreiben. Dadurch sind diese Inhalte als Basis für alle zu konzipierenden Assessments wiederverwendbar.

Das Situationsmodell erlaubt die inhaltliche Beschreibung eines Assessments, das immer als situationsbezogen aufgefasst wird. Die Trennung beider Ebenen erlaubt es, zu jeder Domäne mehrere Leistungsüberprüfungen beschreiben und verwenden zu können. Durch die Trennung von Domäneninhalt und Assessmentbeschreibung ist eine vollständige oder teilweise Wiederverwendung von bekannten Assessments möglich. Der Aufwand für die inhaltliche Dokumentation einer Domäne ist erheblich. Er ist dadurch zu rechtfertigen, dass dazu ein Pool von Assessments mit entsprechenden Items entsteht, der komplett oder teilweise wiederverwendbar ist. So kann bei einem breiteren oder wiederholten Einsatz derselben oder ähnlicher Assessments erheblicher Aufwand eingespart werden.

Das Personenmodell dient der Dokumentation der persönlichen Ergebnisse. Durch die Abtrennung von Kern- und Situationsmodell ist es möglich, die Ergebnisse aller Personen in einem oder auch mehreren Assessments zu dokumentieren. Zugleich bündelt das Modell die Ergebnisse für einen Assessment-Teilnehmer (eine Person), so dass auch assessment-übergreifende Analysen beispielsweise für eine Entwicklungsanalyse grundsätzlich durchführbar sind.

Dreiteiliges Modell

Das Modell wird inhaltlich dreigeteilt. Die Domäneninhalte, die Leistungsüberprüfungen und die personenbezogenen Leistungsmessungen werden in drei Teilmodellen beschrieben, die jeweils eine Rückbeziehung auf das zugrundeliegende Modell erlauben.

Die drei Teile des CSP enthalten jeweils Rückbezüge zum übergeordneten Modell, also vom PERSONAL- zum SITUATION-Modell und vom SITUATION- zum CORE-Modell, um jeweils Auswertungen über die Modelle hinweg durchführen zu können. Damit können insbesondere die von Wilson angesprochenen „Causes“- und „Infer“-Beziehungen realisiert werden.

Bei der Beschreibung des Modells werden die Begriffe „Typ“ und „Beziehung“ verwendet. Statt Typen müsste man eigentlich von Klassen sprechen. Da aber nicht die softwaretechnische Umsetzung, sondern die Strukturierung der Inhalte im Vordergrund steht, wurde eine entsprechende Bezeichnung verwendet. Die Attribute der einzelnen Typen sind nicht im Modell dargestellt, um dieses übersichtlicher zu gestalten. Sie sind im Anhang in den ergänzenden Beschreibungen der jeweiligen Teilmodelle mit einer kurzen Erläuterung komplett aufgeführt. Die Namen der Typen werden in Fettdruck und Großbuchstaben eingefügt. Pluralbildungen werden nach Bedarf für die verwendeten englischen Begriffen vorgenommen, also beispielsweise die **DOMAIN** und die **DOMAINs**. Deutsche Begriffe werden gelegentlich für ein besseres Verständnis ergänzt.

Die Beziehungen zwischen den einzelnen Teilmodellen werden immer nur angedeutet.

6.2 Das CORE-Modell

Das CORE-Modell dient der Dokumentation von Kompetenzstruktur und Kompetenzniveaus und sowie deren Operationalisierungen und somit der Domäneninhalte unabhängig von einem konkreten Assessment.

6.2.1 Überblick

6.2.1.1 Allgemeiner Teil des CSP-Modells

Zunächst werden die Teile des Modells betrachtet, die nicht von einer konkreten Domäne abhängig, sondern generell von allen Domänen unabhängig sind. Diese Teile bilden die Klammer, die der Verwaltung von Domänen dient und generell wiederverwendbare Teile beschreibt. Wichtig ist dabei, dass für die Inhalte jeweils bestimmbar ist, welcher Domäne sie angehören. Diese Möglichkeit wird durch den Typ **DOMAIN** dargestellt, mit dem alle Inhalte des CORE-Modells verknüpft werden können. Durch Änderung der Verknüpfung kann jederzeit eine Zuordnung zu einer anderen Domäne erfolgen. Auch die Zuordnung zu mehreren Domänen ist möglich.

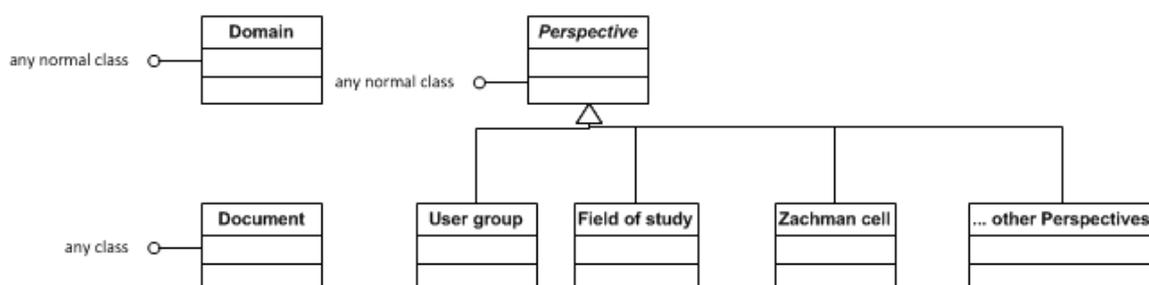


Abbildung 21 Allgemeiner Teil des Modells

Neben der **DOMAIN** sind im allgemeinen Modell weitere Typen vorgesehen, die bei Bedarf an die anderen Modelltypen angebunden werden können, der Übersichtlichkeit wegen aber nicht überall dargestellt werden. Neben der **DOMAIN** spielt die **PERSPECTIVE** eine wichtige Rolle. Sie ermöglicht bei Bedarf die Darstellung derselben Sachverhalte aus unterschiedlichen Perspektiven. Bei den Perspektiven kann es sich um Benutzergruppen (**USER GROUP**), Fachdisziplinen (**FIELD OF STUDY**), strategische Ebenen (**ZACHMAN CELL**) (Burgess und Hokel 1994) oder andere Perspektiven handeln.

Das **DOCUMENT** erlaubt die Integration verschiedener Materialien in verschiedenen Formaten. Es kann ebenfalls mit allen anderen Typen verbunden werden.

6.2.1.2 Die Bereiche des CORE-Modells

Im Weiteren ist die Frage zu klären, wie die Inhalte der einzelnen Domänen beschrieben werden sollen. Dafür ist eine Einteilung des CORE-Modells in verschiedene Modellbereiche vorgesehen. Die einzelnen Modellbereiche sind in Abbildung 22 jeweils als Sechseck dargestellt, um möglichst viele Querbeziehungen zwischen den einzelnen Bereichen bereits optisch symbolisieren zu können. Diese Modellbereiche bilden in ihrer Gesamtheit das CORE-Modell. Sie werden einzeln betrachtet, wobei jeweils die Querbezüge zu den anderen Bereichen ebenfalls dargestellt werden. Die Querbezüge zum SITUATION-Modell werden an dieser Stelle noch nicht dargestellt, sondern im Zusammenhang mit diesem Modell beschrieben.

Tatsächlich umgesetzt wurden im Rahmen der Arbeit nur die ersten sechs Teilmodelle. Die verbliebenen drei Teilmodelle (etwas dunkler gefärbt) sind Ideen für Weiterentwicklungen, die im Ansatz (Process) skizziert beziehungsweise im Ausblick in Abschnitt 10 beschrieben sind.

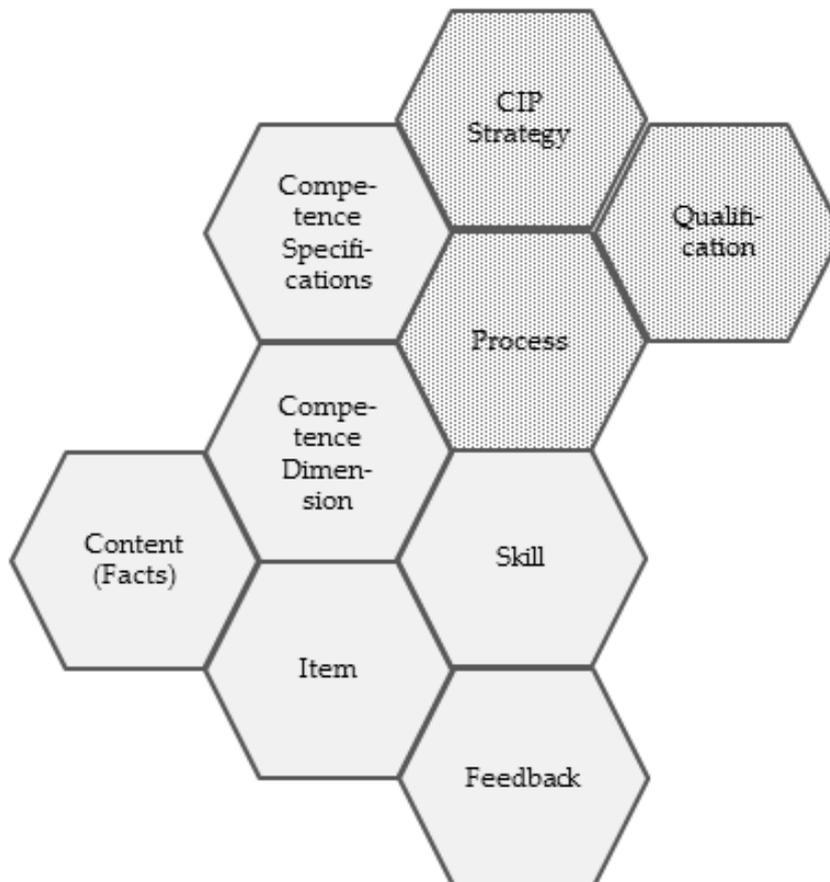


Abbildung 22 Die 9 Teilmodelle des CORE-Modells⁵

- (1) Content - Fakten, Domäneninhalt
- (2) Item – Operationalisierung des Contents
- (3) Skill – Fähigkeiten, Operatoren

⁵ Eine frühere Version dieser Gliederung findet man bei Wieken (Wieken 2011).

- (4) Competence Specifications – Externe Ebene der Kompetenzdefinition
- (5) Competence Dimension – Interne Gliederung der Kompetenzen
- (6) Feedback – Detaillierte Rückkopplung zum Bildungsprozess
- (7) Qualification - Qualifikationen
- (8) Process - Konventioneller Prozess mit Ablauf und Verzweigungen
- (9) CIP/Strategy – Dynamische Struktur eines kompetenzorientierten, flexiblen Prozesses

6.2.2 Modellbereich Content

Der erste und zentrale Bereich für die Beschreibung des Domäneninhaltes ist der Modellbereich Content. Für kognitive Kompetenzmodelle ist die Formulierung einer Domäne grundlegend (Anderson und Krathwohl 2001). Dabei ist wesentlich, dass Kompetenzen in weiten Bereichen kontextspezifisch sind, also an ähnliche Inhalte und Situationen gebunden sind (Winther 2010, 36). Darüber hinaus gilt gemäß dem ACT-R Modell von Anderson (J. Anderson 1993), dass sich Wissen umso leichter auf andere Situationen übertragen lässt, je ähnlicher die deklarativen („chunks“) und prozeduralen („procedures“) Wissensbestände sind. Umgekehrt wächst also die Unabhängigkeit einer Kompetenz von einer Domäne je allgemeiner und somit von der Domäne unabhängiger die sie beschreibenden „chunks“ und „procedures“ sind (Winther 2010, 25).

Das spricht dafür, diese „chunks“ und „procedures“ genauer zu definieren und zu beschreiben, um bei der Konstruktion der Situationen einen Bezug zum Umfang der Anforderungen zu bekommen, wobei hier auf eine Automatisierung wie sie ACT-R anstrebt verzichtet wird. Die Aufteilung in deklaratives und prozedurales Wissen soll auch im vorliegenden Modell nachvollzogen werden, wobei dies durch Untertypen der hier maßgeblichen FACTs realisiert wird (Abbildung 29). In der weiteren Betrachtung soll der deklarative Teil im Vordergrund stehen. Prozedurales Wissen im Sinne der Messung einer Prozessdurchführung wird als Idee integriert. Weitere Ergänzungen wie die bereits erwähnte strategische Kompetenz oder die von Gelmann und Greeno (Gelman und Greeno 1989) erwähnte „Utilizational Competence“ sowie andere Überlegungen, die vor allem auf volitionale und motivationale Aspekte abheben und den Handlungsspielraum des Kontextes beschreiben, werden hier nicht ausmodelliert, könnten aber analog zum deklarativen und prozeduralen Wissen zugeordnet werden, wenn deren Operationalisierung gelingt (in Abbildung 29, als weitere Verfeinerungen der FACTs mit „...“ angedeutet).

Der Domäneninhalt wird als „Wissen“ aufgefasst, das für das Begreifen der Domäne und das Handeln in der Domäne benötigt wird. Das modellierte Wissen soll die für die Teilnehmer relevanten und beobachtbaren Aspekte beschreiben. Es soll nicht im philosophischen Sinne der Versuch gemacht werden, die Dinge als solche zu beschreiben, unabhängig von der Frage, ob diese überhaupt so existieren und ob sie von uns in ihrer Gänze erfassbar sind. Nachdem jede Erkenntnis Stückwerk ist, sollen pragmatisch genau die relevanten Wissensstücke dokumentiert werden. Dies ist insofern wichtig, als dass dies eine Beschränkung unter Nutzbarkeitsaspekten darstellt, die jederzeit nach Bedarf abgebaut werden kann.

6.2.2.1 Ausgangspunkt natürliche Sprache

Zunächst erscheint es unmittelbar klar, was mit dem Domäneninhalt gemeint ist. Es handelt sich um das, was als Ziel des Bildungsprozesses definiert ist. Auch wenn dies in Form von Kompetenzen formuliert wird, wurde schon darauf hingewiesen, dass sich die Kompetenzen letztlich immer zunächst an den Inhalten orientieren müssen (Winther 2010, 36).

Diese Inhalte sind in der Praxis zumeist natürlich-sprachlich formuliert und nicht immer eindeutig verständlich. Somit muss jede Analyse des Inhaltes letztlich von umgangssprachlichen Darstellungen und den damit verbundenen Problemen ausgehen. Daher ist es sinnvoll, zunächst zu betrachten, womit sich die Sprachwissenschaft, genauer das Teilgebiet der Semiotik befasst. Sie untergliedert sich in Syntax, Semantik und Pragmatik [beispielsweise (Brekle 1972, 21ff.)].

Die Syntax beschreibt die Regeln für die Anordnung der möglichen Zeichen/Zeichenketten, die Semantik beschäftigt sich mit deren Bedeutung. Die Pragmatik rückt die Kommunikation zwischen Sender und Empfänger der Zeichenketten in das Zentrum der Betrachtung. Unter Beachtung informationstheoretischer Aspekte kann man bei der inhaltlichen Analyse aus pädagogischer Sicht den strukturellen und den metrischen Aspekt hinzufügen (Schott, Lehrstoffanalyse 1975, 25ff.). So erhält man für eine inhaltliche Beschreibung fünf Aspekte [ebd.]:

- (1) Syntax: beschreibt die erlaubte Reihenfolge von Zeichen und Zeichenketten
- (2) Semantik: beschreibt die inhaltliche Bedeutung der Zeichen und Zeichenketten
- (3) Pragmatik: beschreibt die unmissverständliche und klare Kommunikation der Inhalte, was eine klare, kurze und übersichtliche Darstellung beinhaltet.
- (4) Struktur: Die Vermittlung und Erarbeitung von Inhalten ist für die Pädagogik zentral. Die Dokumentation typischer Strukturen innerhalb der Inhalte erleichtert dies.
- (5) Metrik: Die Quantifizierung von Domäneninhalten erfordert eine Dokumentation der enthaltenen Elemente und deren Beziehungen.

Die geschilderten Aspekte einer inhaltlichen Analyse erfordern offensichtlich die Übertragung einer natürlich-sprachlichen Beschreibung in ein formaleres Beschreibungssystem, denn dass die Analyse natürlicher Sprache erhebliche Probleme verursacht zeigen zwei kleine Beispiele aus der Computerlinguistik.

Das erste Beispiel ist der Satz (Carstensen, et al. 2001): „Time flies like an arrow.“ der sich kontextfrei übersetzen ließe mit

- (1) Zeit fliegt wie ein Pfeil.
- (2) Zeitfliegen mögen einen Pfeil.
- (3) Bestimme die Geschwindigkeit von Fliegen so, wie es ein Pfeil tut.

Das zweite Beispiel zeigt, dass dem Satz: „Hinter dem Betrug werden die gleichen Täter vermutet, die während der vergangenen Tage in Griechenland gefälschte Banknoten in Umlauf brachten.“ durch einen Parser 92 verschiedene Interpretationen zugeordnet werden können (Kuhn und Rohrer 1997).

Die Interpretation natürlicher Sprache kann daher nicht kontextfrei erfolgen, sondern bedarf Zusatzwissens. Dies kann in automatischen Systemen hilfsweise durch Einsatz externen formalisierten Wissens [hierzu beispielsweise (Kuropka 2004)] geschehen. Dies kann auch und wird in der Praxis zumeist durch den Einsatz menschlichen (Vor-)Wissens geschehen. Aber auch diese zusätzliche Kontextkomponente kann nicht immer eindeutige Ergebnisse liefern. Somit kann

nicht von einer eindeutigen Modellierbarkeit des Inhaltes auf der Basis vorhandener natürlich-sprachlicher Beschreibungen ausgegangen werden. Der Sinn der Texte kann unterschiedlich interpretiert werden.

6.2.2.2 Anforderungen an den Modellbereich Content

Kann man also nicht sicherstellen, dass der Sinn eindeutig erkannt und modelliert werden kann, so muss doch bei der Bewertung der Beobachtungen in einer konkreten Situation deren Ergebnis zumindest aus der gewählten Perspektive des Bewertenden eindeutig bewertbar sein. Daher muss die gewählte Perspektive bei der Beschreibung deklarativen, prozeduralen und anderen Wissens derart erfolgen, dass sie eindeutig bewertbar ist, im einfachsten Fall als richtig oder falsch.

Anforderung 1 Bewertbarkeit

Der Domäneninhalt muss so modelliert werden, dass eine eindeutige Bewertung möglich ist.

Dies erfordert eine Formalisierung der Sprache, da wie dargestellt, eine natürliche Sprache nach heutigem Stand zu komplex und nicht eindeutig ist.

Gleichzeitig muss der Inhalt der Beschreibung für die Anwender verständlich sein. Daher sollte die formalisierte Darstellung eine Nähe zu der gewohnten natürlich-sprachlichen Darstellung beibehalten. Hilfreich für das Verständnis einer formalisierten Darstellung ist außerdem die Möglichkeit, eine Darstellung mit Beispielen illustrieren und zumindest teilweise überprüfen zu können.

Anforderung 2 Sprachliche Nähe

Die gewählte Darstellung des Inhaltes soll Nähe zu einer sprachlichen Darstellung aufweisen.

Anforderung 3 Beispiele

Es sollte die Möglichkeit der Illustration der Darstellung mit Beispielen geben.

Hier soll, wie beschrieben, der Domäneninhalt als Modell des Wissens aufgefasst werden, das für das zielgerichtete Handeln benötigt wird. Es soll nicht der Versuch gemacht werden, die Dinge als solche zu beschreiben, sondern nur die relevanten Aspekte. Das bedeutet auch, dass nicht alle denkbaren Zusammenhänge, nicht alle Aspekte modelliert werden.

Anforderung 4 Wissensinseln

Es sollen Fragmente, Inhaltsinseln modelliert werden können ohne ein geschlossenes Ganzes modellieren zu müssen.

Die Modellierung von Wissensinseln kann auch dazu führen, dass Widersprüche in den Inhalten existieren. So kann beispielsweise eine Literaturinterpretation aus marxistischer Sicht anders lauten als eine aus bürgerlicher Sicht. Ohne die zusätzliche Modellierung der geisteswissenschaftlichen Hintergründe, die aber nicht Teil des Inhaltes ist, können hier Widersprüche auftreten.

Anforderung 5 Widersprüche

Es können auch Widersprüche modelliert werden.

Im Zusammenhang mit formalisiertem Wissen spielt die Frage des automatischen Schließens eine wichtige Rolle. Im Rahmen der Inhaltsmodellierung wird bewusst auf die Anforderung eines automatischen Schließens (Inferenz) innerhalb des Systems verzichtet. Der Vorteil der Gewinnung von impliziten Inhalten durch automatisches Schließen wird als nicht zentral für die Aufgabenstellung angesehen. Die Gründe hierfür sind:

- Die impliziten Inhalte, die durch automatische Folgerungen gewonnen werden können, sind für die Gestaltung der Assessments zunächst nicht sichtbar und somit nicht dokumentiert und nicht bewertbar.
- Die impliziten Inhalte sind für die Lehrenden nicht transparent, sie sind in der Praxis nicht überschaubar und mögliche Modellierungsentscheidungen können unüberschaubare Auswirkungen haben.
- Modellerte Widersprüche können in einem Inferenzprozess kaum vorhersagbare Ergebnisse erzeugen.

Dies bedeutet nicht, dass solche Mechanismen künftig nicht gerade vor dem Hintergrund des exemplarischen Lernens hilfreich werden könnten.

6.2.2.3 Formale Beschreibungsmethoden

Vor dem Hintergrund dieser Anforderungen sollen kurz vorhandene Systeme zur Modellierung von Wissen betrachtet werden. Hier bieten sich bekannte Bereiche wie die Prädikatenlogik, formale Sprachen, regelbasierte Systeme oder Ontologien für die formale Beschreibung von Inhalten an.

6.2.2.3.1 Prädikatenlogik

Die Prädikatenlogik ist bei der Inhaltsanalyse sowohl direkt (Kreschnak 1985) als auch als Grundlage formaler Sprachen (Schott 1975) eingesetzt worden. Zugleich bildet sie die Basis bei der Umsetzung semantischer Netze in Ontologien. Die Prädikatenlogik erlaubt die Formalisierung von Aussagen und deren Überprüfung innerhalb eines gegebenen Systems (Diskursuniversum). Prädikate sind Folgen von Wörtern mit definierten Leerstellen. Sie werden oft mit einem Großbuchstaben bezeichnet und enthalten nummerierte Leerstellen, beispielsweise $F_{(1,2)}$ als zweistelliges Prädikat. Die Lücken werden von Individuenkonstanten oder anderen Prädikaten gefüllt. Eine Individuenkonstante bezeichnet ein einzelnes Ding, das im betrachteten Namen als solches aufgefasst wird, beispielsweise „Einstein“ oder „1933“. Es kann sich auch um Klassenausdrücke wie „Eisen“ oder „Frage“ handeln, wenn diese im betrachteten Zusammenhang als atomar und damit individuell aufgefasst werden. Durch die Einbeziehung der Möglichkeit von Prädikaten als Instanziierungen der Lücken entsteht eine Prädikatenlogik höherer Ordnung.

Zusätzlich sind noch die Quantoren, der Allquantor und der Existenzquantor einzuführen, um zu beschreiben, ob eine Aussage für alle oder für mindestens eine Instanziierung gilt.

6.2.2.3.2 Formale Sprachen

Formale Sprachen dienen dem mathematisch korrekten Umgang mit Zeichenketten, wobei es darum geht, eine endliche Darstellung für die unendlich vielen Sätze einer Sprache zu entwickeln. (Salomaa 1973, 3). Sie bilden die Grundlage vieler Programmiersprachen. Ihr Zweck ist aber nicht primär die Kommunikation, wofür sie weniger gut geeignet erscheinen.

6.2.2.3.3 Ontologien

Ontologien werden im Zusammenhang mit Wissensrepräsentation oft genannt. Bei ihnen handelt sich nach Gruber um die „explizite Spezifikation einer Konzeptualisierung“ (Gruber 1993, 200).

Ontologien findet man in verschiedenen Disziplinen, so in der Philosophie, der Linguistik und der Informatik. Letztlich geht es um die Überwindung der Diskrepanz von einer Sache an sich und ihrer sprachlichen Darstellung.

Das Grundproblem sprachlicher Darstellung lässt sich veranschaulichen: Wenn etwas beschrieben werden soll, zeigt die Semiotik, dass zumeist zwischen drei Formen zu unterscheiden ist. Sie werden oft als semiotisches Dreieck dargestellt, wobei die Bezeichnung der Seiten unterschiedlich ist (Wikipedia - Semiotisches Dreieck 2015).

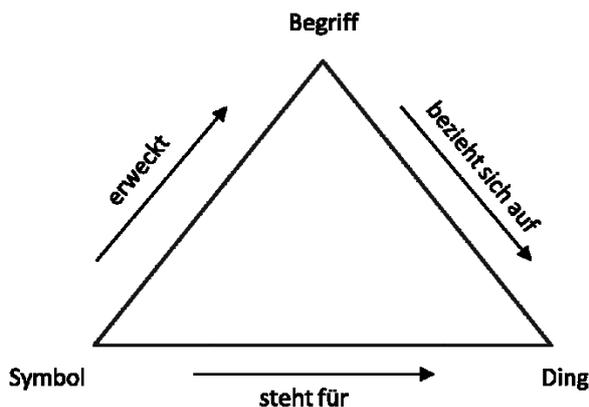


Abbildung 23 Semiotisches Dreieck [nach (Wikipedia - Semiotisches Dreieck 2015)]

Es geht es zunächst um das „Ding“ als solches, das „Bezeichnete“, das im semiotischen Dreieck zumeist unten rechts dargestellt wird. Das Symbol (Wort), das als Bezeichnung gewählt wird, steht zumeist unten links, das „Bezeichnende“. Es existiert nun keine direkte Verbindung zwischen dem Bezeichnenden und dem Bezeichneten. Vielmehr existiert ein Begriff, eine Intention, ein Inhalt, das „Erwartete“, das dem „Bezeichneten“ zugeordnet wird und durch das „Bezeichnende“ ausgedrückt werden soll. Diese Zuordnung ist problematisch. Sie hängt letztlich von erkenntnistheoretischen Fragen ab, die im Wesentlichen dem Relativismus oder dem Realismus zugeordnet werden können. Nimmt man im realistischen Sinne zumindest an, dass eine erkennbare Realität außerhalb des menschlichen Denkens existiert, so gibt es immer noch viele Sichtweisen über die Wahrnehmungsmöglichkeiten. So unterscheiden einige Wissenschaftstheoretiker zwischen Theorien und Modellen und erkennen an, dass mit Modellen gearbeitet wird, die stets nur bestimmte Aspekte der Wirklichkeit erfassen können (Wikipedia - Kritischer Realismus 2015).

Vereinfacht gesagt sollen Ontologien dazu dienen, genau dieses verbindende Element zwischen „Bezeichnendem“ und „Bezeichneten“, also das „Erwartete“ zu formalisieren (Stuckenschmidt 2009, 8). Während die Philosophie eher die Prinzipien dieses vermittelnden Elementes zu analysieren versucht, sucht die Informatik in einem konstruktivistischen Ansatz nach Möglichkeiten Folgerungen auf Basis der Ontologie erstellen zu können (Inferenz). Dazu werden im Wesentlichen semantische Netze formalisiert, indem deren Struktur in Formeln umgesetzt wird, wobei wiederum auf die Prädikatenlogik zurückgegriffen wird (Stuckenschmidt 2009, 27ff.). Konkrete Sprachen sind hierzu beispielsweise RDF, OWL und F-Logic.

Typische Kennzeichen von Ontologien sind (Stock und Stock 2008, 256f.):

1. Eine standardisierte Ontologie-Sprache
2. Mechanismen zum automatischen Schlussfolgern

3. Berücksichtigung von Allgemeinbegriffen und Individualbegriffen
4. Verwendung unterschiedlicher, typisierter Relationen.

Ontologien sind also aus dem Interesse der automatisierten Informationsverarbeitung entstanden und auf Inferenzmechanismen ausgerichtet. Gemäß einer pragmatischen Sicht auf Ontologien ist „[...] die Funktion einer Ontologie weniger die korrekte Beschreibung der Realität, als vielmehr die Unterstützung einer korrekten Interpretation eines gegebenen Datenbestandes.“ (Stuckenschmidt 2009, 97).

6.2.2.3.4 Folgerungen für die Auswahl

Aus den hier beispielhaft dargestellten grundlegenden formalen Beschreibungen ergeben sich folgende Folgerungen. Die Anforderung 1 aus Abschnitt 6.2.2.2 wird durch formale Darstellungen klar erfüllt. Die Stärke dieser Systeme liegt in den intendierten Abfrage- und Folgerungsmöglichkeiten für implizites Wissen. Die Anforderungen 3 und 4 sind grundsätzlich umsetzbar, allerdings mindert die Umsetzung von Wissensinseln in erheblichem Maße die gerade erwähnten Stärken der Systeme.

Wesentlich ist aber, dass sich im Bereich der Anforderungen 5 und insbesondere 2 erhebliche Schwierigkeiten ergeben. Keine der Darstellungsformen ist leicht lesbar oder in der Nähe einer natürlichen Sprache. Der Modellierungsaufwand und die mangelnde Lesbarkeit stehen einer Anwendung im hier betrachteten Zusammenhang entgegen. Auch Widersprüche sind schwer mit den Konstruktionsprinzipien von Ontologien oder anderen formalen Ansätzen vereinbar.

Für das Ziel dieser Arbeit ist daher auf die Verwendung eines streng formalen Ansatzes verzichtet worden. Zusammenfassend sind die Gründe

- Der pragmatische Kommunikationsaspekt erfordert eine leicht lesbare Form.
- Automatische Schlussfolgerungen sind zunächst nicht notwendig.
- Es soll keine komplette Weltsicht oder Domänensicht modelliert werden, sondern der relevante Ausschnitt soll abgebildet werden.
- Durch die Ausschnittbildung bedingt, sollen auch widersprüchliche Aussagen modelliert werden können.

Daher werden im Folgenden semi-formale Darstellungen betrachtet. Die Nutzung solcher Darstellungen soll nicht als komplette Ablehnung formaler Darstellungen aufgefasst werden, sondern als Einstieg, der durchaus eine weitere Formalisierung nicht ausschließt. Dieser Weg wurde beispielsweise auch von der KMDL bei der Beschreibung wissensintensiver Prozesse eingeschlagen (Gronau und Fröming 2006).

6.2.2.4 *Semi-formale Darstellung*

Der „pragmatische“ Ansatz, der für eine semi-formale, zumeist graphische Sprache spricht lässt sich durch typische Anforderungen der Anwender konkretisieren. Als Ergebnisse einer Metaanalyse von anwendungsbezogenen Kriterien für den Einsatz semi-formaler Darstellungen findet man beispielsweise (Hogrebe, Gehrke und Nüttgens 2010) als Top 7 der Anforderungen aus Anwendersicht: Einfachheit, Flexibilität, Genauigkeit, Verständlichkeit, Vollständigkeit, Zeiterfordernis und Zweckmäßigkeit. Diese Begriffe sind schwer zu definieren und auch der übergeordnete Begriff der Usability ist widersprüchlich definiert (Schalles, Rebstock und Creagh 2010).

Allerdings zeigt die Praxis der weiten Verbreitung von grafischen Modellierungssprachen insbesondere im Bereich der Anwendungssysteme mit der UML, der Prozessmodellierung beispielsweise mit EPK, BPMN, YAWL und ähnlichen Sprachen sowie der ungebrochenen Nutzung von ERM im Bereich der Datenmodellierung beispielhaft deren Akzeptanz. Somit kann man davon ausgehen, dass die Top 7 zumindest in ausreichendem Maß berücksichtigt werden. Daher soll hier als ergänzende Anforderung die Verwendung einer graphischen Darstellung postuliert werden ohne den obigen Ausgangspunkt einer sprachnahen Darstellung fallenzulassen.

In der Informatik sind im Rahmen der Entwicklung von Informationssystemen bereits zahlreiche graphische Modellierungsmethoden von Wissens- oder Informationsdomänen entwickelt worden. Derartige Modelle sind bereits früh im Rahmen der Softwareentwicklung entstanden. Die vielleicht prominentesten Vertreter sind die Datenmodelle, die in ihrer ursprünglichen Form von Chen (Chen 1976) entwickelt und dann in zahlreichen anderen Varianten weiterentwickelt wurden. Für die Softwareentwicklung seien die strukturierten Methoden [eine Übersicht findet man beispielsweise in (Pressman 1989)] und die objektorientierten Methoden, wie sie heute in der im Wesentlichen aus den Methoden von Booch, Rumbaugh und Jacobson entstandenen UML zum Einsatz kommen, genannt (Booch, Rumbaugh und Jacobson 2006). Im Rahmen der Wissensabbildung sind neben den bekannten Modellen zahlreiche weitere Modellierungen entstanden, so seien beispielhaft ältere Ansätze wie die Nijssen Information Analysis Method (NIAM) (Nijssen und Halpin 1989), spezialisierte Darstellungen wie die Toulmin-Diagramme für Argumentationsketten (Toulmin 1958) oder Wissenanwendungskarten (Lehner 2009, 196) erwähnt.

Sollen die zumeist graphischen Darstellungen in Datenbanken oder anderen Datenträgern strukturiert abgelegt werden, bedarf es dazu eines entsprechenden Metamodells oder einer entsprechenden strukturierten Sprachdefinition.

Anforderung 6 Graphische Darstellung

Es soll eine graphische Darstellung verwendet werden, die eine ausreichende Nähe zu einer sprachlichen Darstellung und die prinzipielle Möglichkeit einer Formalisierung beinhaltet. Daher soll sie sich in einem Metamodell abbilden lassen.

Detaillierte Anforderungen hierfür sind also

1. (6A) graphische semi-formale Darstellung
2. (6B) Möglichkeit einer Formalisierung
3. (6C) Abbildbarkeit in einem Metamodell

6.2.2.5 Auswahl von NIAM als Basis

Als graphische Beschreibung wurde die schon etwas ältere NIAM-Methode (Nijssen und Halpin 1989) gewählt, da sie viele der bisher erörterten Anforderungen erfüllt. Es wurden die Grundzüge und graphischen Darstellungen übernommen und in der Folge den Anforderungen entsprechend modifiziert und erweitert. Zunächst soll die Idee der NIAM kurz dargestellt werden.

Die Grundidee der NIAM ist die Abbildung des Universe of Discourse in ein System von Entities und Fakten. Der grundsätzliche Umfang des Universe of Discourse wird durch den Domäneninhalt vorgegeben. Unter Entities werden konkrete oder abstrakte Basiselemente (Begriffe) verstanden, die für die Dokumentation des Domäneninhaltes relevant sind. Die Modellstruktur soll aus der Abbildung der den Begriffen entsprechenden Entities und ihren Beziehungen in Form

der Fakten deutlich werden. Diese Vorstellung wird auch in der Pädagogikliteratur in einer dort geäußerten Sicht aus der Praxis ähnlich formuliert: Der Kompetenzaufbau „[...] zeigt sich darin, inwieweit die gelernten Begriffe, Informationen zu einem in differenzierter Weise verknüpften sachlichen Beziehungsnetz (einem semantischen Netzwerk) aufgebaut wurden.“ (Beyen 2013, 286).

Entities werden zu Entitytypen abstrahiert. In NIAM wird jedes Entity einem Entitytyp zugeordnet, beispielsweise (Nijssen und Halpin 1989, 39)

- **Einstein** -> **WISSENSCHAFTLER**
- **USA** -> **LAND**
- **1933** -> **JAHR**

Dies entspricht der klassischen Datenmodellierungssicht. Andererseits ist diese Unterscheidung manchmal auch umständlich. Klassenbegriffe wie „Stuhl“, „Durst“ oder „Gold“ müssen nicht zwingend instanziiert werden. Es kann im Sachzusammenhang auch unwichtig sein, zu welcher Klasse ein Entity gehört. Daher sollen sowohl individuelle Entities als auch Klassenbegriffe zulässig sein. Es können auch Klassen als Typisierung von Entities gebildet werden. Generell soll hier verallgemeinernd von Elementen gesprochen werden.

Ein Fakt stellt Beziehungen zwischen Entitäten her. Er übernimmt eine zentralere Aufgabe als in Datenmodellen, er ist der Kern der Darstellung, nicht nur eine Beziehung zwischen Entitäten. Außerdem hat er eine sprachliche Bedeutung. Ein Fakt ist anschaulich eine Aussage mit mehreren Rollen. Jede Rolle wird durch ein Entity instanziiert. Ein Fakt kann also ein oder mehrere Rollen besitzen, die wie Platzhalter für die entsprechenden Entitäten wirken. Ein Fakt entspricht somit einem Prädikat und ein oder mehreren Elementen, die als Variablen die Rollen instanziiieren. Die Verbindung zur Prädikatenlogik erscheint intuitiv und soll später konkretisiert werden.

So kann mit den oben definierten Elementen ein Fakt formuliert werden (Nijssen und Halpin 1989, 39):

- **Der WISSENSCHAFTLER Nachname Einstein**
- **zog**
- **im JAHR 1933**
- **in LAND USA.**

In diesem Beispiel ist das Element „Einstein“ vom Typ WISSENSCHAFTLER und wird über seinen Nachnamen identifiziert. Das Element „1933“ ist vom Typ JAHR und das Element „USA“ vom Typ LAND. Die Elemente können zusätzlich mit Qualifiern wie Nachname beschrieben und graphisch wie in Abbildung 24 dargestellt werden.



Abbildung 24 Zwei Elemente mit ihren Qualifern

Das Element „Einstein“ instanziiert die Rolle „zog“, und die beiden anderen Rollen „im“ und „in“ werden in dieser Reihenfolge von den beiden Elementen „1933“ und „USA“ instanziiert. In seiner Rohform wäre der Fakt als

- $\text{zog}(_ ; \text{im } _ ; \text{in } _)$

beschrieben. Instanziiert ergibt sich

- $\text{zog}(\text{Einstein}; \text{im } 1933; \text{in } \text{USA})$ oder $\text{Einstein zog im } 1933 \text{ in } \text{USA}$.

Die Qualifier können nach Bedarf ergänzt werden. Es ergibt sich die Darstellung in Abbildung 25.

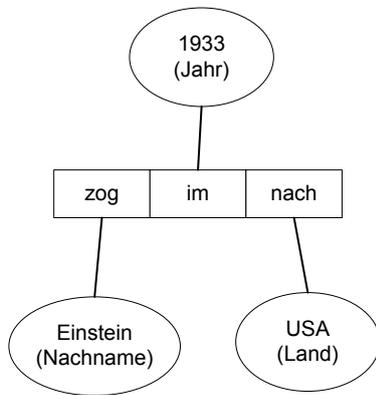


Abbildung 25 Fakt nach NIAM

6.2.2.6 Modellierung des Contents

Die vorgenommenen Modifikationen beginnen mit der Art der graphischen Darstellung. Die graphische Repräsentation des obigen Faktensatzes ist in Abbildung 26 wiedergegeben. Die Darstellung ist der NIAM-Darstellung sehr ähnlich. Der Fakt mit seinen Rollen ist als Rechteck wiedergegeben und die als Ellipsen dargestellten Entitäten instanziiert die Rollen. Qualifier wie „Nachname“ können optional angegeben werden. Andere Aspekte der NIAM, die stärker datentechnisch motiviert sind entfallen, da es nicht um die Umsetzung in ein Datenbanksystem geht. Zu diesen Aspekten gehören insbesondere Uniqueness-Constraints, Entity-Type-Constraints, mandatory und optional roles sowie occurrence frequencies.

Die Notation ist in ihrer grundsätzlichen Logik erhalten geblieben, es sind allerdings Änderungen vorgenommen worden, die sich entsprechend der Intention an Sprache und Darstellung von Inhalten statt an der technischen Umsetzung orientieren. Der Fakt selbst hat auch einen Namen („Beispiel 1“) erhalten, um die Fakten bezeichnen zu können und die Betonung liegt auf seiner sprachlichen Aussage. Die Rollen werden jentsprechend der sprachlichen Logik als Valenzen, als zu füllende Leerstellen, bezeichnet. Die Instanziiierung der Valenzen ist gerichtet. Die Richtung wird entsprechend der Quantorenlogik bestimmt. Die Modellierung eines auf eine Valenz hinzeigenden Pfeils entspricht einem Allquantor, also der Aussage „Für alle ... gilt“. Ein wegführender Pfeil entspricht der Nutzung einer Existenzquantors in der Art „es gibt mindestens ein ... für das gilt“ zugeordnet. Bei Einzelementen ist dies weniger entscheidend als bei Begriffsklassen.

Die Richtung kann bei der sprachlichen Formulierung (Nennung der Instanz vor oder nach der Rollenbezeichnung) genutzt werden. Die Modellierung weist darauf hin, dass es bei auf die Valenz zeigenden Pfeilen zumeist sinnvoll ist, das Entity vor die Valenz zu setzen, bei wegführenden Pfeilen dahinter. Damit wird die Reihenfolge im Satz geregelt und die Lesbarkeit erhöht. Dies ist aber nur ein Hinweis für die praktische Nutzung.

Ein Element, dessen Instanziierungspfeil auf den Rollennamen zeigt ist zumeist ein „aktives“ Element oder das Element über das eine Aussage getroffen wird. Sinnvoll ist es die erste Rolle mit einem Prädikat zu belegen, so dass die erste Entität die Rolle des Subjekts in der Aussage einnimmt.

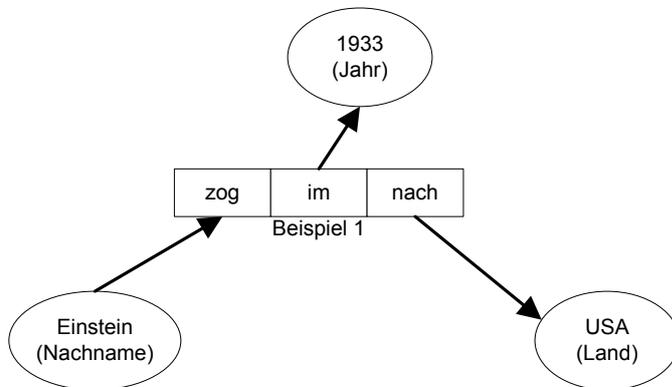


Abbildung 26 Fakt in graphischer Darstellung

Statt von Rollen soll von Valenzen (Wertigkeiten) gesprochen werden. Sie werden als Platzhalter für die Elemente aufgefasst. Der Begriff wird aus der Linguistik übernommen und beschreibt die Fähigkeit eines Wortes, andere Wörter an sich zu binden (Agel 2000). Statt von Fakten kann auch von Sätzen oder Aussagen gesprochen werden.

Man könnte auch sagen, dass es darum geht, Platzhalter oder Leerstellen im Satz zu schaffen und festzulegen, wie diese zu füllen sind. Wird als erste Valenz ein Prädikat gewählt, bestimmt dies die weiteren möglichen Valenzen. Dabei können auch direkte Objekte auftreten, so dass nicht alle Valenzen zwingend benannt werden müssen. Zudem können Valenzen auch durch Fakten, also Aussagen und nicht nur durch Entitäten belegt werden. So kann eine Aussage als kausale Ergänzung mit der Valenzbezeichnung „weil“ die Valenz einer anderen Aussage belegen.

Als Beispiel für ein daraus ableitbares Modell für einen Ausschnitt eines Domäneninhaltes ist ein Bereich aus dem Mathematikunterricht ausgewählt worden, wobei es thematisch um lineares und exponentielles Wachstum geht. Der Ausschnitt ist in Abbildung 27 dargestellt. Dort finden sich einige Fakten, die beispielhaft verbalisiert wurden.

- **(F1104): Alle Linearer Prozess werden beschrieben durch Gleiche Veränderung pro Periode**
- **(F1105): Alle Exponentieller Prozess werden beschrieben durch Exponentialfunktion $f(x)=a \cdot b^x$**
- **(F1103); Alle Exponentieller Prozess beschreiben Proportionale Veränderung zum aktueller Bestand.**

Natürlich-sprachlich sind diese Verbalisierungen syntaktisch nicht vollkommen korrekt, aber doch noch verständlich und sie erscheinen nahe an einer sprachlichen Darstellung. Hier könnten beispielsweise durch Singular und Pluralnamen, Qualifier oder veränderte Schreibweisen weitere Verbesserungen hinsichtlich der Lesbarkeit erreicht werden.

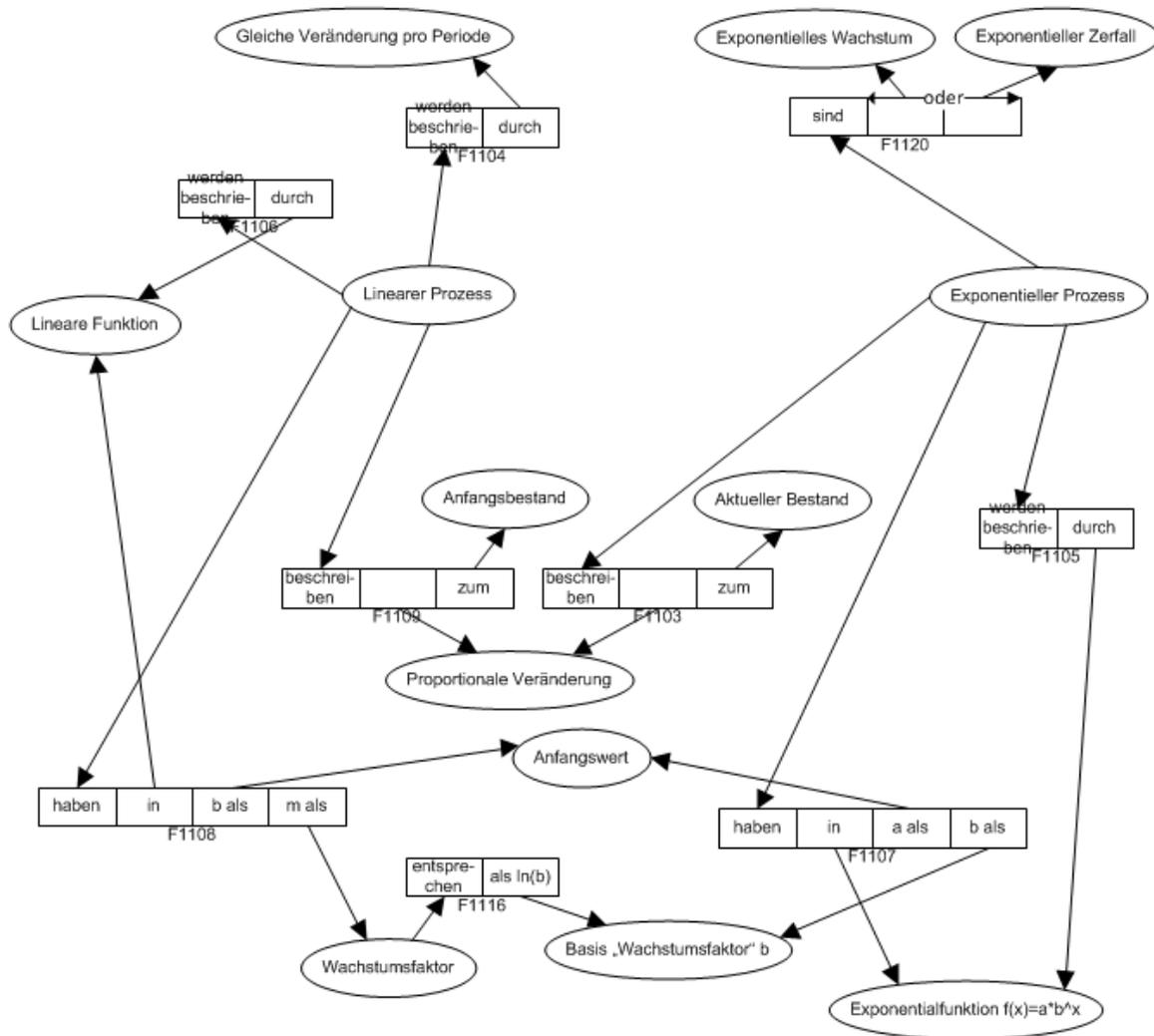


Abbildung 27 Ausschnittweise Modellierung von Wachstumsprozessen

Eine Besonderheit sieht man in Abbildung 27 beim Fakt F1120.

- **(F1120): Alle Exponentieller Prozess sind Exponentielles Wachstum oder Exponentieller Zerfall**

Das Besondere ist das „oder“, in der Graphik durch einen waagerechten Pfeil angedeutet, der zwei Valenzen umfasst. Dies ist ein Beispiel für Constraints, die hier für beliebige typisierte benannte Bedingungen zugelassen werden. Für eine oder mehrere Valenzen können solche Constraints definiert werden, die Regeln vorgeben, die die entsprechenden Valenzen erfüllen müssen. Die Einhaltung eines Constraints kann bewertet werden. Die Menge der möglichen Typisierungen wird über das zugrundeliegende konzeptionelle Modell (CSP) definiert. Die logischen Operatoren AND, OR, NOT sind sinnvolle Beispiele. Die in NIAM eingeführten Constraints zwischen Fakten wie equality oder exclusion sind erweitert worden. Es können sowohl logische Constraints als auch andere algorithmisch definierte Constraints verwendet werden. Diese können zwischen mehreren Valenzen desselben oder unterschiedlicher Fakten ebenso bestehen wie zwischen Fakten, zwischen Fakten und Valenzen oder zwischen Entitäten. Es wird wie in Abbildung 27 beim Fakt 1120 beispielhaft zu sehen ein Doppelpfeil verwendet, der die betroffenen Teile umfasst oder auf sie zeigt und eine entsprechende Beschriftung vorgenommen.

Die Einbettung in den Gesamtzusammenhang sieht vor, dass die Aussagen (Fakten) aus den Kompetenzen abgeleitet werden. Dies stellt den eigentlichen Kern der Operationalisierung dar. Mit den Fakten müssen gleichzeitig die Valenzen, die Entitäten, die Constraints und die sie ergänzenden Informationen modelliert werden.

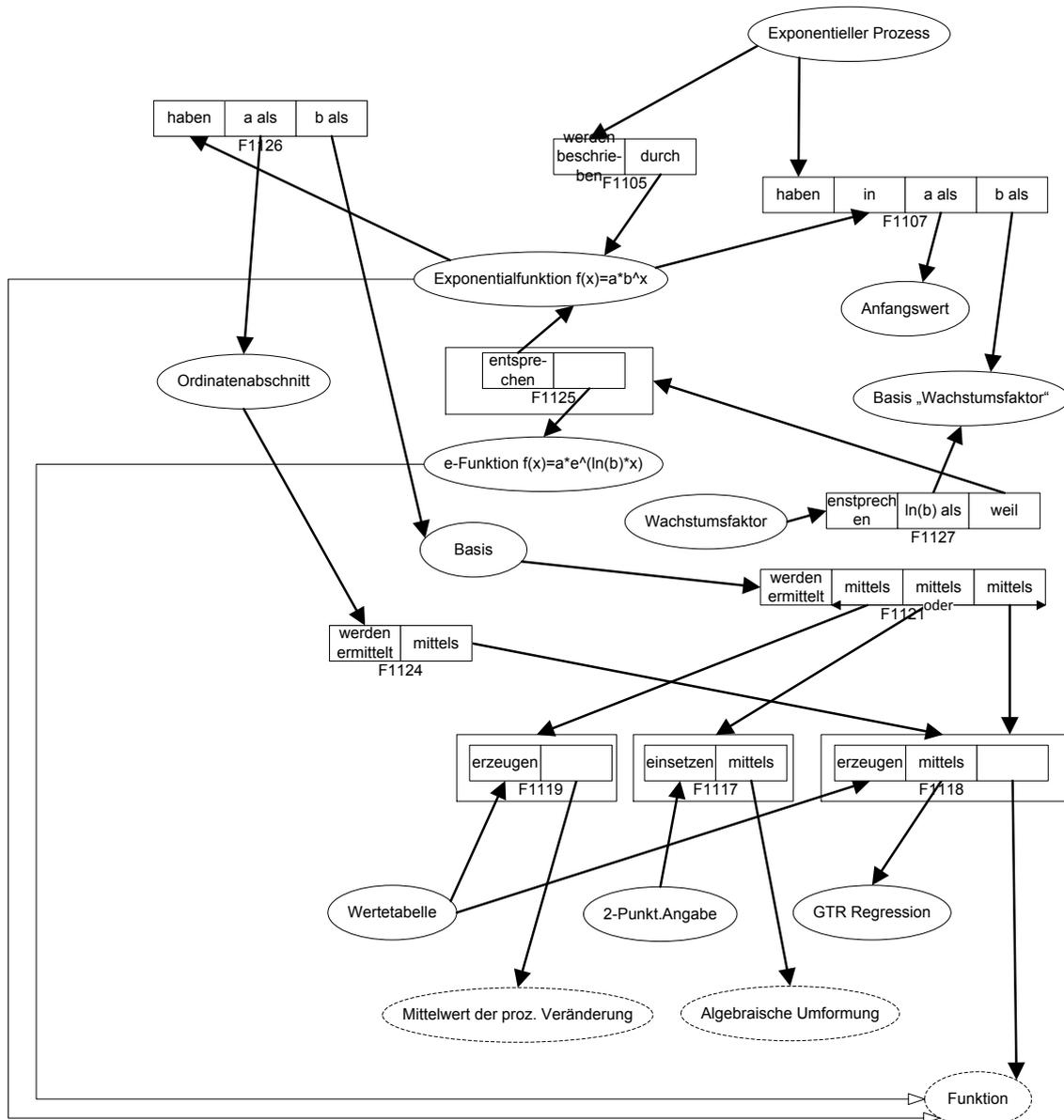


Abbildung 28 Ausschnittweise Modellierung von Exponentialfunktionen

In Abbildung 28 ist ein weiterer Ausschnitt eines Domäneninhaltes dargestellt, diesmal aus dem Bereich der Exponentialfunktionen, wobei es beispielsweise durch die Fakten F1105 und F1107 Überlappungen mit dem Ausschnitt in Abbildung 27 gibt. An dieser Stelle werden weitere Modellierungsmöglichkeiten sichtbar. So ist in Fakt F1127 eine Valenz „weil“ mit dem Fakt 1125 belegt, so dass eine Kausalität dargestellt werden kann. Im Fakt 1121 werden sogar drei Valenzen mit verschiedenen Fakten belegt und durch einen ODER-Constraint miteinander verbunden. Drei Entitäten, darunter die Entität „Funktion“ sind mit gestrichelten Ellipsen verbunden, wodurch sie als Entitäten markiert sind, für die eigene Bereiche modelliert werden. Außerdem ist Funktion als Oberbegriff der Exponentialfunktion und der e-Funktion gekennzeichnet.

6.2.2.7 Das konzeptionelle Modell für den Bereich Content

Die graphische Darstellung basiert auf dem Modellbereich Content im CSP. NIAM selbst schlägt strukturell ein solches Modell vor (Nijssen und Halpin 1989, 297), das als Basis dienen kann. Die im CSP verwendete, stark modifizierte Version, ist in Abbildung 29 als UML-Klassendiagramm dargestellt.

Im Zentrum stehen hier die **FACTs**. Der Begriff **FACT** oder Aussage verdeutlicht, dass ein **FACT** immer wahr oder falsch sein muss. Er wird daher hier gebraucht, obwohl seine Verwendung schwierig ist, weil er im pädagogischen Umfeld leicht mit „Faktenwissen“ in Verbindung gebracht werden kann. **FACTs** dienen als Oberbegriff für verschiedene Wissensarten.

Wie oben dargestellt, kann ein **FACT** oder eine Aussage einem umgangssprachlichen Satz ähneln. Er besteht aus einer oder mehreren **VALENCES**, oder deutsch Valenzen. Eine Valenz kann im Regelfall durch ein **ENTITY** aber auch durch einen anderen **FACT** instanziiert werden, was in Abbildung 28 beispielsweise anhand des Faktus F1121 verdeutlicht wird. Damit kann sie sprachlich mit verschiedenen Konstrukten und kompletten Aussagen (Sätzen) gefüllt werden.⁶

Eine **ENTITY** kann eine Individualkonstante oder eine Klasse repräsentieren. Sie besitzt zusätzlich eine Abstraktionsbeziehung, die der direkten Abbildung von Abstraktionen und Typisierungen dient, beispielsweise der Abbildung, dass Einstein ein Wissenschaftler ist. Grundsätzlich wäre dies auch über die Fakten abbildbar, erscheint so aber klarer. Eine **ENTITY** kann auch einen **QUALIFIER** besitzen. Ein Beispiel für einen Qualifier ist der „Nachname“ beim Wissenschaftler Einstein.

Die **FACTs** können typisiert werden. So kann beispielsweise zwischen deklarativem Wissen, prozeduralem Wissen [„chunks“ und „procedures“ (J. Anderson 1993)] und der Fähigkeit zielorientiert zu handeln (Winther 2010, 57) oder anderen Wissensarten unterschieden werden. Dies geschieht durch die **FACT-TYPES**. **FACT-TYPES** dienen der Klassifizierung der **FACTs** beispielsweise als deklaratives oder prozedurales Wissen. Zusätzlich können für bestimmte Wissenstypen Untertypen definiert werden, die eine weitere detaillierte Modellierung erlauben. Im Modell sind die **CHUNKs** und die **PROCEDUREs** ausgewiesen und weitere mögliche Arten durch die drei Punkte angedeutet. Die Begriffe **CHUNK** und **PROCEDURE** wurden bereits im Zusammenhang mit Andersons Modell ACT-R verwendet und sollen hier übernommen werden. Die Klassifikation in verschiedene Wissenstypen ist nicht abschließend vorgenommen, um Anpassungen an die heterogenen Klassifikationen und künftige Entwicklungen zu erleichtern.

⁶ Es ist auch denkbar beispielsweise prozedurales Wissen abzubilden und die Valenzen mit Prozessen oder Prozessschritten zu füllen.

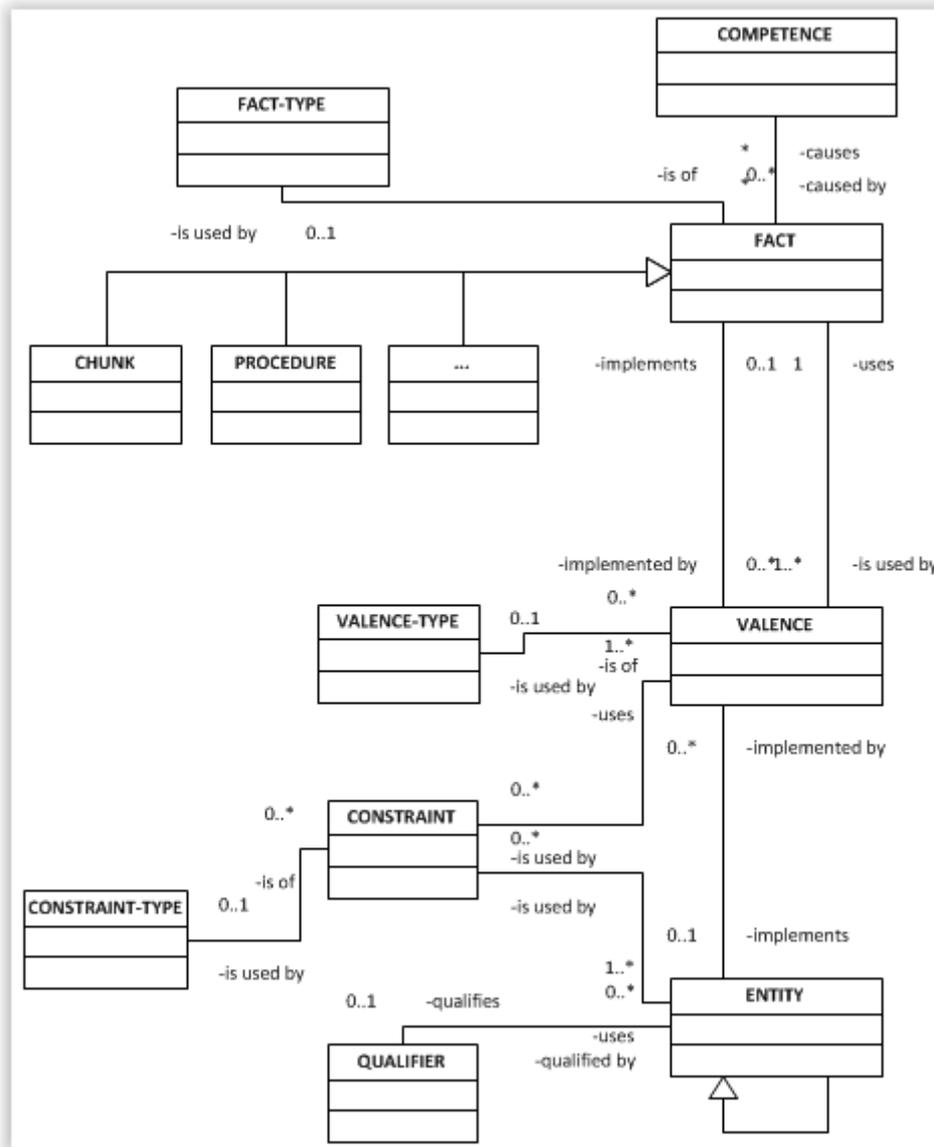


Abbildung 29 Modellbereich Content - deklaratives Wissen

Auch VALENCEs können typisiert werden. Dies erlaubt beispielsweise syntaktische Klassifikationen wie Präpositionen mit Kategorisierungen wie „lokal“, „temporal“, „kausal“ oder „modal“ abzubilden.

Die CONSTRAINTs sind Bedingungen, die ein oder mehrere VALENCEs oder ENTITYs betreffen können. Im Beispiel ist bereits ein ODER-Constraint verwendet worden. Über den Constraint-Type können die entsprechenden Typen wie ODER, UND, EXKLUSIV-ODER, NICHT hinterlegt oder Bedingungen für die Gültigkeit des FACT (WENN) oder andere prozedurale Bedingungen modelliert werden.

6.2.2.8 Bereich Content – Wissenstyp prozedurales Wissen

In Abbildung 29 ist der Type FACT in verschiedene Wissenstypen untergliedert worden. Eine dieser Untergliederungen ist der Typ PROCEDURE, der der Abbildung prozeduralen Wissens dienen soll. Hier soll exemplarisch gezeigt werden, wie ein weiteres Teilmodell PROCESS daran anknüpfen kann. Dies ist im hier betrachteten Kontext als Idee, nicht als Teil des eigentlichen CSP zu verstehen. Dabei wird an den klassischen Prozessbegriff angeknüpft und ein solcher Prozess als

SIMPLE PROCESS oder Elementarprozess bezeichnet. Bei der weiteren Verfeinerung wird auf die klassischen Komponenten eines Prozesses, insbesondere **ACTIVITY**(Aktivität), **CONNECTION** (Verbindung), **OPERATION** und **EVENT** zurückgegriffen. Sie könnten in einer gängigen Modellierungssprache wie BPMN oder YAWL modelliert werden.

Im Sinne der Allgemeingültigkeit erscheint es sinnvoll, für die Beschreibung des Prozesses im CSP auf ein möglichst von einer konkreten Prozesssprache unabhängiges, beziehungsweise auf eine solche Sprache abbildbares Modell zurückzugreifen. Das vorliegende Modell zur Abbildung eines **SIMPLE PROCESS** ist angelehnt an das Modell von Shazad, Elias und Johannesson (Shazad, Elias und Johannesson 2009), das diese Voraussetzungen erfüllt (Abbildung 30).

Es gibt eine Reihe anderer Ansätze. So ist vom WfMC die XPDL (XML-based Process Definition Language) (WfMC 2002) zum Austausch von Prozessdaten als Standard vorgeschlagen worden. Ein umfangreicher Vergleich mit anderen Sprachen (Aalst 2003) zeigt die Problematik unterschiedlicher Sprachelemente und Werkzeugansätze auf. Die Sprache scheint zumindest in dieser Form nicht allen Anforderungen eines Metadatenaustausches zu genügen. Ein neuerer ontologiebasierter Ansatz ist die BPMO (Cimpian, et al. 2008), die Business Process Modeling Ontology, die stärker auf die Umsetzung fokussiert und zudem eng an die BPMN angelehnt ist (Cimpian, et al. 2008, 8). Daneben existieren verschiedene proprietäre Ansätze. Für die vorliegenden Zwecke erscheint der sprachübergreifende Ansatz am sinnvollsten. Zudem handelt es sich um ein Teilmodell, das an definierten Schnittstellen in das Gesamtmodell eingebettet ist und so bei Bedarf austauschbar ist. Die Grundelemente des Modells entspringen wie angegeben dem Universal Process Repository wie von Shazad u.a. beschrieben. Die Autoren geben auch ein Mapping an mit dem die grundlegenden Elemente in die bekannten Prozessbeschreibungssprachen abgebildet werden können.

	Generic Metamodel	Activity Diagram	BPMN	EPC	YAWL
Subprocess	Activity	Activity	Subprocess Task	Subprocess Function	Task
Event	Start Event	Initial node	Start Event	Pre-activity Event	Input condition on activity
	End Event	Final node (process)	End Event	Post-activity Event	Output condition on activity
Control Flow	Operator	Fork node, join, decision, merge	AND/OR(XOR), Complex	XOR, AND, OR connector	AND, XOR, OR split & join
	Connection	Control flow, Object flow	Sequence Flow Association, Message flow	Control flow	No formal name of the construct
Participant			Pool	Org. Unit, Org. Role	
Ressource	Informational Ressource	Object Node	Data objects	Information objects	

Abbildung 30 Mapping der Metatypen in gängigen Prozessbeschreibungssprachen [nach (Shazad, Elias und Johannesson 2009, 181)]

Das Teilmodell **PROCESS** ist in Abbildung 31 dargestellt. Ein darin beschriebener Elementarprozess besteht aus einer oder mehreren Aktivitäten. Zusätzlich können auch ein oder mehrere

Teilprozesse existieren, die ihrerseits aus Aktivitäten bestehen. Die Aktivitäten sind durch einfache Verbindungen oder Operatoren miteinander verknüpft. Sie können durch Ereignisse ausgelöst werden und ihrerseits Ereignisse auslösen. Ferner werden sie durch die Prozesssubjekte ausgeführt beziehungsweise betreffen diese und arbeiten auf der Basis von Ressourcen.

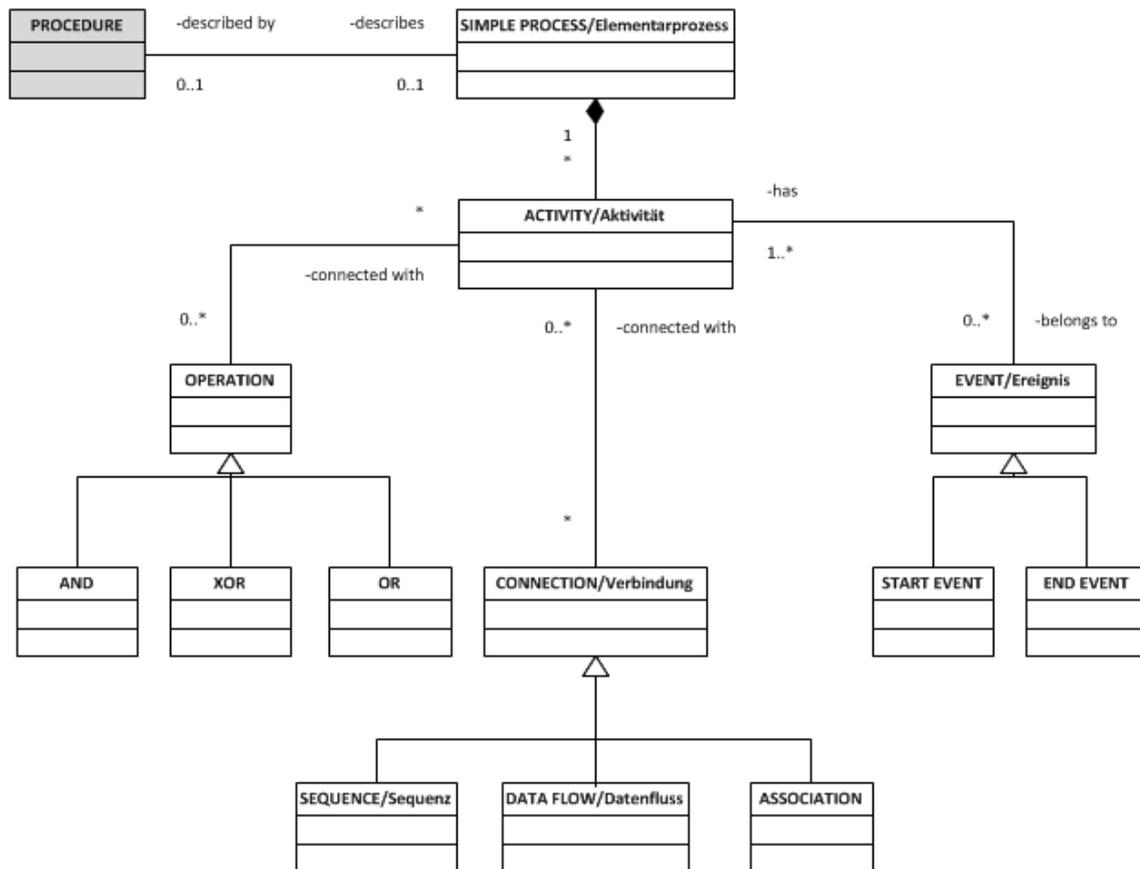


Abbildung 31 Wissenstyp PROCEDURE beispielhaft modelliert

Dieses Modell wird in das Gesamtmodell eingebettet als Verfeinerung des Wissenstyps **PROCEDURE**. So kann im Rahmen der Mathematikdomäne beispielsweise die Berechnung eines Extremums beschrieben oder es kann die Vorgehensweise bei der Nutzung einer Gerätefunktion hinterlegt werden.

6.2.2.9 Bezug zu den detaillierten Anforderungen

In Abschnitt 6.2.2.4 sind Anforderungen an die Darstellung formuliert worden, auf die nach Einführung der Beschreibung und des Modellbereichs Content noch einmal Bezug genommen werden soll.

6.2.2.9.1 Bewertbarkeit (1)

Aussagen sind stets so formuliert, dass sie als wahr oder falsch bewertbar sind. Die Entscheidung über die Erkenntnisse der Assessment-Teilnehmer ist abhängig von der Operationalisierung und der Entscheidbarkeit auf dieser operationalisierten Ebene. Dies wird in der weiteren Entwicklung des SITUATION-Modells und des PERSONAL-Modells gezeigt.

6.2.2.9.2 Sprachliche Nähe (2)

Das Faktenkonstrukt mit seinen Valenzen ist sprachlich nah an der Konstruktion eines Aussagesatzes. Es nimmt das Konzept von Subjekt und einem Prädikat, das die weiteren Valenzen bestimmt, auf. Im Fokus stehen die durch das Prädikat bestimmten Valenzen, auch wenn andere Bestandteile ebenfalls Valenzen besitzen. Dies kann genutzt werden, wenn Präpositionen als Valenznamen verwendet werden, die durch entsprechende Begriffe ihre Valenzen besetzen.

In einem pragmatischen Ansatz werden Fakten beschrieben, die der Sprachlichkeit, dem „Bezeichnenden“ verhaftet bleiben und deren Begrifflichkeit in möglichst guter Annäherung von verschiedenen Personen verständlich aufgenommen werden kann. Zugleich bleibt die Granularität im Bereich des Satzes und erscheint daher praktikabel.

6.2.2.9.3 Beispiele (3)

Die Fakten können tabellenartig durch Angabe der Beispiele ergänzt werden. Beispielhaft kann der Fakt F1106 in Abbildung 27 mit Beispielen wie folgt ergänzt werden.

LINEARER PROZESS	wird beschrieben durch	LINEARE FUNKTION
Geschwindigkeit bei Erdbeschleunigung in Abhängigkeit der Zeit t.	wird beschrieben durch	$v(t) = 9,81 t$
Erlöse bei einem Preis von 4 Euro in Abhängigkeit der Menge x.	wird beschrieben durch	$E(x) = 4 x$
Holzbestand eines fiktiven Waldes in Abhängigkeit der Zeit t.	wird beschrieben durch	$B(t) = 1,4 t$

Tabelle 4 Beispiele für einen Fakt

Weitere Beispiele können analog für andere Fakten erstellt werden.

6.2.2.9.4 Wissensinseln (4)

Die Modellierung von Wissensinseln ist bereits dadurch sichergestellt, dass die Modelle nicht zusammenhängend sein müssen.

6.2.2.9.5 Widersprüche (5)

Es gibt keinerlei Regeln, die sich inhaltlich widersprechende Aussagen verbieten. So erfordern beispielsweise in der Volkswirtschaft das Keynesianische und das neoklassische Modell in derselben Situation sich widersprechende Maßnahmen. Die erste Aussage entspricht der nachfrageorientierten Theorie (Keynes), die zweite der angebotsorientierten Theorie (Neoklassik):

Staat betreibt aktive Konjunkturpolitik in Depressionsphase.

Staat betreibt keine aktive Konjunkturpolitik in Depressionsphase.

Der Widerspruch lässt sich durch den entsprechenden Kontext auflösen:

Staat betreibt aktive Konjunkturpolitik in Depressionsphase bei Nachfragepolitik

Staat betreibt keine aktive Konjunkturpolitik in Depressionsphase bei Angebotspolitik

Es besteht keine Notwendigkeit diesen Kontext herzustellen. Beide Aussagen können im Modell enthalten sein. Die Bewertung beider Aussagen kann im SITUATION-Modell gleich gehandhabt oder unterschiedlich gewichtet werden.

6.2.2.9.6 Graphische semi-formale Darstellung (6A)

Die graphische Darstellung ist in obigen Beispielen erkennbar. Bei entsprechender Modellierung kann ein **FACT** von links nach rechts gelesen werden und die Instanziierungen können in den **FACT** eingesetzt werden. Es entsteht eine lesbare Aussage allerdings mit syntaktischen Mängeln.

Allein die Darstellung erzeugt bereits eine Formalisierung des Ausdrucks. Durch die Typisierung der verschiedenen Bestandteile kann eine weitere Formalisierung erreicht werden.

Lässt man die Typisierung beispielsweise der **CONSTRAINTs** offener, besteht die Möglichkeit einer größeren Flexibilisierung der Valenzen und damit der durch diese ausgedrückten Beziehungen. Eine strengere Typisierung erzwingt eine höhere Formalisierung. Somit hängt der Grad der Formalisierung letztlich von der konkreten Ausgestaltung des Modells ab.

6.2.2.9.7 Formale Umsetzbarkeit (6B)

Trotz der semi-formalen Beschreibung ist es sinnvoll eine Umsetzung in formale Darstellungen nicht vollständig zu vernachlässigen.

Hier ist zunächst die auch als Grundlage anderer Modellierungen verwendete Prädikatenlogik zu betrachten. Die Prädikatenlogik bietet außerdem eine hohe semantische Aussagefähigkeit [Fill09, S.31].

Die graphische Darstellung weist bereits unmittelbar auf die Prädikatenlogik hin. Der hier definierte **ENTITY**-Begriff erlaubt die Übersetzung in die Individualkonstanten. Handelt es sich beispielsweise wegen der Typisierung bereits um Prädikate erster Ordnung, können diese problemlos verwendet werden, da auch die Prädikate als Instanziierungen der Valenzen erlaubt sind. Die **ENTITYs** können somit als Individualkonstanten oder Prädikate betrachtet werden, die Fakten sind immer Prädikate. Einstellige Prädikate sind Prädikatoren der **ENTITYs** und erlauben beispielsweise die Typisierung der Elemente.

Funktoren können die weiteren verwendeten Elemente abbilden. Die Bezeichnung über Qualifier kann über namensbestimmende Funktoren erfolgen, während die **CONSTRAINTs** allgemein auf aussagenbestimmende Funktoren abgebildet werden können. Die Quantoren als Funktoren 2. Stufe sind in der graphischen Darstellung indirekt in der Pfeilrichtung enthalten. Da im CORE-Modell keine automatische Folgerung erfolgen soll, erscheint dies ausreichend, um zumindest prinzipiell eine Übertragung in die Prädikatenlogik zu erlauben.

6.2.2.9.8 Abbildbarkeit in einem Metamodell (6C)

Ein entsprechendes konzeptuelles Modell ist als Basis des Modellbereichs Content im Überblick dargestellt worden (Abbildung 29). Die detaillierte Beschreibung der Typen des Modellbereichs Content sowie aller anderen Typen mit ihrer Umsetzung in ein relationales Schema ist mit Attributen im Anhang A enthalten.

6.2.3 Modellbereich Item

Der Bereich Item beschreibt die Operationalisierung des Contents. Dabei geht es darum, die verschiedenen Aspekte der einzelnen Fakten zu modellieren und die messbaren Aspekte in Form von Items daraus abzuleiten. Diese abgeleiteten Items bilden die Basis für die Gestaltung der Assessments. Der Modellbereich Item knüpft an die **FACTs** des Content-Bereichs an. Die

Beschreibung der Vorgehensweise für die Umsetzung erfolgt im Rahmen des Content Modeling (Abschnitt 7.7).

Die **FACTs** formulieren Aussagen über den Inhalt einer Domäne. Sie bedienen sich dazu der **ENTITYs** und anderer **FACTs**. Für die Umsetzung in konkrete Assessments ist die Frage zu klären, wie ein solcher **FACT** so operationalisiert werden kann, dass die Messung im Assessment Rückschlüsse auf die entsprechende Fähigkeit bezüglich dieses **FACTs** erlaubt. Nimmt man das obige Beispiel

Einstein zog im Jahr 1933 in die USA

so erfordert die Untersuchung auf deren Wahrheitsgehalt, also der Bedeutung, zumindest folgendes Wissen:

- 1) wer mit dem Namen „Einstein“ bezeichnet wird,
- 2) was mit dem Jahr „1933“ bezeichnet wird,
- 3) was mit dem Namen „USA“ bezeichnet wird,
- 4) ob eine Relation $_1$ zog im $_2$ in $_3$ mit den entsprechenden Instanziierungen besteht und
- 5) ob die Relation bereits in der Vergangenheit besteht (und nicht erst in der Zukunft bestehen könnte).

Im Folgenden werden diese einzelnen Aspekte **SNIPPETs** genannt, da es sich sozusagen um kleine „Schnipsel“, Aspekte der Grundaussage handelt, in die diese zerschnitten wird und die für sich überprüfbar sein müssen. Der Inhalt von **SNIPPETs** ließe sich auch in Form von Fakten modellieren, was die Modelle aber verkomplizieren und unübersichtlich werden ließe.

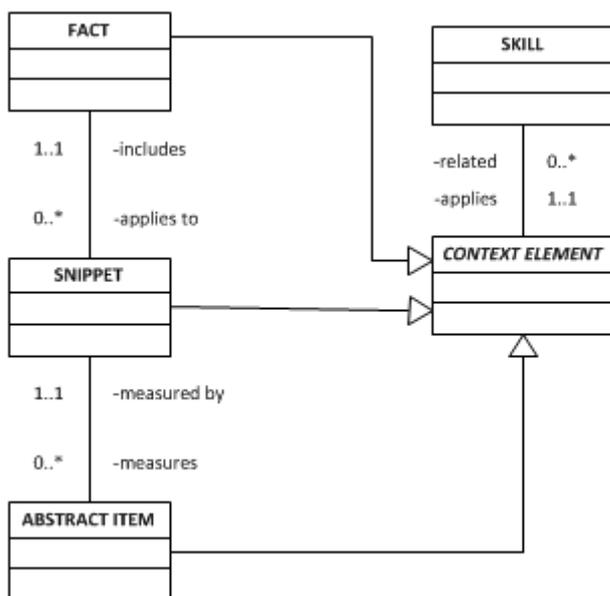


Abbildung 32 Modellbereich Item

Die Ableitung der **SNIPPETs** aus den **FACTs** erfolgt auf Grund fachdidaktischer, kausaler Überlegungen, ebenso deren nächste Verfeinerung. Die tatsächliche Überprüfung der **SNIPPETs** in Assessments wird dann in **ABSTRACT ITEMs** vorbereitet. Der Begriff **ABSTRACT ITEM** wird in Abgrenzung zu den Items im tatsächlichen Assessment verwendet. **ABSTRACT ITEMs** stellen

Vorschläge für Aspekte dar, die in tatsächlichen Items in einer konkreten Situation überprüft werden sollen.

Die **ABSTRACT ITEMS** sind die Basis für die konkrete Konstruktion des Assessments. Alle Elemente vom **FACT** über die **SNIPPET**s zu den **ABSTRACT ITEMS** werden zum **CONTEXT ELEMENT** abstrahiert, das also alle drei Typen repräsentieren kann. Diese Elemente können als Anker für die spätere Bewertung verwendet werden, die dann wahlweise auf allen drei Ebenen, also von der groben Granularität der **FACT**s bis in die feinste Granularität der **ABSTRACT ITEMS** erfolgen kann. Der Anknüpfungspunkt wird hier der **SKILL** sein, der die Handlung in Form von Operatoren mit einbezieht.

Die detaillierte Beschreibung der Typen des Modellbereichs Item mit Attributen ist im Anhang enthalten.

6.2.4 Modellbereich Skill (Fähigkeiten)

Dieser Bereich erweitert den Content um die Fähigkeiten (Skills), die im Zusammenhang mit dem Inhalt beispielsweise in Form von **FACT**s beschrieben, beobachtbar sein sollen. Da wie erläutert, die Messung des Contents nur in der Beobachtung des Einsatzes von Fähigkeiten erfolgen kann, weil ohne beobachtbare Handlung Kompetenzen nicht gemessen werden können, werden dem Content Operatoren zugeordnet, die dessen Verwendung beobachtbar machen.

Die **FACT**s haben eine Bedeutung (meaning) und einen Sinn (sense). Bei der Bedeutung geht es darum, dass Aussagen wahr oder falsch sind. Geht man zunächst von wahren Aussagen aus, so konkretisieren sie in diesem Zusammenhang eine Tatsache des Domäneninhaltes. Eine Person, die den Domäneninhalt beherrscht, sollte auch diese Aussage beherrschen. Das „Beherrschen“ einer Aussage kann verschiedene Aspekte haben, es kann bedeuten, dass eine Person die Aussage

- beschreiben,
- bestimmen,
- ausführen,
- zeigen,
- interpretieren,
- problematisieren,
- ...

kann. Es handelt sich jeweils um aktive Tätigkeiten, die eine Person hinsichtlich dieses **FACT**s beherrschen soll. Ohne mindestens einen der aktiven Operatoren⁷ zu beherrschen, nützt das „Kennen“ eines **FACT**s nichts. Es ist nicht aktivierbar, nicht praktisch nutzbar. Außerdem ist das „Kennen“ nur zu beobachten, wenn mindestens einer der anderen Operatoren angewandt wird, da es sonst für einen außenstehenden nicht erkennbar ist. Das Wissen bleibt implizit (tacit)⁸ und ist somit nicht messbar. Diese die Tätigkeit beschreibenden Verben werden als **OPERATOR** bezeichnet.

⁷ Operatoren haben im pädagogischen Umfeld bei der Frageformulierung insbesondere in zentralen Prüfungen einen wichtigen Stellenwert erreicht, erläutert beispielsweise für das Fach Deutsch (Teachsam 2010).

⁸ Siehe SECI-Modell zu der Bedeutung von implizitem und explizitem Wissen, (Nonaka und Takeuchi 1995).

Es ist also im Rahmen der Handlungskompetenz davon auszugehen, dass stets eine beobachtbare Fähigkeit notwendig ist, um Nutzen zu erzeugen. Diese soll hier als **SKILL** (Fähigkeit) bezeichnet werden. Ein **SKILL** ist immer mit einem **CONTENT ELEMENT** verbunden auf das er sich bezieht. Dies kann beispielsweise ein **FACT** sein. Der **OPERATOR** gibt an, welche Fähigkeit hinsichtlich des **FACT** angesprochen ist.

Beispiele:

BESCHREIBE, Gleiche Veränderung pro Periode beschreibt linearen Prozess.
oder

INTERPRETIERE, Gleiche Veränderung pro Periode beschreibt linearen
Prozess.

An den Beispielen sieht man, dass die Fähigkeit als Verb angegeben ist. Es ist zu unterscheiden von dem Verb, das im Fakt selbst verwendet wird. Das Verb im **SKILL** wird über den **OPERATOR** systematisiert.

Man kann auch sagen, dass die Aussagen(**FACTs**) den Inhaltsaspekt der Kompetenzen beschreiben, während über die Fähigkeiten der Verhaltensaspekt betont wird. Beides ist im Model getrennt modelliert (Schott und Ghanbari 2008, 85).

Die **OPERATORS** ihrerseits sind zumeist mit bestimmten Niveaus verbunden. Klassisch sind dies die Anforderungsbereiche I bis III. Im Sinne dieses Modells handelt es sich bei ihnen um **COMPETENCE NIVEAUs**. Durch die Verwendung eines bestimmten **OPERATORS** wird der Fähigkeit ein Kompetenzniveau zugeordnet. Das bedeutet, dass man erwartet, dass wenn Aufgaben gelöst werden, die diese Fähigkeit erfordern, das entsprechende Kompetenzniveau vorliegt. Derartige deterministische Zusammenhänge sind in der Praxis schwierig zu isolieren. So wird hierfür beispielsweise ein stochastisches Modell vorgeschlagen (Beyen 2013, 282). Auch wenn der **OPERATOR** als Verb angegeben wird, muss er nicht wörtlich im Assessment verwendet werden. Wesentlich ist vielmehr, dass er im Assessment sinngemäß verwendet wird.

Ein im Assessment für einen Assessment-Teilnehmer gemessener **SCORE** kann über das **SCORING** im noch zu beschreibende **SITUATION**-Modell einem bestimmten **SKILL** zugeordnet werden. Das bedeutet, dass der Assessment-Teilnehmer bezüglich dieses **SKILLS** eine bestimmte, deterministisch bestimmte Punktzahl (absolut und relativ) erreicht hat. Fasst man alle **SKILLS** eines **CONTEXT ELEMENTs** zusammen, kann eine Aussage über dieses **CONTEXT ELEMENT**, beispielsweise einen **FACT** getroffen werden. Er kann auch genutzt werden, um über **SKILL_INDICATING** mittels der **COMPETENCE NIVEAU METRICS** die Zuordnung eines **SPECIFIC COMPETENCE NIVEAUs** für einen Teilnehmer durchzuführen, das heißt, ein bestimmtes Kompetenzniveau hinsichtlich einer bestimmten Kompetenz wird einem bestimmten Assessment-Teilnehmer zugeordnet. Diese Zuordnung kann zusätzlich mittels **CNM AGGREGATE** hierarchisch geschachtelt werden.

In den **COMPETENCE NIVEAU METRICS** können beispielsweise die in Abschnitt 7.16 beschriebenen Competence Niveau Maps abgelegt (CNM) werden. Bei ihnen handelt es sich um eine Folge von Niveaustufen und Schwellenwerten (**LEVEL INDICATION**), deren Überschreiten das Erreichen eines bestimmten Niveaus anzeigt. In Abbildung 33 ist eine solche Competence Niveau Map schematisch dargestellt. Dabei werden ordinal geordnete Stufen, die inhaltlich beschrieben werden und Schwellenwerte im Intervall [0,1] für deren Erreichen angegeben. Diese Schwellenwerte werden mit den pro Person erreichten Scores verglichen und daraus das Niveau bestimmt.

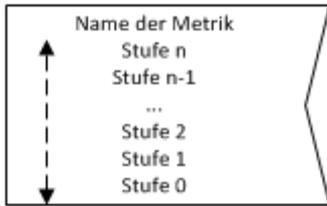


Abbildung 33 Competence Niveau Map (CNM) schematisch

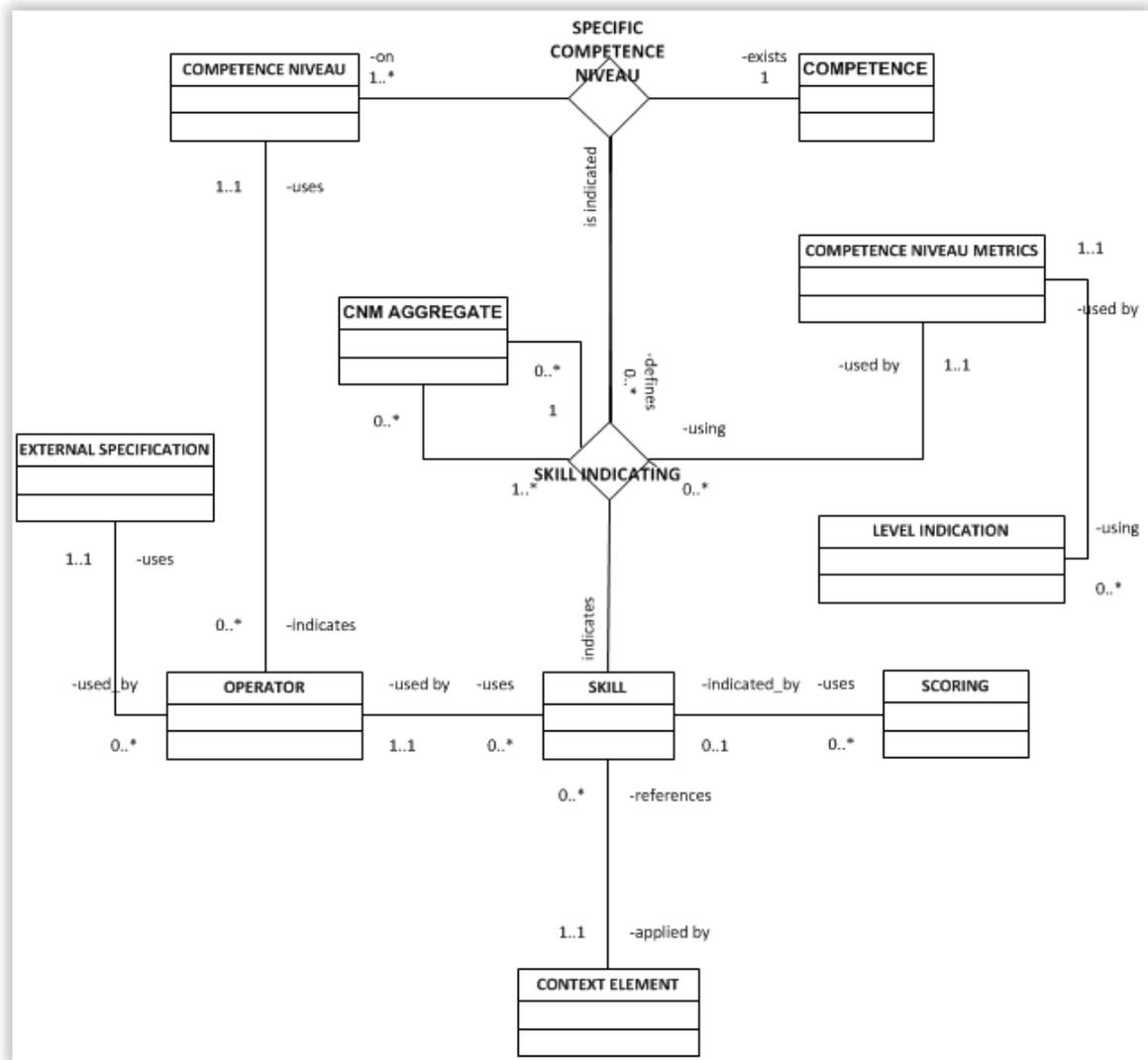


Abbildung 34 Modellbereich Skill

Bei der Berechnung können unterschiedliche SKILLS berücksichtigt werden, die über verschiedene Competence Niveau Maps auf Kompetenzniveaus abgebildet werden, wobei die erreichten Bewertungen mittels Gewichten für die Skills oder mittels eines anderen Algorithmus bestimmt werden können. Die Vorgehensweise ist in Abschnitt 7.8 im Rahmen des Vorgehensmodells beschrieben.

Die detaillierte Beschreibung der Typen des Modellbereichs Skill mit Attributen ist im Anhang enthalten.

6.2.5 Modellbereich Competence Specifications

Die Kompetenzen entstehen in der Regel aus systemexternen Anforderungen. Aus dem Bereich der Bildung wären hier Curricula, einheitliche Prüfungsanforderungen oder Bildungsstandards zu nennen. Diese haben in den letzten Jahren einen Weg von inhaltlichen zu kompetenzorientierten Beschreibungen genommen, beinhalten aber letztlich eine Mischung aus beiden, da sich die Kompetenzen nicht vollständig von den Inhalten trennen lassen. Diese Beschreibungen stellen aus Sicht des CSP die Anforderungen an die Assessments dar und sind im Bereich der Competence Specifications zu dokumentieren.

Die Modellierung der Kompetenzen wird durch die Beschreibung der zu modellierende Domäne bestimmt. Im hier vorliegenden Modell wird davon ausgegangen, dass für die Domäne definierte Kompetenzen vorliegen, die allerdings im Allgemeinen nicht detailliert spezifiziert sind. Typische Beispiele findet man in der Pädagogik in den Curricula, beispielsweise im Kerncurriculum Mathematik für die niedersächsischen Gymnasien (Kultusministerium 2009). Es muss also anknüpfend an diese aus Modellsicht extern vorgegebenen Spezifikationen (**EXTERNAL SPECIFICATION**) modellintern eine Detaillierung der Kompetenzen erfolgen, die nach einem einheitlichen Muster gestaltet werden soll. Das CSP unterscheidet daher einen externen und einen internen Modellbereich.

Aus den **EXTERNAL SPECIFICATION**s ergeben sich die externen Ziele (**EXTERNAL GOAL**) für die Domäne. Generell wird hier zwischen Kompetenzen und Qualifikationen unterschieden, wie im Abschnitt 3.3 erläutert. So wird sowohl im schulischen als auch im beruflichen Bereich vielfach von Kompetenzen gesprochen, während aus der unmittelbaren beruflichen Sicht von Qualifikationen gesprochen wird. Im Rahmen der Aufgabenstellung wurde auch dargestellt, dass im CSP zunächst auf die Kompetenzen eingegangen wird, während die Qualifikationen nur im Ausblick erwähnt werden.

Der Modellbereich der Competence Specifications modelliert die Gewinnung der zu messenden Kompetenzen aus den extern spezifizierten Kompetenzen. Im Bereich der Schulen sind dies die Standards die von der Kultusministerkonferenz, den Kultusministerien und deren untergeordneten Stellen und Gremien erarbeitet werden. Zentral für die Kompetenzorientierung sind beispielsweise die Bildungsstandards (Kultusministerkonferenz 2012). Typische Materialien für den hier beispielhaft untersuchten Bereich der gymnasialen Oberstufe in der beruflichen Bildung sind die Einheitlichen Prüfungsanforderungen für das Abitur (EPA) sowie die Kerncurricula (KC). Beide werden fachspezifisch ausgeprägt.

Gemeinsam ist allen Standards, dass sie in

- Inhaltliche Standards (Content standard oder **CURRICULUM STANDARD**),
- Standards für Lehr- und Lernbedingungen (**OPPORTUNITY-TO-LEARN-STANDARD**) und
- Leistungs- oder Ergebnisstandards (performance oder **OUTPUT STANDARD**)

sowie Niveauanforderungen unterschieden werden (Bildungsstandards der Kultusministerkonferenz 2005, 8).

Das konzeptionelle Modell beinhaltet daher die Möglichkeit, die entsprechenden Standards zu hinterlegen. Die Bildungsstandards bilden einen national und international vergleichbaren Rahmen für die Ausgestaltung der Bildungsziele und gleichzeitig für die Schulen und Lehrkräfte

einen Orientierungsrahmen. Sie überlassen andererseits die konkrete Ausgestaltung den Schulen mit ihren zuständigen Teams und Konferenzen.

Diese Standards definieren zu erreichende Ziele (EXTERNAL GOAL) die sich wiederum kategorisieren lassen. Im Modell sind beispielhaft Ziele hinsichtlich der zu erreichenden Kompetenzen (COMPETENCE GOAL), der zu erreichenden Qualifikationen (QUALIFICATION GOAL) und der praktischen Umsetzbarkeit (UBIQUITY GOAL) definiert worden. Sie lassen sich bei Bedarf entsprechend erweitern.

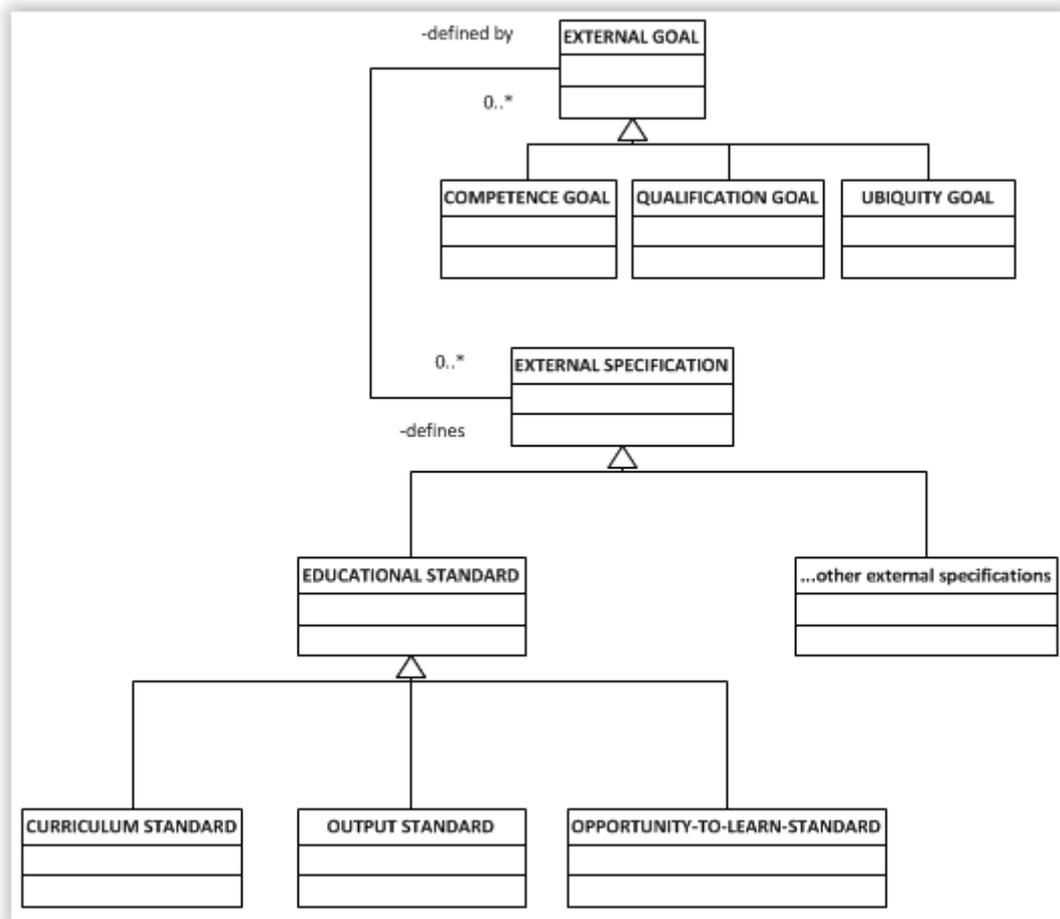


Abbildung 35 Modellbereich Competence Specifications

Die detaillierte Beschreibung der Typen des Modellbereichs Competence Specifications mit Attributen ist im Anhang enthalten.

6.2.6 Modellbereich Kompetenzdimension

Der nächste Bereich des Modells dient der modellinternen Modellierung der Kompetenzen mit ihren Dimensionen. Für die Definition von Kompetenzen ist typisch, dass Strukturierungen wie die in Selbstkompetenz, Sachkompetenz und Sozialkompetenz vorgeschlagen werden. Weitere Kompetenzarten wurden in Abschnitt 3.3.1 angesprochen.

Wenn diese Aufteilungen auch durchaus kontrovers sind, so ist die weitere Detaillierung von Kompetenzen in der Literatur durch hierarchische Begriffsgliederungen nach unterschiedlichsten Aspekten, die sich allerdings oft überlappen, ein charakteristisches Merkmal.

Ein anderer Aspekt ergibt sich aus der Betrachtung der Reichweite von Kompetenz. Im Wesentlichen ist dabei zu unterscheiden, ob es sich um domänenspezifische oder verallgemeinerte, nur domänenverbundene Kompetenzen (auch Schlüsselqualifikationen) handelt. So lassen sich

- domänenspezifische Kompetenzen (conceptual),
- bereichsübergreifende Kompetenzen (beispielsweise Problemlösekompetenz) (procedural) und
- fächerübergreifende Kompetenzen (beispielsweise das Selbstkonzept oder die Motivation) (utilizational)

unterscheiden (Winther 2010, 30ff.).

Die Schwierigkeit nicht domänenspezifische Kompetenzen zu messen wurde bereits angesprochen. Konzentriert man sich daher zunächst auf diese Kompetenzen, bedeutet das in der Konsequenz, dass die definierte und gemessene Kompetenz an einen Kontext gebunden ist. Die Schlussfolgerung ist: „Kompetenzen sollten sich auf spezifische Kontexte beziehen, deren Handlungsspielraum möglichst klar definiert werden kann.“ (Winther 2010, 36). Wenn hinreichend genaue Anforderungen in verschiedenen Situationen beschrieben werden sollen, muss es entsprechend genaue Vorstellungen von den Prozessen zum Bearbeiten und Lösen dieser Anforderungssituationen geben [ebd.]. Dies ist wiederum ein klares Argument dafür, im hier beschriebenen Kontext zunächst von domänenspezifischen Kompetenzen auszugehen.

Allerdings sollten auch übergreifende oder sich in verschiedenen Dimensionen in ähnlicher Form wiederholende Kompetenzdimensionen einfach abgelegt werden können. Oft kann von einer Untergliederung in Sach-, Methoden-, Sozial-, und Selbstkompetenz oder einer ähnlichen Struktur ausgegangen werden. Die Sachkompetenz kann unter anderem eine deskriptive, eine prozedurale und oder eine andere Kompetenz beinhalten. Diese Aspekte werden im Typ **ABSTRACT DIMENSION** mit den entsprechenden Untergliederungen abgebildet. Dieser Typ dient als „Überbau“ zwischen den allgemeinen in der Theorie formulierten Dimensionen und den konkreten Spezifikationen.

Somit stellt sich die Frage, wie domänenspezifische Dimensionen im CSP abgelegt werden können. Aus den Überlegungen zur Umsetzung der Bildungsstandards sind beispielsweise das Kompetenzmodell und die Kompetenzmatrix von Klinger als multidimensionale Ansätze entstanden. Dieses multidimensionale Modell zeigt gut die Anforderungen, die sich an die interne Kompetenzabbildung im CSP ergeben. Wiederum zurückgreifend auf die Kompetenzdefinition von Weinert leitet er zwei zentrale Dimensionen ab (Klinger 2005, 137):

- Wissen, kognitive Strukturen, Einstellungen, Haltungen, Einsicht und
- Können, Handeln, operationale Ebene, konkretes Tun, Aktivitäten

Beiden Dimensionen sind zunächst keine Hierarchien oder einzelne Dimensionselemente zugeordnet. Die weiteren Überlegungen, basierend auf nationalen Bildungsstandards für Physik und Deutsch sowie der Ansatz zu einer Verallgemeinerung speziell für die Naturwissenschaften führen für die Handlungsdimension zu den Elementen (Klinger 2005, 139)

- Erkenntnisgewinnung, Anwendung, Kommunikation und Bewerten,
- Deutsch: Sprache und Sprachgebrauch untersuchen, Sprechen und Zuhören, Schreiben, Lesen – mit Texten und Medien umgehen und
- Mathematik, sechs allgemeine mathematische Kompetenzen und die Leitideen.

Von hier kann man in das weit verbreitete Konzept der Operatorennutzung übergehen.

In seinem Kompetenzmodell stellt Klinger die Handlungsdimension in einer Matrix einer weiteren Dimension gegenüber, dem Anforderungsbereich. Diese Dimension in Form von Kompetenzstufen (Anforderungsbereiche) beinhaltet die bekannten Dimensionselemente

- Anforderungsbereich I
- Anforderungsbereich II
- Anforderungsbereich III

Schließlich wird die Dimension Wissen in

- Alltagswissen
- Praktisch-technisches Wissen
- Fachwissen
- Metawissen

unterteilt. Die einzelnen Dimensionselemente werden mit Hilfe von beispielhaften Formulierungen verdeutlicht. Das zeigt, dass auch Versuche existieren mehrere, zueinander orthogonale Dimensionen zu verwenden. Die gängigen hierarchischen, baumähnlichen Strukturen sind im Allgemeinen in ihrer Struktur eindimensional, da sie jeweils „nur“ weiter verfeinert werden. Der Ansatz von Klinger ist dagegen ein Beispiel, das zeigt, dass auch mehrdimensionale Strukturen und innerhalb der Dimensionen Dimensionsstrukturen mit mehreren Ebenen modelliert werden können. Zu beachten ist allerdings, dass mit dem Anforderungsbereich eine wertende hierarchische Struktur vorliegt, die eher auf unterschiedliche Kompetenzniveaus hindeutet.

Die Multidimensionalität zeigt sich auch in den erläuternden Aspekten mit denen Kompetenzen dargestellt werden. Die Aspekte (und eigentlich Dimensionen) der Selbstkompetenz oder personalen Kompetenz (Edelmann und Tippelt 2007, 133) werden beispielsweise charakterisiert durch (Achtenhagen und Baethge 2007, 57)

- Einstellungen und Werte,
- Wahrnehmungen, Anreize,
- Motivation,
- Metakognition und
- Lernstrategien.

Beginnt man diese Aspekte jeweils zu strukturieren, so ergibt sich eine neue Menge von Kompetenzdimensionen, die in sich zu strukturieren sind. Es entsteht in jedem Fall eine mehrdimensionale Struktur.

Im hier vorgeschlagenen Modell werden zur Abbildung und zusammenfassenden Analyse Dimensionsstrukturen verwendet, die die Freiheit zur Abbildung der jeweils relevanten Aspekte bieten. Damit kann das Modell dem jeweils relevanten Blickwinkel angepasst werden.

Im Folgenden werden exemplarisch einige Dimensionsstrukturierungen aufgezeigt, ohne dass die Begriffe erschöpfend erarbeitet und abgegrenzt werden. Es soll auch keine Bewertung der verschiedenen Dimensionen vorgenommen werden, sondern nur deren Abbildung gezeigt werden.

Generell werden die domänenspezifischen Dimensionen aus den in den **EXTERNAL SPECIFICATIONS** definierten Kompetenzen als Kompetenzdimensionen **COMPETENCE DIMENSION** übernommen und können mit den **ABSTRACT DIMENSIONS** verbunden werden.

Sie bilden so den Übergang in den Bereich des internen Kompetenzmodells. Der Startpunkt der Dimensionshierarchien **COMPETENCE DIMENSION** entspricht zumeist einer extern vorgegebenen Kompetenz.

Der Bereich **COMPETENCE DIMENSION**, **COMPETENCE LEVEL** und **COMPETENCE** beschreibt die Dokumentation der dem Modell zugrunde gelegten internen Kompetenzstrukturen. Entsprechend den unterschiedlichen Vorstellungen von Kompetenzhierarchien muss eine Modellierung unterschiedlicher Vorstellungen erlaubt sein. Grundsätzlich werden Kompetenzen jeweils als Hierarchien modelliert. Mehrere Kompetenzen bilden mehrere Hierarchien.

Dabei werden sowohl Hierarchien ohne feste Ebenenstruktur als auch mit Ebenenstruktur erlaubt, auch wenn dies zusätzliche Kontrollen bei der Modellierung erfordert, um keine Widersprüche zwischen der Struktur der Einzelkompetenzen und der Ebenenstruktur zu dokumentieren. So ist eine Hierarchie innerhalb der **COMPETENCES** möglich, ohne einen Bezug zum **COMPETENCE LEVEL** herzustellen. Diese Hierarchie wird immer gebraucht, um die Struktur zwischen den einzelnen Kompetenzen zu dokumentieren. Es ist auch möglich, zusätzlich eine Hierarchie der **COMPETENCE LEVEL** aufzubauen und die **COMPETENCES** jeweils einer Ebene zuzuordnen. Die Dimensionsebenen sollen dann jeweils „gleichwertige“ Kompetenzen zusammenfassen.

Das folgende Beispiel zeigt die dem Kerncurriculum Mathematik (Kultusministerium 2009, 13) entnommenen externen Kompetenzen C109 und C110 für die Einführungsphase des Beruflichen Gymnasiums Wirtschaft.

Die **COMPETENCE DIMENSION** „Inhaltsbezogene Kompetenzen Mathematik für das Fachgymnasium Niedersachsen“ kann unter anderem untergliedert werden in die **COMPETENCE**:

(C109) Die SuS bestimmen eine Funktionsgleichung aus gegebenem Graphen für Potenz- und Exponentialfunktion in der Form $y=a*f(b*x+c)+d$

und

(C110) Die Schülerinnen und Schüler grenzen in Sachzusammenhängen lineares und exponentielles Wachstum gegeneinander ab, auch unter Verwendung der eingeführten Technologie.

Bis hierhin ist die Struktur durch das Kerncurriculum vorgegeben. Unterhalb dieser Kompetenzen können weitere Untergliederungen innerhalb des Modells erfolgen, die nicht mehr mit der externen Spezifikation verknüpft sind.

Daraus ließe sich beispielsweise die folgende unterhalb der Kompetenz C110 angesiedelte Hierarchie ableiten:

1. Lineares und exponentielles Wachstum (C110)
 - 1.1. Zuwachs/Abnahme beinhalten dieselbe Logik
 - 1.2. Wachstum wird durch Veränderung gegenüber einem Anfangswert, einem Bestand oder einem noch fehlenden Bestand beschrieben.
 - 1.2.1. Lineares Wachstum bezieht sich auf einen Anfangsbestand
 - 1.2.2. Exponentielles Wachstum bezieht sich auf den aktuellen Bestand
 - 1.2.3. Lineares Wachstum wird durch konstante Veränderung pro Zeiteinheit beschrieben
 - 1.2.4. Exponentielles Wachstum wird durch gleiche relative (prozentuale) Veränderung beschrieben
 - 1.3. Wachstum kann in verschiedenen mathematischen Funktionstypen beschrieben werden
 - 1.3.1. Die Funktionstypen beinhalten Aussagen wozu die Veränderung gemessen wird
 - 1.3.2. Proportionalität kann in einer Differenzialgleichung verdeutlicht werden
 - 1.3.3. Die Parameter der Funktionen haben logische Bedeutungen für das Wachstum

Es existiert also ein Verfeinerungsbaum auf Ebene der **COMPETENCE** bis hin zu den detailliertesten auszumodellierenden Einzelkompetenzen.

Die **COMPETENCE LEVEL** können entsprechend den Gliederungsebenen modelliert werden. Im oben angegebenen Beispiel gibt es drei Kompetenzebenen. Sie sollen mit 1), 2) und 3) bezeichnet werden und sie entsprechen der Gliederungstiefe. Jede Kompetenzebene ist mit den umgebenden Kompetenzebenen verbunden. Die eigentlichen Gliederungspunkte sind die Kompetenzen, die mit der jeweiligen Ebene verknüpft sind. So ist beispielsweise der Punkt „1.2.1 Lineares Wachstum bezieht sich auf den Anfangsbestand“ eine Instanz von **COMPETENCE**. Sie ist verbunden mit der übergeordneten Instanz „1.2 Wachstum wird durch Veränderung gegenüber einem Anfangswert, einem Bestand oder einem noch fehlenden Bestand beschrieben.“. Sie ist außerdem verbunden mit dem Competence Level „3“, da sie der dritten Stufe angehört, wenn man davon ausgeht, dass C110 der ersten Stufe entspricht.

Die **COMPETENCES** der untersten Detaillierungsstufe werden durch konkrete **FACTs** repräsentiert, deren Beherrschung als Indikator für die verfügbare Kompetenz angesehen wird. Diese **FACTs** werden also basierend auf den Kompetenzen formuliert und so die Beziehung zum Modellbereich Content hergestellt. Das CSP-Modell erlaubt die Zuordnung verschiedener Fakten zu einer Kompetenz und umgekehrt mehrerer Kompetenzen zu einem Fakt. Beispielsweise kann der Kompetenz „Die Schülerinnen und Schüler grenzen in Sachzusammenhängen lineares und exponentielles Wachstum gegeneinander ab, auch unter Verwendung der eingeführten Technologie.“ (Lineares und exponentielles Wachstum, C110) eine Reihe von Fakten wie F1103, F1104 oder F1107 zugeordnet werden, die das Verständnis für die Basis des Wachstums (Proportionalität) beinhalten. Auch die Zuordnung der mathematischen Funktionstypen (Fakten F1105 und F1106) oder von Anwendungsbereichen (beispielsweise F1111) ist sinnvoll. Bei dieser Zuordnung handelt es sich zunächst um rein fachpädagogische, kausale Entscheidungen. Damit sind auch mehrdimensionale Kompetenzbeschreibungen möglich, indem die entsprechenden Kompetenzen aus verschiedenen Dimensionen mit einem **FACT** verbunden werden.

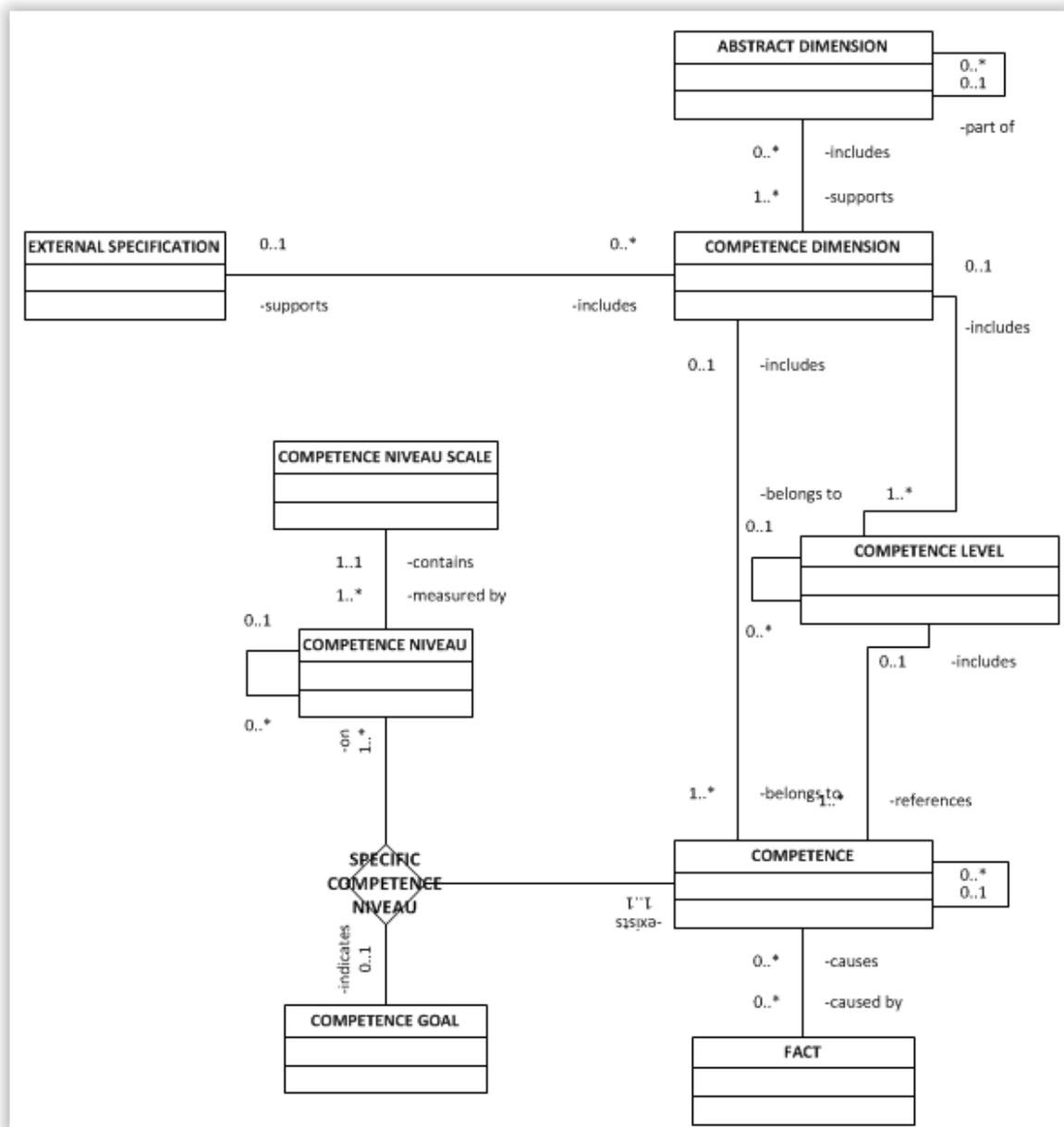


Abbildung 36 Modellbereich Competence Dimension

Ein anderer Aspekt ist der des Kompetenzniveaus. Kompetenzniveaus sind zumeist ordinal gemessene Niveaustufen (Kompetenzstufen) innerhalb einer Dimension. Diese Niveaus können unabhängig von den Kompetenzen definiert werden. Ein typisches Beispiel sind die Anforderungsniveaus I bis III⁹. In diesem Fall ist „Anforderungsniveau“ eine Instanz des Typs **COMPETENCE NIVEAU SCALE**, während die drei Niveaus Instanzen des **COMPETENCE NIVEAU** sind. Die hierarchische Beziehung beschreibt den Zusammenhang der drei Niveaus. Diese **COMPETENCE NIVEAU**s (Kompetenzniveaus) gibt es in unterschiedlichen Ausprägungen (Saldern 2011, 162ff.), was die Modellierung der entsprechenden Typen im CSP zur gleichberechtigten Abbildung dieser verschiedenen Niveauskalen rechtfertigt. Da in jeder Skala jede

⁹ Diese klassische Einteilung ist nicht unbedingt mit dem Kompetenzbegriff verbunden. Sie kann auch auf Qualifikationsniveaus oder andere Einteilungen bezogen werden.

Kompetenzstufe das Erreichen der untergeordneten Stufen implizieren soll, ergibt sich die hierarchische Beziehung.

Durch die Modellstruktur können auch mehrere Skalen nebeneinander für unterschiedliche Niveauhierarchien abgebildet und genutzt werden, indem verschiedene **COMPETENCE NIVEAU SCALEs** mit ihren jeweils zugehörigen **COMPETENCE NIVEAU**s dokumentiert werden. So könnten auch die oft vernachlässigten anderen Aspekte einbezogen werden, so der affektive, kommunikative oder psychomotorische, die auch in der Schule verwendet werden könnten (Saldern 2011, 166). Eine gute Übersicht mit sieben nebeneinander betrachteten sogenannten Kategorien, die auch eine mehrdimensionale Zuordnung erlauben findet man beim Institut für Erziehungswissenschaften der Universität Tübingen (Universität Tübingen 2012).

Durch die Kombination eines **COMPETENCE NIVEAU** mit einer **COMPETENCE** ergibt sich ein **SPECIFIC COMPETENCE NIVEAU**. Gemeint ist damit ein auf eine konkrete Kompetenz bezogenes Niveau, beispielsweise wird Anforderungsbereich II für die Kompetenz C110 (Lineares und exponentielles Wachstum) definiert. Im **SPECIFIC COMPETENCE NIVEAU** ist die Verknüpfung zwischen Kompetenzstruktur- und Kompetenzniveaumodell realisiert. Dadurch entsteht eine weitere mehrdimensionale Matrixbeschreibung, die Kompetenzdimensionen und Kompetenzniveauskalen kombiniert.

Zusätzlich kann das Erreichen eines **SPECIFIC COMPETENCE NIVEAU** mit dem Erreichen eines kompetenzorientierten Ziels (**COMPETENCE GOALS**) verbunden werden. Damit können umgekehrt diese Ziele auch auf das Erreichen bestimmter Niveaus in bestimmten Kompetenzdimensionen bezogen auf bestimmte Kompetenzen definiert werden.

Die detaillierte Beschreibung der Typen des Modellbereichs Competence Dimension mit Attributen ist im Anhang A enthalten.

6.2.7 Modellbereich Feedback

Ein weiterer wichtiger Aspekt ist die Modellierung des Feedbacks. Schließlich stellt die Rückkopplung zum Bildungsprozess selbst einen wesentlichen Nutzen dar, da so konstruktive Wege für eine Behebung von Defiziten in der betrachteten Domäne beziehungsweise der nutzbringenden Handlung in diesem Bereich aufgezeigt werden können. Der Bildungsprozess selbst ist im CSP nur im Ansatz modelliert, da er nicht das primäre Ziel ist. Gleichwohl kann ohne eine gewisse Strukturierung kaum ein gezieltes Feedback erreicht werden. Es ist ausdrücklich nicht das Ziel, die Qualität eines solchen Kompetenzbildungsprozesses zu beschreiben oder gar zu bewerten. Ein allgemeiner Kompetenzbildungsprozess wird hier als **COMPETENCE CREATION PROCESS** bezeichnet.

Eine klassische Spezialisierung des **COMPETENCE CREATION PROCESS** in der Bildung wird als **EDUCATION COURSE** bezeichnet und ist als eine neben anderen möglichen (andersartigen) Ausprägungen eines solchen Prozesses angelegt. Er besteht aus einem oder mehreren **EDUCATION MODULEs**. Entsprechend den verschiedenen Ansätzen der Pädagogik können diese Module an unterschiedlichen Gesichtspunkten und Zielvorstellungen orientiert sein. Eine aktuelle Form sind beispielsweise Lernsituationen. Diese Module können auch als Ausprägungen der **STEPs** des klassischen Prozessmodells aufgefasst und dementsprechend in unterschiedlichen Reihenfolgen angeordnet werden. Dies beinhaltet, dass die Reihenfolge der Module innerhalb des

Kurses variieren kann, aber nicht muss. In jedem Modul kann **EDUCATION MATERIAL** verwendet werden.

Die Ziele der **EDUCATION MODULES** können durch verschiedene Beziehungen beschrieben werden, insbesondere durch Verknüpfungen zu den entsprechenden **EXTERNAL GOALS**, zu den **SPECIFIC COMPETENCE NIVEAUs**, die erreicht werden sollen und am konkretesten zu den **CONTEXT ELEMENTS**.

Wichtig im Sinne eines Feedbacks ist es, diesen Prozess mit den Ergebnissen eines Assessments so zu verbinden, dass sich aus dessen Ergebnissen Rückschlüsse und zielgerichtete Maßnahmen ableiten lassen. Dies kann teilweise vor der Konstruktion des Assessments geschehen, indem mögliche Antworten der Assessment-Teilnehmer antizipiert und die entsprechenden Ergebnisse bestimmten Fehlertypen (**ERROR TYPEs**) zugeordnet werden. Ein solcher **ERROR TYPE** kann gedanklich bereits aus einem **FACT** oder einer Fähigkeit bezüglich **FACTs** abgeleitet werden.

Wichtiger für das CSP ist der Bezug zu **SPECIFIC COMPETENCE NIVEAUs**. Hier können die für das Erreichen eines bestimmten Niveaus relevanten Feedbacks dokumentiert werden. Diese Festlegung kann im Vorfeld als kausale Ableitung aus der Modellierung erfolgen. Sie kann aber insbesondere auch nach der Durchführung eines Assessments durch die Analyse der mit den entsprechenden Kompetenzen verbundenen Items erfolgen. Im Ergebnis wird ein Zusammenhang zwischen dem hinsichtlich bestimmter Skills erwarteten Ergebnis und den spezifischen Kompetenzniveaus dokumentiert, die ihrerseits bestimmte Niveaustufen hinsichtlich konkreter Kompetenzen repräsentieren. Durch die Verbindung dieser Niveaus mit den typisierten Feedbacks können entsprechend dem bezüglich einer bestimmten Kompetenz erreichten Niveau zielgerichtete Feedbacks festgelegt werden.

Dann ist es auch möglich, hieraus ein mögliches **FEEDBACK** abzuleiten und **FEEDBACK MATERIAL** bereitzustellen. Entsprechendes gilt für die erst nach der Durchführung des Assessments definierten Antworten. Damit entstehen sowohl Informationen für den Kursleiter als auch für die Assessment-Teilnehmer, denen zusätzliches **FEEDBACK MATERIAL** zur Verfügung gestellt werden kann, das nicht unmittelbar mit dem Prozess verknüpft sein muss. So entsteht ein lernender Kreislauf im Bildungsprozess.

Der **ERROR TYPE** beschreibt damit eine typisierte Fehlersituation, die auf eine zugrundeliegende Ursache zurückgeführt wird. Das **FEEDBACK** beschreibt diese Ursache. Dabei kann auch auf die betroffenen **SKILLS** und **CONTEXT ELEMENTS**, also **FACTs**, **SNIPPETs** und **ABSTRACT ITEMs** Bezug genommen werden.

Aus dem **FEEDBACK** können mögliche **ACTIONs** abgeleitet werden, also Unterstützungsmaßnahmen für die Lernenden und die Lehrenden. Diese Aktionen können ihrerseits durch **FEEDBACK MATERIAL** unterstützt werden.

Die detaillierte Beschreibung der Typen des Modellbereichs Feedback mit Attributen ist im Anhang enthalten.

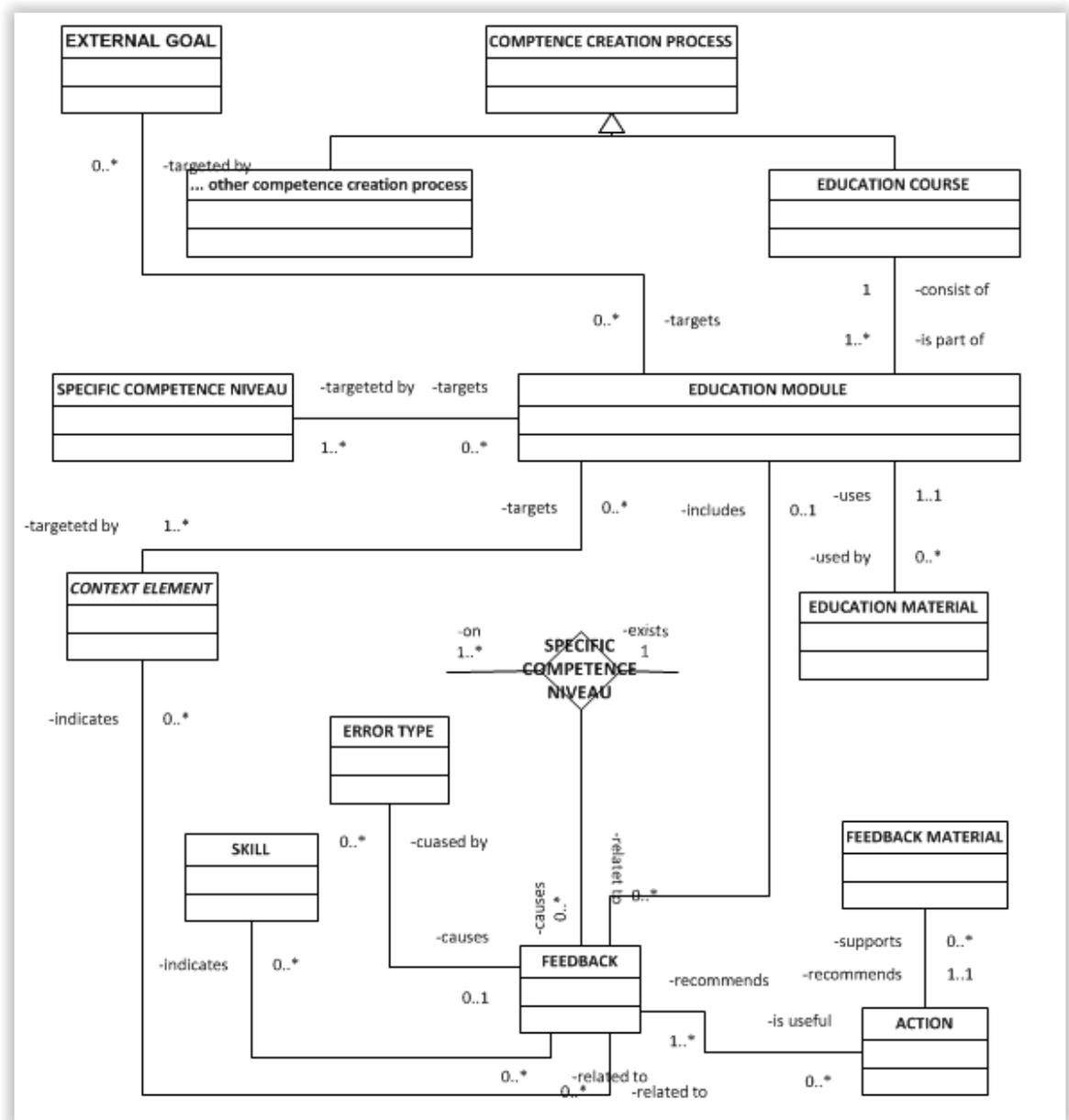


Abbildung 37 Modellbereich Feedback

6.3 Das Situation-Modell

Das SITUATION-Modell ist die zweite Ebene des CSP, also das „S“, wie es in Abbildung 20 dargestellt ist. Dieses Modell beschreibt die Ebene des Assessments, also die Konstruktion einer Situation, die der Leistungsüberprüfung dient oder die Handlungsfähigkeit in einer konkreten Situation beschreibt. Seine Grundeinheit ist somit zunächst das Assessment, nicht die Assessment-Durchführung. Das Modell erlaubt auch die Dokumentation der mehrfachen Durchführung desselben Assessments wie auch die mehrfache Nutzung von Teilen in verschiedenen Assessments. Somit ist hier ein Pool-Gedanke realisiert, der eine Wiederverwendbarkeit einmal konstruierter Items erlaubt.

Es wird zwischen den Bereichen Assessment, Score und Statistik unterschieden. Der Modellbereich Assessment stellt den Kern des eigentlichen Assessments dar. Hier wird das Assessment als solches mit allen Items beschrieben, also die Aufgabenstellung. Der Modellbereich Score enthält

die Beobachtungen in Form von Antworten (Outcoming space) und Bewertungsvorschriften mittels derer ein Bezug zum Erfolg einer Person hergestellt werden kann. Er stellt also gewissermaßen einen erweiterten Erwartungshorizont dar. Der Modellbereich Statistik stellt den Bezug zur statistischen Kompetenzmessung mit Beschreibung der entsprechenden statistischen Analysemodelle und deren Parametern her. Der Modellbereich Situation process model stellt eine Idee zur Erweiterung von Assessments um die Beschreibung von prozessorientierten Aufgabenstellungen und den damit verbundenen Entscheidungen dar, der im Rahmen dieser Arbeit aber nicht vertieft wird.

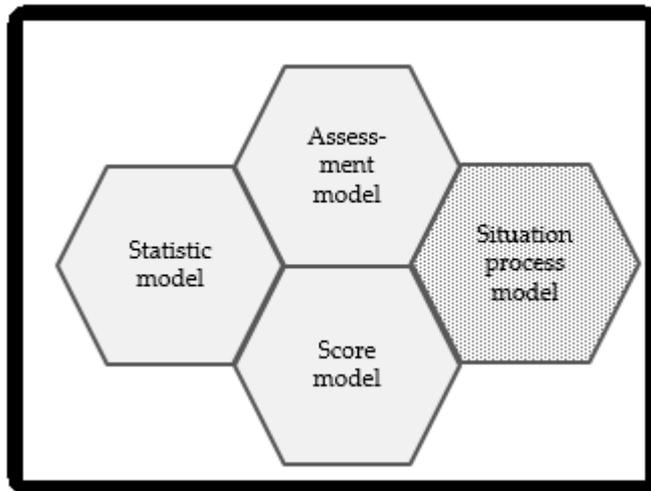


Abbildung 38 Bereiche des SITUATION-Modells

6.3.1 Modellbereich Assessment

Der Modellbereich beinhaltet das **SETTING** als Grundlage der Beschreibung eines Assessments und als Anker für die weiteren Bestandteile der Dokumentation eines Assessments. Der Begriff wurde abstrakter gewählt als der Begriff Assessment, damit er nicht sofort mit klassischen Klausuren oder Fragebögen assoziiert wird, auch wenn diese oft damit beschrieben werden. Diese Begriffe werden verwendet, um das Verständnis intuitiver zu gestalten. Ein **SETTING** kann aus mehreren „Fragebögen“ **QUESTIONNAIRES** bestehen, wenn diese beispielsweise als Variante A und B für verschiedene Lernende erstellt werden.

Die **SETTING EXECUTION** beschreibt die konkrete Durchführung eines Settings, also des Assessments. Es wird davon ausgegangen, dass dasselbe Assessment mehrmals durchgeführt werden kann. Dann wird für jede Durchführung eines Assessments, genauer für jede Nutzung eines **QUESTIONNAIRES**, da hier spezifische Probleme enthalten sein können oder Parameter zu beachten sind, eine **SETTING EXECUTION** angelegt.

Die **EXECUTION PARAMETERS** beinhalten Hinweise zur organisatorischen Gestaltung des Assessments also zu dessen räumlicher und zeitlicher Durchführung, der Art der Präsentation und weitere Komponenten. Es handelt sich also um planbare und geplante Randbedingungen. Sie sind von der **SETTING EXECUTION** abgespalten, um sie in verschiedenen **SETTING EXECUTIONS** wiederverwenden zu können und um sie bei Bedarf weiter ausmodellieren zu können. Entsprechendes gilt für die Qualitätsbeschreibungen zur Objektivität, Validität und Reliabilität, die ebenfalls hier aufgenommen werden können.

Die **EXECUTION PROBLEMS** beschreiben die während der Durchführung möglicherweise auftretenden Probleme. Die tatsächlich aufgetretenen Probleme und die getroffenen Maßnahmen

werden in **EXECUTION PROBLEM ACTIONs** beschrieben. Diese werden zusammen mit den im Rahmen des Assessments getroffenen Maßnahmen dokumentiert.

Dementsprechend beschreiben die **RAW ERROR TYPEs** die möglichen Fehlertypen, die bei der Gestaltung des Assessments auf Grund von Missverständnissen, technischen Defiziten oder anderen Einflussfaktoren möglicherweise auftreten, die aber nicht in die Bewertung des Assessments einfließen sollen. Dies können in e-Assessments beispielsweise fehlerhafte Leerzeichen, Verwechslung von Buchstaben und Ziffern aber auch wie in allen Assessments Rundungsfehler oder ähnliche Fehlertypen sein. Die tatsächlich getroffenen Maßnahmen hinsichtlich dieser Fehlertypen werden in **RAW ERROR TYPE ACTIONs** beschrieben.

Der weitere Teil des Modells beschreibt den inhaltlichen Aufbau von **QUESTIONNAIREs**, wobei das zentrale Element die kleinste elementare Aufgabe, das **ASSESSMENT ITEM** ist. Die Struktur soll die Wiederverwendung einzelner **ASSESSMENT ITEMs**, aber auch größerer Einheiten wie **ASSESSMENT ITEM SETs**, **EXERCISEs** bis hin zum kompletten **QUESTIONNAIREs** einschließlich des begleitenden **MATERIALs** erlauben. Damit soll neben der bereits beschriebenen kompletten Wiederverwendung von Assessments als **SETTINGs** auch die Nutzung beliebiger Teilmengen zum Aufbau neuer **SETTINGs** möglich sein, da nur eine solcher „Aufgabenpool“ letztlich den verhältnismäßig großen Aufwand für deren Erstellung und Dokumentation rechtfertigt.

Im Mittelpunkt steht dabei das **ASSESSMENT ITEM**. Bei ihm handelt es sich um eine einzelne atomar beobachtbare und bewertbare Aufgabe oder Teilaufgabe. Es wird davon ausgegangen, dass aus den **ABSTRACT ITEMs** des CORE-Modells konkrete Aufgaben in Form von **ASSESSMENT ITEMs** kausal im Sinne der „Four Building Blocks“ abgeleitet werden können. Dabei können aus einem Item mehrere Aufgaben abgeleitet werden, es kann aber auch eine Aufgabe als geeignet zur Überprüfung mehrerer **ABSTRACT ITEMs** angesehen werden. Insofern wird hier ein Bezug zum CORE-Modell hergestellt, der auf dem Prinzip der (fach-)pädagogischen Herleitung beruht (causality). Im Rahmen der Modellierung ist daher vorgesehen, dass das **ABSTRACT ITEM** die Erstellung eines **ASSESSMENT ITEMs** bedingen kann. Mit **CAUSES ITEM** wird diese Beziehung hergestellt und zusätzlich mit der erwarteten gemessenen Fähigkeit (**SKILL**) verbunden.

Andererseits zeigt sich unter Umständen erst bei der Nutzung des Items in einem Assessment dessen Eignung zur Messung sowohl von **ABSTRACT ITEMs** als auch deren übergeordneten Strukturen in Form von **SKILLs** und **FACTs** (inference). Somit muss bei der Analyse des Items im Rahmen des Assessments eine solche Dokumentation erfolgen. Der Prozess ist in Abschnitt 7.11 im Rahmen des Outcome Modeling beschrieben. Zur Messung der Antworten der Assessment-Teilnehmer für ein **ASSESSMENT ITEM** wird auf eine oder mehrere **MANIFEST VARIABLE** zurückgegriffen. Diese sind Teil des Score-Modells und können dort für die Inference die Beziehungen zu den **SKILLs** und dann weiter über das **CONTEXT ELEMENT** zu Typen des Fact-Modells herstellen. Ebenfalls zum Modellbereich Score gehören die für die Speicherung der möglichen Antworttypen zuständigen **RESPONSE TYPEs**.

Schließlich ist als prozessorientierte Erweiterung eine Zuordnung zu einer **ACTIVITY** vorgesehen, indem das Item als beobachtbarer und bewertbarer Teil eines auszuführenden Prozesses aufgefasst wird. Wie beschrieben ist dies nicht Teil der Arbeit. Das Gleiche gilt auch für die Möglichkeit von

Die Zusammenfassung von **ASSESSMENT ITEMSETs** wird als **EXERCISEs** bezeichnet, also Aufgaben. Diese können bei Bedarf als gröbere Zusammenfassungen thematisch ähnlicher Teile genutzt werden. Sollen beispielsweise Situationsbeschreibungen erfolgen und wird entsprechendes **MATERIAL** für eine Situation erstellt, kann es sich auf eine **EXERCISE** beziehen, die aus mehreren **ASSESSMENT ITEMSETs** besteht. Die **EXERCISEs** bilden wiederum die Grundlage für die gesamte Assessment-Beschreibung, die hier als **QUESTIONNAIRE** bezeichnet wird. Sollten weitere Untergruppierungen notwendig werden, ist eine entsprechende Strukturierung auf Basis der **ASSESSMENT ITEMSETs** vorgesehen.

Die verschiedenen Strukturierungstypen werden zu **ASSESSMENT ELEMENTs** abstrahiert, ähnlich wie dies mit den **FACTs**, **SNIPPETs** und **ABSTRACT ITEMs** sowie dem **CONTEXT ELEMENT** erfolgt ist. Damit kann für jeden dieser Typen begleitendes **MATERIAL** zur Verfügung gestellt werden, etwa Situationsbeschreibungen, Lösungshinweise oder Ähnliches. Schließlich dient das **ASSESSMENT SET** zu weiteren Gruppierungen, die unabhängig von einem konkreten Assessment, etwa für einen Aufgabenpool verwendet werden können. Hier kann die Modellierung auch weiter verfeinert werden, etwa durch eine weitere hierarchische Gliederung, bessere Nutzung von Schlüsselwörtern und Ähnlichem.

Die detaillierte Beschreibung der Typen des Modellbereichs Assessment mit Attributen ist im Anhang enthalten.

6.3.2 Modellbereich Score

Der Modellbereich Score dient zur Beschreibung der deterministischen Bewertung der Assessment-Ergebnisse. Er bildet zugleich die Basis für die stochastische Bewertung.

Jedes **ASSESSMENT ITEM** besitzt einen Outcome Space, der aus einer Menge qualitativ unterschiedlicher Antworten besteht (Wilson 2004b, 63). Der Outcome Space beschreibt so unterschiedlich zu bewertende, aber typisierte Antworten. Jede Antwort wird als ein **RESPONSE TYPE** bezeichnet. Im einfachsten Fall sind die möglichen Antworten „Ja“ und „Nein“. Aber auch komplexere Antworten müssen möglich sein. Dies wird im Rahmen des Feedback Modells in Abschnitt 7.12 am Beispiel der einfachen Subtraktion gezeigt. Dieses Beispiel würde 10 verschiedene **RESPONSE TYPEs** beinhalten. Die im CSP verwendeten **RESPONSE TYPEs** können bereits bei der Entwicklung des Assessments festgelegt werden (predefined) oder sie werden basierend auf den tatsächlichen Antworten ermittelt (ex-post). Gerade dann, wenn es sich um relativ offene Antwortformate handelt, sind unerwartete Antworten nach der tatsächlichen Durchführung des Assessments zu ergänzen (Wilson 2004b, 14). Bei mehrfacher Nutzung derselben Assessments sind hier Synergieeffekte zu erwarten und bisherige „Ex-post“-Antworten können für weitere Durchführungen als „predefined“ vorgegeben werden. So können beispielsweise auch die Ergebnisse polyvalenter Aufgaben mit dezidierten Scores verbunden werden.

Jede tatsächliche Antwort, jede tatsächliche **PERSONAL ITEM RESPONSE** aus dem PERSONAL-Modell muss in ein **PERSONAL ITEM RESPONSE TYPE** als Antwort umgewandelt werden, der wiederum genau einem **RESPONSE TYPE** entspricht. Die Korrektur des eigentlichen **PERSONAL ITEM RESPONSE** zum **PERSONAL ITEM RESPONSE TYPE** kann über die dabei involvierten **RAW ERROR TYPEs** dokumentiert werden. Eine solche Korrektur kann mehrstufig erfolgen, sodass auch mehrere dieser **RAW ERROR TYPEs** genutzt werden können, beispielsweise ein falsches Dezimalzeichen oder Rundungsfehler. Die Zuordnung der **PERSONAL ITEM RESPONSE TYPEs**

zu den **RESPONSE TYPE**s erfolgt, indem alle als gleichwertig angesehenen Antworten demselben **RESPONSE TYPE** zugeordnet werden.

Ein **SCORING** ist die Zuordnung einer Bewertung (**SCORE**) aus einer Bewertungsskala (**MEASUREMENT SCALE**) für einen standardisierten **RESPONSE TYPE** unter dem Blickwinkel eines bestimmten **SKILL** zu einer **MANIFEST VARIABLE**. Somit werden vier Aspekte einbezogen. Das **ASSESSMENT ITEM** wird dabei durch eine **MANIFEST VARIABLE** repräsentiert, die genau eine solche Bewertung unter einem bestimmten Blickwinkel durch die Verbindung zu dem **SCORING** erlauben soll. Für ein **ASSESSMENT ITEM** können mehrere **MANIFEST VARIABLE** definiert werden. Außerdem bildet die **MANIFEST VARIABLE** die Basis für entsprechende Gruppierungen in der Bewertung.

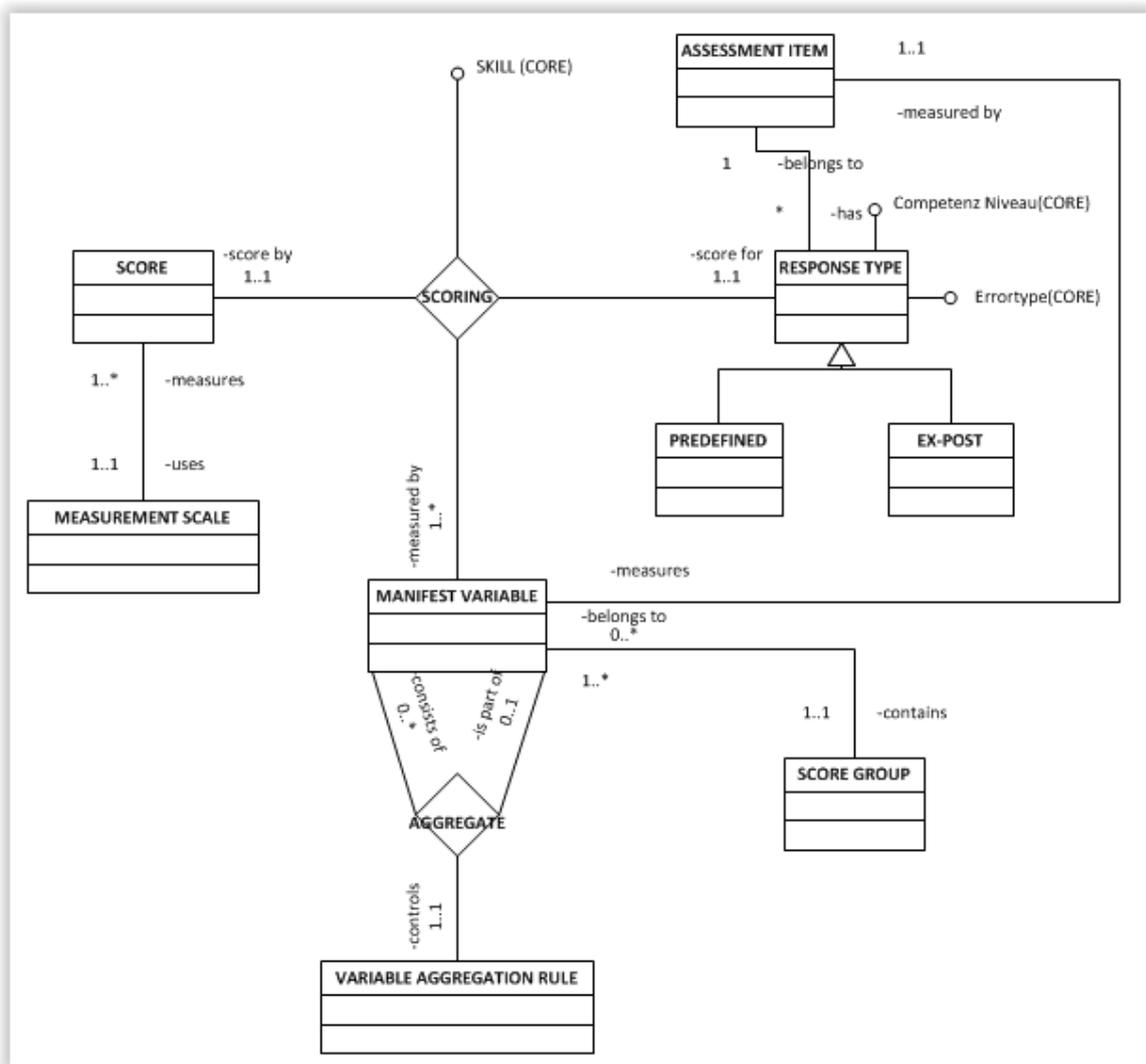


Abbildung 40 Modellbereich Score

Die Einbeziehung des **SKILL** in die Bewertung ist erforderlich, da die Bewertung derselben Antwort auf ein Item aus unterschiedlichen Perspektiven erfolgen kann. Beispielsweise kann ein Item unter Berücksichtigung seines Beitrags zu einem **SKILL** unterschiedlich gesehen werden, weil die Antworten in unterschiedlicher Weise auf das Vorhandensein des **SKILL**s schließen lassen.

Auch die für eine Bewertung verfügbaren Werte, also die Skala kann in unterschiedlichen Settings unterschiedlich sein. Derartige Skalen können von einer zentralen Instanz wie dem Kultusministerium aber auch von Organen der Schule, beispielsweise bestimmten Konferenzen oder letztlich auch von der einzelnen Lehrkraft festgelegt worden sein. Daher können verschiedene **SCO_MEASUREMENT_SCALES** definiert werden, denen dann die entsprechenden **SCORES** zugeordnet werden.

Eine **MANIFEST VARIABLE** beschreibt die Messvariable für ein **ASSESSMENT ITEM**. Für ein **ASSESSMENT ITEM** können mehrere **MANIFEST VARIABLES** definiert werden. Grundsätzlich bezieht sich eine **MANIFEST VARIABLE** immer auf ein **ASSESSMENT ITEM**. Die Variablen bilden die Basis sowohl für Aggregationen zu anderen **MANIFEST VARIABLES** als auch beispielsweise für die statistische Analyse mit **LATENT VARIABLES**.

Sollen die Variablen aggregiert werden, kann dies mittels **AGGREGATES** erfolgen. Die **VARIABLE AGGREGATION RULES** dienen dazu festzulegen, wie die Wertaggregation für die **MANIFEST VARIABLES** erfolgen soll, die rekursiv aus anderen **MANIFEST VARIABLES** aufgebaut werden können. Im Gegensatz zu den Beziehungen bei der Konstruktion des Assessments sind diese nicht kausal aus der Konstruktion des Modells hergeleitet, sondern repräsentieren die Schlusskette (infer) nach Durchführung des Assessments. Diese kann, muss aber nicht zwingend, mit der Konstruktion des **CORE-Modells** und des **ASSESSMENT-Modells** übereinstimmen. Insbesondere erfolgt die Gruppierung so, dass eine **MANIFEST VARIABLE** maximal einem **SKILL** zugeordnet werden kann. Diese gruppierten Variablen können durch die Verbindung zum übergeordneten **SKILL** auch **CONTEXT ITEMS** und somit **ABSTRACT ITEMS**, **SNIPPETS** oder **FACTS** für die manifeste Analyse zugeordnet werden. Mittels der **MANIFEST VARIABLES** können außerdem im **PERSONAL-Modell** die **PERSONAL AGGREGATED SCORES** berechnet und gespeichert werden. So können die berechneten Scores für die leichtere Analyse personenbezogen materialisiert werden. Die zusammengehörigen Variablen werden für eine Analyse in **SCORE GROUPS** zusammengefasst.

Die detaillierte Beschreibung der Typen des Modellbereichs Score mit Attributen ist im Anhang enthalten.

6.3.3 Modellbereich Statistik

Das Statistikmodell dient der Ableitung von kompetenzorientierten Zahlen mittels eines statistischen Modells, das entweder auf der Item Response Theory (IRT) oder der Classical Test Theory (CTT) beruht. Im **STATISTICAL MODEL TYPE** wird die Modellbeschreibung mit den entsprechenden Modellierungsentscheidungen und Parametern hinterlegt. Eine komplette statistische Analyse wird als **STATISTICAL MODEL** hinterlegt. Ein solches Modell ist immer einem bestimmten Typ zugeordnet, der die Grundlagen beschreibt. Es bezieht sich auf eine bestimmte **SETTING EXECUTION**, also die konkrete Durchführung eines Assessments. Für eine **SETTING EXECUTION** können mehrere statistische Modelle, auch basierend auf demselben Typ angelegt werden, da die Durchführung die Wahl unterschiedlicher Parameter oder Entscheidungsschritte erlauben kann. Ein Modell beschreibt also die tatsächliche Analyse des Ergebnisses der Durchführung eines Settings.

Das Modell dient der Messung von **LATENT VARIABLES**. Die latenten Variablen beruhen auf Mengen von manifesten Variablen und sind damit letztlich **ASSESSMENT ITEMS** zugeordnet. Die manifesten Variablen beschreiben die direkt beobachtbaren Ergebnisse im Assessment, die

latenten Variablen beruhen auf diesen und beschreiben dementsprechend die nicht beobachtbaren zugeordneten Kompetenzen. Eine latente Variable beschreibt so die Kompetenzschätzung bezogen auf einzelne oder Gruppen von Items. Letztlich kann ein komplettes Assessment als Messung einer einzigen latenten Variablen angesehen werden. Es können aber auch Teile, bestimmte Items auf Grund inhaltlicher Kriterien gruppiert und gemeinsam einer latenten Variablen zugeordnet werden. Das Modell lässt hier eine beliebige Zuordnung bis hin zu einer latenten Variablen pro Item zu.

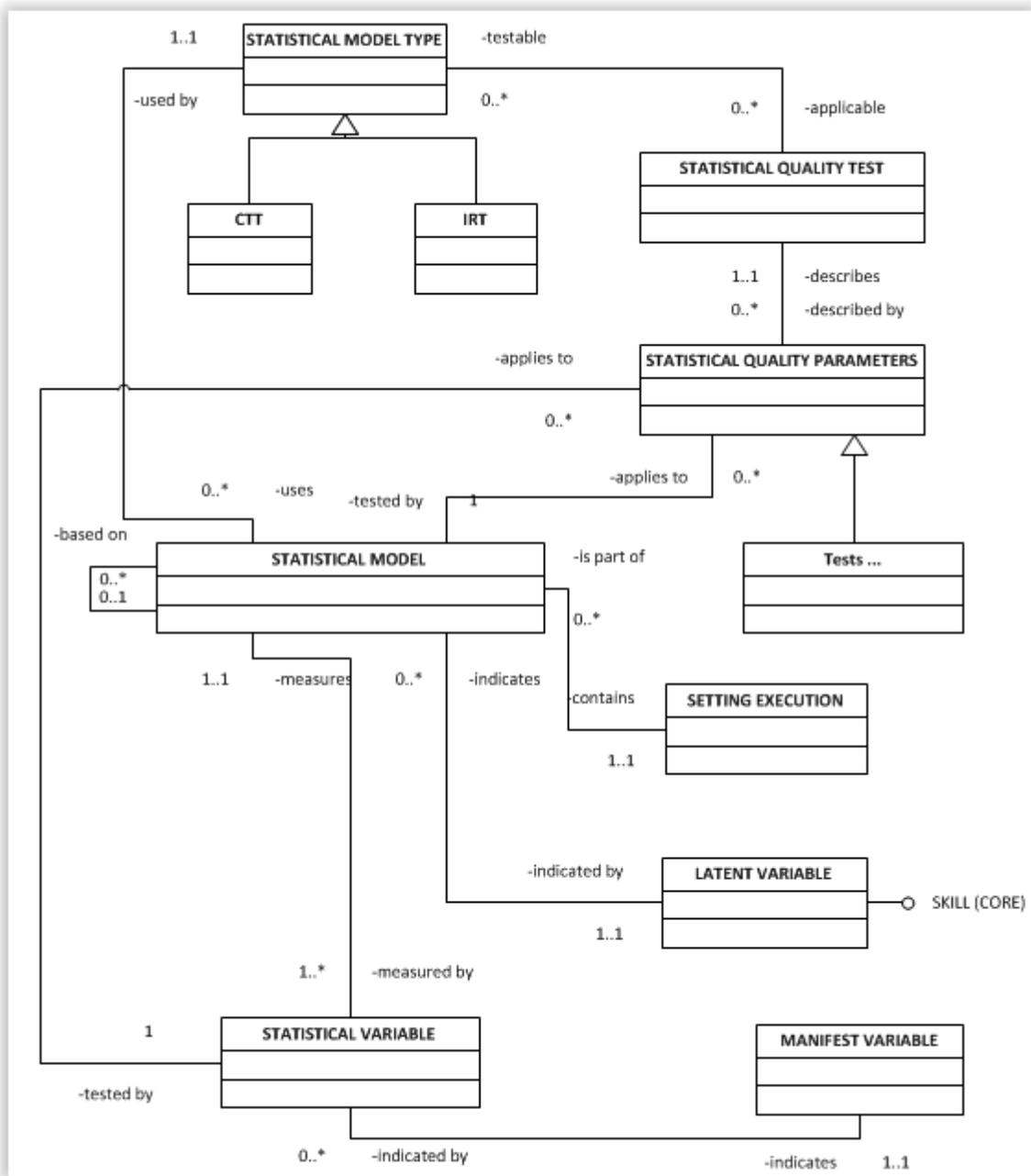


Abbildung 41 Modellbereich Statistik

Dabei werden für alle potenziell relevanten **MANIFEST VARIABLES** jeweils **STATISTICAL VARIABLES** angelegt. Eine Gruppe wird durch Gruppierung der passenden **STATISTICAL VARIABLES** zu einem **STATISTICAL MODEL** gebildet. Das **STATISTICAL MODEL** dient der Schätzung des Traits für eine **LATENT VARIABLE**, wobei Werte der **STATISTICAL VARIABLES**

die Grundlage für diese Schätzung bilden. So kann sichergestellt werden, dass eine latente Variable immer auf einer Gruppe von manifesten Variablen beruht ohne die Allgemeingültigkeit der Gruppenbildung zu beeinträchtigen. Umgekehrt kann eine manifeste Variable von mehreren latenten Variablen verwendet werden, beispielsweise für die Analyse in verschiedenen Modellen. Insbesondere kann dieselbe **LATENT VARIABLE** iterativ in verschiedenen Modellen geschätzt werden. Die Zwischenschaltung der statistischen Variablen dient der Dokumentation, welche dieser Variablen tatsächlich und in welcher Form in die Schätzung eingegangen sind. Im Falle des RASCH-Modells können hier beispielsweise der Schwierigkeitsgrad aber auch die Standardabweichung für dessen Schätzung dokumentiert werden. Daher sind entsprechende Möglichkeiten in der **STATISTICAL VARIABLE** vorgesehen.

Das Modell dient der Einschätzung der Kompetenz der **PERSONS**, die an dem Assessment teilgenommen haben. Diese Kompetenzen werden als Traits bezeichnet. Die Ergebnisse werden letztlich im **PERSONAL-Model** im Typ **PERSONAL_TRAIT_SCORE** mit Beziehung zu der relevanten **LATENT VARIABLE** abgelegt.

Neben der Schätzung des Traits für die Personen muss das statistische Modell selbst auf seine Gültigkeit überprüft werden. Die jeweils angewendeten Verfahren zur Überprüfung der Qualität der Modellschätzung werden in **STATISTICAL QUALITY TEST** dokumentiert, insbesondere handelt es sich dabei um typische statistische Testverfahren. Die Ergebnisse dieser Tests können in den entsprechenden **STATISTICAL QUALITY PARAMETERS** hinterlegt werden. Hier ist die Ergänzung um dem jeweiligen Test entsprechende Typen vorgesehen, da die Parameter zum Teil sehr stark variieren.

Die detaillierte Beschreibung der Typen des Modellbereichs Statistik mit Attributen ist im Anhang enthalten.

6.4 Das **PERSONAL-Modell**

Das **PERSONAL-Modell** dient der Speicherung der Daten, die sich auf eine **PERSON**, einen Assessment-Teilnehmer, beziehen, insbesondere der tatsächlichen Assessmentergebnisse dieser Person. Der Typ **PERSON** ist daher zentral in diesem Modell. Er wird überall in dem Modell referenziert, auch wenn nur zwei Beziehungen dargestellt sind. Personen können zu verschiedenen Gruppen **GROUP** zusammengefasst werden, um gruppenspezifische Analysen zu ermöglichen. Die persönlichen Daten werden auch aus datenschutzrechtlichen Gründen in **PERSONAL DATA** ausgelagert. Dieser Typ kann im Modell oder extern realisiert sein. Neben der **PERSON** ist die **SETTING EXECUTION** zentral, die zwar zum **SITUATION-Modell** gehört, aber ebenfalls überall referenziert wird, um die Zuordnung der Ergebnisse zur Ausführung des Assessments sicherzustellen. Viele Informationen werden voneinander abgeleitet und sind oft materialisierte Views, die auch dynamisch bestimmt werden könnten. Eine solche Materialisation erhöht aber die Übersichtlichkeit und erleichtert viele weitere Auswertungen.

Die Reihenfolge der Ableitung dieser Views ist hier wichtiger als die Beziehungen, die sich zumeist auf die **PERSON** und einige Typen des **SITUATION-Modells** beziehen. Die Reihenfolge der Erstellung ist daher durch gestrichelte Pfeile dargestellt. Zunächst wird für jede **PERSON** die **PERSONAL ITEM RESPONSE** gespeichert. Sie soll so exakt wie möglich der beabsichtigten Originalantwort der Person entsprechen. Einerseits bedeutet dies die tatsächliche Antwort abzulegen. Andererseits sind eventuell durch technische Probleme verursachte Fehler in der

Durchführung so weit zu korrigieren, dass die tatsächlich beabsichtigte Antwort möglichst gut rekonstruiert werden kann (**RAW ERROR TYPES**). Sonst werden die tatsächlichen Antworten nicht verändert. Es werden also die tatsächliche, unveränderte Antwort und die gegebenenfalls rekonstruierte, beabsichtigte Antwort dokumentiert.

Da die Antwort sich auf die Messung eines bestimmten **ASSESSMENT ITEMS** bezieht, wird sie mit der zugehörigen **MANIFEST VARIABLE** verknüpft. Diese Antwort muss dann einem **RESPONSE TYPE** des entsprechenden **ASSESSMENT ITEMS** zugeordnet werden, um sie bewerten zu können. Dabei werden wie im Rahmen des Outcome Modeling die vorhandenen Antworttypen bei Bedarf ergänzt.

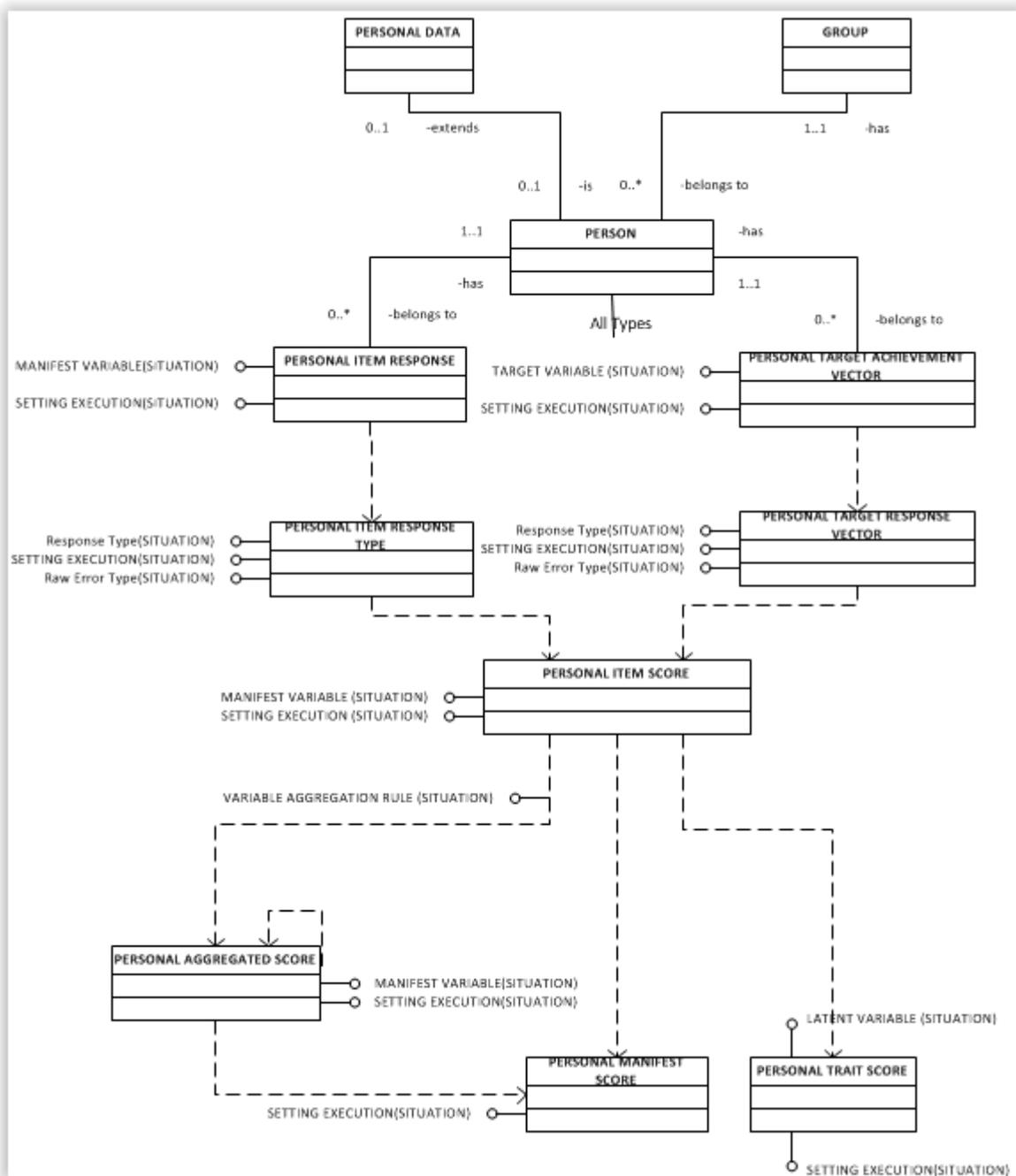


Abbildung 42 PERSONAL-Modell

In jedem Fall muss die Antwort einem **RESPONSE TYPE** zugeordnet und die zugeordnete Antwort im **PERSONAL ITEM RESPONSE TYPE** abgelegt werden.

Damit ist der Schritt der Beobachtung abgeschlossen und es wird zur Bewertung übergegangen. Dafür wird auf die im SCORE-Modell hinterlegte Bewertungsdokumentation zurückgegriffen. Die Bewertung insbesondere mittels des **SCORING** führt dazu, dass der relevante Score hier im PERSONAL-Modell im **PERSONAL ITEM SCORE** eingetragen werden kann.

Dabei handelt es sich erstmalig um eine Bewertung der Antwort auf ein bestimmtes Item aus einer bestimmten Perspektive (Skill) für eine bestimmte Person. Dies kann daher auch als „Scored response“ aufgefasst werden. Diese „Scored responses“ lassen sich mittels der Aggregation der **MANIFEST VARIABLEs** zu **PERSONAL AGGREGATED SCOREs** verdichten, den entsprechenden aggregierten **MANIFEST VARIABLEs** zuordnen und über diese in verschiedenen Verdichtungsstufen beispielsweise pro **SKILL** oder **FACT** bestimmen und bei Bedarf in **PERSONAL AGGREGATED SCORE** materialisieren.

Auf Basis der **PERSONAL ITEM SCOREs** oder der **PERSONAL AGGREGATED SCOREs** kann durch Aggregation ein **PERSONAL MANIFEST SCORE** für das Assessment berechnet werden.

Außerdem bildet der Typ **PERSONAL ITEM SCORE** die Grundlage für Analysen im Modellbereich Statistik für die entsprechenden Kompetenzschätzungen. Dabei können wie in Abschnitt 7.15 beschrieben, auch zusätzliche Bewertungen entsprechend den Anforderungen des verwendeten statistischen Modells hinterlegt werden. Die dort bestimmten Trait-Schätzungen können dann wiederum im Typ **PERSONAL TRAIT SCORE** gespeichert werden.

Der rechte Ast des Modells mit den Typen **PERSONAL TARGET ACHIEVEMENT VECTOR** und **PERSONAL TARGET RESPONSE VECTOR** beinhaltet mögliche Erweiterungen für die Berücksichtigung von Prozessen, deren Erfolg in einem Zielerreichungsvektor dokumentiert wird.

Die detaillierte Beschreibung der Typen des Modellbereichs Personal mit Attributen ist im Anhang enthalten.

7 CSP - Vorgehensmodell

Ein wie im vorherigen Kapitel entwickeltes Informationsmodell bedarf einer Anweisung wie es zu nutzen ist. Damit ist der Prozess der Population des Modells, die Durchführung des Assessments, dessen Nachbearbeitung und Auswertung gemeint, also ein Vorgehensmodell.

Wie in Abschnitt 4.3 dargestellt beschreiben sowohl Wilson im Rahmen der „Four Building Blocks“ als auch Mitlevy et al. im Rahmen des Evidence-Centered Design eine solche Vorgehensweise zur Entwicklung, Durchführung und Analyse eines Assessments. Diese muss sich hier zusätzlich auf die Erstellung der Beschreibung des Domäneninhaltes beziehen. Außerdem erlaubt das detaillierte Informationsmodell eine genauere Beschreibung der zu erarbeitenden Inhalte. Grundsätzlich handelt es sich also um einen ähnlichen, aber detaillierteren Ablauf. Allerdings erfordert das CSP auch eine etwas andere Vorgehensweise, da das Messkonstrukt unabhängig vom konkreten Assessment vordefiniert wird. Außerdem werden in der Auswertung vielfältigere Wege vorgeschlagen und in allen Schritten ist eine ausführlichere Dokumentation der Zusammenhänge unter Bezugnahme auf das CSP-Informationsmodell vorgesehen.

Zentrale Aspekte sind also die durch die getrennte, aber aufeinander aufbauende Entwicklung der Domänenbeschreibung, des Assessments und seiner Durchführung angestrebte Wiederverwendbarkeit der jeweils anderen Teile sowie die Anknüpfung an eine strukturierte Dokumentation im Informationsmodell.

7.1 Überblick

In Abbildung 43 sind zunächst die wichtigsten Schritte und die grundsätzliche Vorgehensweise für diese Schritte dargestellt. Zu jedem Schritt wird im Folgenden das erwartete Ergebnis angegeben. Dieses ist letztlich wichtiger als die Vorgehensweise innerhalb des Schritts.

Wie bei Wilson die „Construct Map“ oder bei der ECD die Domänenanalyse steht eine Beschäftigung mit dem inhaltlichen Ziel am Anfang der Entwicklung. Zuvor ist allerdings ein möglicher Initialisierungsschritt sinnvoll, der domänenübergreifende Aspekte beinhaltet.

(0) Initialization - Es werden domänenübergreifende Aspekte dokumentiert, soweit dies sinnvoll ist, beispielsweise allgemeine Kompetenzhierarchien, Bewertungsskalen oder Vorgehensweisen bei Problemen in der Assessmentdurchführung.

(1) External Analysis - Der Ausgangspunkt ist eine Analyse der extern vorgegebenen Ziele und Kompetenzanforderungen. Dazu sind die relevanten Dokumente zu analysieren und das Ergebnis ist im CSP abzulegen. Das Ergebnis ist ein strukturierter Anforderungskatalog.

(2) Competence Modeling – Im nächsten Schritt wird dieser Katalog in Form von Kompetenzen mit verschiedenen Niveaus intern strukturiert und in Abgrenzung gegeneinander weiter ausformuliert. Das angestrebte Verfeinerungsniveau hängt von den angestrebten Messergebnissen ab. Das Ergebnis ist eine Liste zu messender Kompetenzen mit Niveaus und individuellen Algorithmen zur Bestimmung der Niveaus.

(3) Content Modeling – Sind die angestrebten Kompetenzen ausreichend verfeinert, müssen sie in den eigentlichen Domäneninhalt in Form von deklarativem, prozeduralem und anderem „Wissen“ detailliert werden. Dabei wird wie bei Wilson das Prinzip der vermuteten Kausalität verwendet. Es wird also mittels Expertenwissen die plausibelste Beschreibung und Zuordnung zu Kompetenzen

getroffen. Hier sind auch Auswahlentscheidungen hinsichtlich des zu berücksichtigenden Umfangs erforderlich. Das Evidence-Centered Design spricht hier von zentralen Ideen und typischen Prozessen. Dies ist das Ergebnis des Content Modeling.

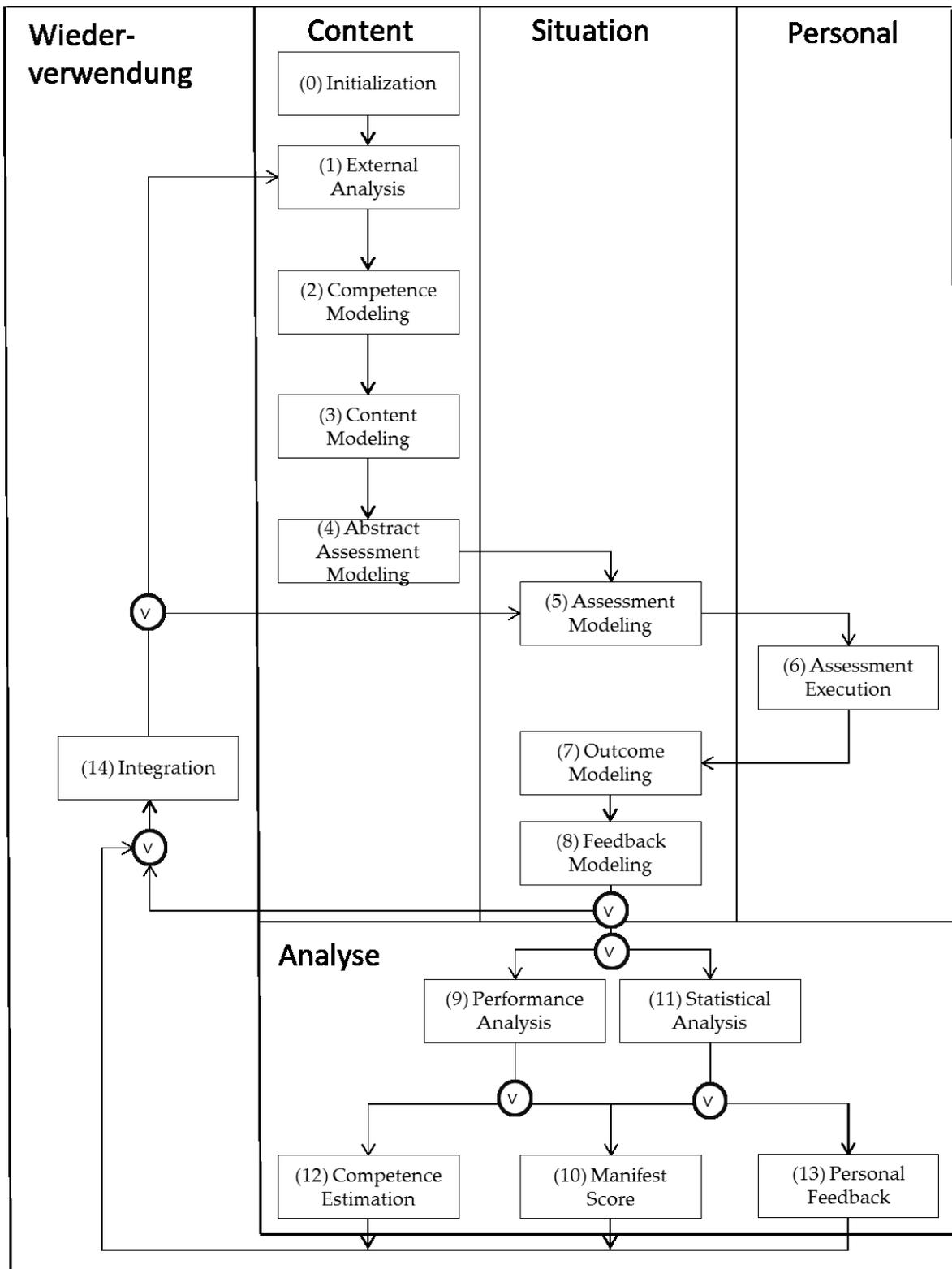


Abbildung 43 Grundstruktur der CSP-Assessment-Modellierung

(4) Abstract Assessment Modeling – Im Gegensatz zu den geschilderten Ansätzen erfolgt nicht unmittelbar eine Umsetzung in ein konkretes Assessment, sondern es werden in einer weiteren

Verfeinerung die zentralen Aspekte des im Content definierten Wissens modelliert und Vorschläge für abstrakte Items entwickelt mit denen diese Aspekte gemessen werden können. Das soll die Formulierung ähnlicher Items zur Erfassung derselben Aspekte wie auch die Gestaltung anderer vergleichbarer Assessments erleichtern. Das Ergebnis sind Muster für Items.

(5) Assessment Modeling – Die konkrete Modellierung des Assessments beinhaltet die Gestaltung von fünf verschiedenen Aspekten. Zum ersten ist die Festlegung der Rahmenbedingungen des Assessments notwendig. Dies entspricht zusammen mit dem zweiten Aspekt, der Definition der Variablen und des statistischen Modells der Konstruktion des Assessment-Modells im Sinn des ECD beziehungsweise der Definition des vierten „Building Blocks“, dem „Measurement Model“ bei Wilson. Der dritte Schritt besteht in der Modellierung der Items wie sie auch das ECD kennt und wies es Wilson in seinem zweiten Building Block, „Item Response“ beschreibt. Zusätzlich scheint eine Gruppierung der Items sinnvoll. Außerdem ist die Modellierung erwarteter Lösungen, also der Response Types hier zugeordnet.

(6) Assessment Execution – Die eigentliche Durchführung ist der erste Teil dessen was das ECD zusammenfassend als Einsatz des Assessments beschreibt. Hier geht es um die Durchführung und die Gewinnung der Rohdaten als Ergebnis.

(7) Outcome-Modeling – Die Ergebnisse des Assessments in Form von Rohdaten sind das, was Wilson als „Outcome Space“ bezeichnet. Es ist zumindest bei nicht vollkommen geschlossenen Fragestellungen nicht davon auszugehen, dass alle möglichen Antworten und Vorgehensweisen komplett antizipiert worden sind. Es sind immer unerwartete Lösungen möglich oder in den Lösungen werden unerwartete Gedankengänge sichtbar. Daneben sind auch unerwartete Probleme bei der Durchführung, Unklarheiten oder sonstige Störungen zu beachten. Daher muss das Ergebnis nachbearbeitet werden und diese Besonderheiten müssen in die Auswertung einbezogen werden. Das Ergebnis ist ein unmittelbar analysierbares Assessment-Ergebnis in Form eindeutiger Beobachtungen für die einzelnen Items.

(8) Feedback Modeling – Liegen alle möglichen Response Types vor können sie verschiedenen Fehlertypen zugeordnet werden. Diese Fehlertypen können genauso wie zu erwartende Defizite bei Skills oder Fakten mit möglichen Ursachen und Maßnahmen zu deren Behebung verbunden werden. Diese Zuordnung ist das Ergebnis des Feedback-Modeling.

(9) Performance Analysis – Die Performance Analysis umfasst die Bewertung der Ergebnisse mit dem Ziel eines quantitativen Assessment-Ergebnisses. Das Ergebnis ist eine Bewertung der einzelnen Beobachtungen für die Assessment-Teilnehmer.

(10) Manifest Score – Der Manifest Score ist das direkte, summative Ergebnis des Assessments für jeden einzelnen Assessment-Teilnehmer, sozusagen das „Klausurergebnis“.

(11) Statistical Analysis – Die Statistical Analysis umfasst die Anwendung des gewählten statistischen Modells zur Ermittlung der Kompetenzschätzungen für die einzelnen Personen. Das Ergebnis ist eine Kompetenzschätzung für jeden Assessment-Teilnehmer.

(12) Competence Estimation – Die Competence Estimation verwendet die Schätzungen der vorhergehenden Stufe zur Ermittlung von Kompetenzaussagen übergeordneter Aspekte, also im Bereich des Contents, der Skills oder der dokumentierten Kompetenzen mit ihren Hierarchieebenen. Hier geht es um die „Indikatoren“ (Wilson spricht von „infer“), die auf den über die

Kausalität abgeleiteten tatsächlichen Zusammenhang mit den modellierten Kompetenzanforderungen hinweisen. Das Ergebnis sind Zuordnungen der Kompetenzniveaus hinsichtlich der modellierten Kompetenzen für die Assessment-Teilnehmer.

(13) Personal Feedback – Sowohl die Performance Analysis als auch die Statistical Analysis können genutzt werden, um auf Basis der Einzelergebnisse und dem dokumentierten Zusammenhang zu den modellierten Strukturen individuelle Stärken und Schwächen zu erkennen, zu dokumentieren, der Person zu kommunizieren und im Sinn eines formativen Assessments zielgerichtete Maßnahmen vorzuschlagen. Entsprechendes gilt für die gesamte Lerngruppe.

(14) Integration – Abschließend kann an verschiedenen Stellen eine Analyse vorgenommen werden, inwieweit bestimmte Ergebnisse verallgemeinert für eine Wiederverwendung im Rahmen des Modells dokumentiert werden können.

7.2 Bezüge zu anderen Assessment-Modellen

7.2.1 Bezug zum Assessment Triangle

Das „Assessment Triangle“ lässt sich mit der Gliederung in Abschnitt 4.1.1 wie auch mit der CSP-Struktur in Übereinstimmung bringen, wobei das Triangle auch das Vorgehensmodell visualisiert. Der Bereich „cognition“ entspricht dem Kernmodell und dessen Entwicklung in den Schritten (1) bis (4). Die Kante von „cognition“ zu „observation“ entspricht der Entwicklung des SITUATION-Modells und - insofern die Ergebnisse tatsächlich gespeichert werden - letztlich dem PERSONAL-Modell. Dies entspricht auch den Schritten (5) bis (8). Die „interpretation“ beginnt mit der Performance Analysis (9) und endet mit Schritt (13). Die Kante zurück zu „cognition“ kann als Rückbezug zum Domäneninhalt aufgefasst werden und entspricht der Idee, die Ergebnisse des PERSONAL-Modells bis ins Kernmodell zurückzuverfolgen und geeignet zu aggregieren und zu analysieren, also den Schritten (10) und (12). Insofern stellt das Vorgehen im CSP eine Verfeinerung eines „Umlaufs“ um das „Assessment Triangle“ dar.

7.2.2 Bezug zum ECD-Modell

Die Ergebnisse der Domänenanalyse, wie sie das ECD vorsieht, können im Core-Modell im Modellbereich Competence specifications, Competence dimension und Content abgelegt werden. Dabei handelt es sich aus Sicht des CSP-Modells nicht nur um die für ein spezielles Assessment relevanten Kompetenzen, sondern um alle für die Domäne als relevant eingeschätzten Kompetenzen. Im Rahmen des Modells wurde dieser Bereich auch als der externe Kompetenzbereich definiert, was gut zu der Vorstellung eines Außenbereichs passt.

Die Domänenmodellierung des ECD knüpft an die Modellierung des Content an. Dabei werden im Sinne einer Operationalisierung auch die **SNIPPETs** und **ABSTRACT ITEMS** des Core-Modells definiert. Der gesamte Bereich ist inhaltlich mit dem „Student Model“ vergleichbar, soll aber nicht nur auf ein Assessment bezogen, sondern übergreifend sein. Die Entwicklung des „Statistic Model“ ist ebenso wie die des „Task Model“ Teil des SITUATION-Modells, da beide auf ein konkretes Assessment bezogen sind. Der Bereich des „Task Model“ wird im CSP insbesondere von dem Modellbereich Assessment abgedeckt, der die grundlegende Struktur des Assessments und seiner Rahmenbedingungen beinhaltet.

Dementsprechend werden die Tätigkeiten der „Konzeption des Assessment-Modells“ und der „Entwicklung des Assessments“ im ECD im Wesentlichen im Assessment Modeling (Schritt 5) abgebildet und im Assessment-Modell dokumentiert. Zusätzlich werden im Scoring-Modell Bewertungsmöglichkeiten und im Statistikmodell stochastische Modelle hinterlegt.

Das „Evidence Model“ des ECD ist letztlich in der Struktur des gesamten CSP an vielen Stellen implementiert und verfeinert. Man könnte es auch als Gesamt-„Evidence Model“ bezeichnen, da gerade die Dokumentation und Auswertung der entsprechenden Regeln in allen Bereichen möglich sein soll. Insbesondere findet man den Teilbereich des „Statistic Models“ im Modellbereich Statistik im CSP wieder, der der Idee einer Basis für eine psychometrische Analyse entspricht. Viele der „Evidence Rules“ des ECD sind in den Modellbereichen „Scoring“, „Skill“ und „Feedback“ enthalten, die die Rückkopplung bei der Analyse ermöglichen.

Die beim Einsatz des Assessments gewonnenen Daten werden im CSP-Modell durch die Instanziierung des PERSONAL-Modells dokumentiert. Das ECD fasst die gesamte Analyse pauschal im „Einsatz des Assessments“ zusammen, während im CSP die Schritte (9) bis (13) vorgesehen sind.

7.2.3 Bezug zu den Four Building Blocks

Die „Four Building Blocks“ sind vom Grundsatz so aufgebaut, dass jeweils das benötigte Ergebnis definiert wird, die Mittel aber weitgehend frei wählbar sind. Dies entspricht dem Gedanken der einzelnen Schritte im CSP. Jeder Vorgehensschritt hat ein bestimmtes Ergebnis zum Ziel. Durch das im CSP definierte Informationsmodell kann das Ergebnis direkt auf Basis dieses Modells definiert werden. Die Werkzeuge sind jedoch nicht festgelegt.

Wilson liefert mit seinen „Four Building Blocks“ auch tiefergehende Hinweise zur Konstruktion des Assessments selbst. Das Messkonstrukt wird im Block „Construct“ definiert, das Messinstrument ist bei Wilson die „Construct Map“. Diese bereitet bereits die spätere Kompetenzbewertung bei der Erstellung des Domäneninhaltes vor. Sie ist allerdings stark an einer psychometrischen Analyse orientiert. Das CSP definiert in seinem Core-Modell ebenfalls das Messkonstrukt in Form des Domäneninhaltes. Dafür ist aber nicht die komplette Nutzung des Inhaltes in einem Assessment notwendig, vielmehr können die Inhalte für ein Assessment ausgewählt werden. Die Umsetzung einer Bewertung in eine Kompetenz ist ebenfalls vorgesehen, insbesondere im Bereich des „Skills“ könnte sie auch als eine „Construct map“ dokumentiert werden. Insgesamt umfasst die Konstruktion des Blocks „Construct“ die Schritte (1) bis (4) des CSP.

Das Prinzip der „Causality“ gilt in beiden Modellen, insbesondere bei der Ableitung des Blocks „Item Responses“, also dem Design der Items für ein konkretes Assessment. Beim CSP erfolgt auch die Auswahl des Bereichs des Domäneninhaltes, der zum Messkonstrukt werden soll. Die „Four Building Blocks“ definieren hier die Items einschließlich möglicher Antwortalternativen. Diese werden im CSP im Assessment-Modell und im Score-Modell hinterlegt. Es können auch Annahmen über die latenten Variablen im Statistik-Modell dokumentiert werden (Schritt 5).

Der dritte Block „Outcome Space“ enthält eine strukturierte und bewertete Darstellung der tatsächlichen Beobachtungen. Dazu können im CSP die möglichen Antworten und deren Bewertung ergänzt werden. Die Strukturierung der Antworten ist insbesondere durch die **RESPONSE TYPEs** im Score-Modell und die entsprechenden individuellen unstrukturierten und strukturierten Beobachtungen im PERSONAL-Modell möglich. Hier kann also die Dokumentation

auch unerwarteter Antworten ergänzt werden. Gleichzeitig wird sie so Bestandteil der Dokumentation des Assessments und kann bei späteren Nutzungen als Teil der bereits vorgegebenen Antworten wiederverwendet werden (Schritt 7).

Aus den so gefundenen Antworten wird bei den „Four Building Blocks“ per „Inference“ die Bewertung und Analyse der Ergebnisse im vierten Block, dem „Measurement Model“ vorgenommen. Diese Inference wird im CSP abgebildet, indem aus den per „Causality“ abgeleiteten Items Bezüge zwischen den Antworten und den Skills hergestellt und verändert werden können. Ebenso können im Rahmen des Score-Modells neue Bewertungen vorgenommen werden und die Ergebnisse des Statistik-Modells einbezogen werden. Dies entspricht den weiteren Schritten bis Schritt 13. Insofern liegen die beiden Modelle hier eng beieinander.

7.3 Bezug zum Informationsmodell

In der folgenden Tabelle 5 ist überblickartig der Bezug der Schritte des Vorgehensmodells zu den Typen des Informationsmodells dargestellt. Dabei geht es darum, in welchen Schritten des Vorgehensmodells welche Typen des Informationsmodells dokumentiert werden. In einigen Fällen sind beispielhaft Hinweise und Hilfsmittel zu deren Befüllung dokumentiert. Der Name ist der Name des Typs im CSP-Informationsmodell. Hier wird bereits auf das physisch umgesetzte Modell Bezug genommen, um die praktische Nutzbarkeit zu erhöhen. Mit dem „Schritt“ ist jeweils der Arbeitsschritt wie in Abbildung 43 dargestellt gemeint, hier erfolgt also der direkte Bezug zum Vorgehensmodell.

Name	Schritt	Hinweise	Hilfsmittel
DIMENSION SPECIFICATIONS (CSP)			
EXTERNAL_GOAL	(1)		Manuell
CURRICULUM_STANDARD	(1)		Manuell
DEFINE_GOAL	(1)		Manuell
COMPETENCE DIMENSIONS (CSD)			
ABSTRACT_DIMENSION	(0)		Manuell
COMPETENCE_DIMENSION	(1,2)		Manuell
COMPETENCE_LEVEL	(2)		Manuell
COMPETENCE	(2)		Manuell
COMPETENCE_NIVEAU_SCALE	(2)		Manuell
COMPETENCE_NIVEAU	(2)		Manuell
SPECIFIC_COMPETENCE_NIVEAU	(2)		Manuell
FACTS (FCT)			
FACT	(3)		Manuell
FACT_TYPE	(3)		Manuell
VALENCE	(3)		Manuell
VALENCE_TYPE	(3)		Manuell
CONSTRAINT	(3)		Manuell
CONSTRAINT_TYPE	(3)		Manuell
ENTITY	(3)		Manuell
LABEL_TYPE	(3)		Manuell
ITEMS (ITM)			
SNIPPET	(3,4)		Manuell
ABSTRACT ITEM	(4)		Manuell
CONTEXT_ELEMENT	(4)		Datenbank

SKILL (SKI)			
SKILL	(4)		Manuell
OPERATOR	(0,1,4)		Manuell
COMPETENCE_NIVEAU_METRICS	(4)		Manuell
SKILL_INDICATING	(4)		Manuell
CNM_AGGREGATE	(4)		Manuell
LEVELINDICATOR	(4)		Manuell
FEEDBACK (FBCK)			
EDUCATION_COURSE	(0,8)		Manuell
EDUCATION_MODULE	(0,8)		Manuell
ERROR_TYPE	(5,7,8)		Manuell
EDUCATION_MATERIAL	(0,8)		Manuell
FEEDBACK_MATERIAL	(8)		Manuell
FEEDBACK	(5,7,8)		Manuell
FEEDBACK_SCN	(5,7,8)		Manuell
ACTION	(8)		Manuell
ASSESSMENT MODEL (ASS)			
SETTING	(0,5)		Manuell
SETTING_EXECUTION	(5)		Manuell
EXECUTION_PARAMETERS	(5)		Manuell
EXECUTION_PROBLEM	(0,5,14)		Manuell
EXECUTION_PROBLEM_ACTION	(6)		Manuell
RAW_ERROR_TYPE	(0,5,14)		Manuell
RAW_ERROR_TYPE_ACTION	(8)		Manuell
QUESTIONNAIRE	(0,5)		Manuell, zur Generierung des Assessments
CAUSES_ITEM	(5)		Manuell/Datenbank
ASSESSMENT_ITEM	(5)		Manuell, zur Generierung des Assessments
ASSESSMENT_ITEM_SET	(5)		Manuell, zur Generierung des Assessments
EXERCISE	(5)		Manuell, zur Generierung des Assessments
MATERIAL	(5,14)		Manuell
ASSESSMENT_ELEMENT	(5,14)		Manuell
ASSESSMENT_POOL	(14)		Manuell
SCORE MODEL (SCO)			
AGGREGATE	(10,12)		Manuell
MEASUREMENT_SCALE	(0,5,9)		Manuell
SCORE	(0,5,7,9)		Manuell
RESPONSE_TYPE	(5,7)		Manuell und mit R aus den Antworten
MANIFEST_VARIABLE	(5,9,10,12)		Aus Items
VARIABLE_AGGREGATION_RULE	(10,12)		Manuell
SCORING	(5,7,9)		Datenbank
STATISTICS MODEL (STA)			
STATISTICAL_MODEL_TYPE	(0,5)		Manuell
STATISTICAL_QUALITY_TEST	(0,5)		Manuell
APPLICABLE_TEST	(0,5)		Manuell

STATISTICAL_MODEL	(5,11,12)		Manuell
STATISTICAL_QUALITY_PARAMETERS	(11,12)	LR Wald Graphic	Aus R Aus R Aus R
LATENT_VARIABLE	(11,12)		Verknüpfung über Datenbank, Werte über R-Import
STATISTICAL_VARIABLE	(11,12)		Datenbank, Import aus R
AGGREGATE	(11,12)		Manuell, Import
PERSONAL MODEL (PERS)			
GROUP	(6)		Manuell
PERSON	(6)		Import
PERSONAL_DATA			Manuell/Import
PERSONAL_ITEM_RESPONSE	(6,7)	Indicator	Import
PERSONAL_ITEM_RESPONSE_TYPE	(7)	Indicator	Import
PERSONAL_ITEM_SCORE	(9)		Datenbank
PERSONAL_TRAIT_SCORE	(11,12)		Import aus R
PERSONAL_AGGREGATED_SCORE	(10)		Datenbank
PERSONAL_MANIFEST_SCORE	(10)		Datenbank

Tabelle 5 Zusammenhang zwischen Informations- und Vorgehensmodell

7.4 Schritt 0 – Initialization

Vor der erstmaligen Anwendung des Modells ist es sinnvoll, einige vorbereitende Maßnahmen durchzuführen, um die Standardwerte im Modell zu hinterlegen, auf die beim Start zurückgegriffen werden kann.

Der erste Bereich ist der Bereich der **ABSTRACT DIMENSION**. Hier geht es um die Hinterlegung allgemein akzeptierter Kompetenzdimensionen, auf die bei der Festlegung der domänenspezifischen Kompetenzen Bezug genommen werden kann. Ein typisches Beispiel ist die Untergliederung in Fachkompetenz, Methodenkompetenz, Sozialkompetenz und Selbstkompetenz.

Auch die in **OPERATOR** dokumentierten Operatoren können unter Umständen domänenübergreifend genutzt werden. Es bleibt allerdings zu beachten, dass gerade gängige Operatoren wie „bestimmen“, „erstellen“ oder ähnliche Verben in verschiedenen Domänen teilweise mit unterschiedlicher Bedeutung belegt sind.

Im Bereich der eigentlichen Assessments können unter Umständen bereits Rahmenbedingungen für Assessment-Typen im **SETTING** und die Verwendung unterschiedlicher Aufgabenbögen (**QUESTIONNAIRE**) vordefiniert werden. Schließlich kann auch die Vorgehensweise bei Problemen in Assessment-Durchführungen (**EXECUTION PROBLEM**) oder bei der Behandlung standardisierter, nicht domänenspezifischer Fehler (**RAW ERROR TYPE**) vor der eigentlichen Durchführung bestimmt werden.

Im Bereich des Scorings können allgemeingültige Messskalen, beispielsweise dichotome Punktevergaben vordefiniert werden. Schließlich können im statistischen Modell die zu verwendenden Modelltypen, beispielsweise das RASCH-Modell und die zugehörigen Testverfahren ebenfalls bereits dokumentiert werden.

Außerdem kann hier bereits die optionale Modellierung des Bildungsprozesses bis hin zum **EDUCATION MODULE** erfolgen.

COMPETENCE DIMENSIONS (CSD)			
ABSTRACT_DIMENSION	(0)		Manuell
SKILL (SKI)			
OPERATOR	(0,1,4)		Manuell
ASSESSMENT MODEL (ASS)			
SETTING	(0,5)		Manuell
EXECUTION_PROBLEM	(0,5,14)		Manuell
RAW_ERROR_TYPE	(0,5,14)		Manuell
QUESTIONNAIRE	(0,5)		Manuell, zur Generierung des Assessments
SCORE MODEL (SCO)			
MEASUREMENT_SCALE	(0,5,9)		Manuell
SCORE	(0,5,7,9)		Manuell
STATISTICS MODEL (STA)			
STATISTICAL_MODEL_TYPE	(0,5)		Manuell
STATISTICAL_QUALITY_TEST	(0,5)		Manuell
APPLICABLE_TEST	(0,5)		Manuell
FEEDBACK (FBCK)			
EDUCATION_COURSE	(0,8)		Manuell
EDUCATION_MODULE	(0,8)		Manuell
EDUCATION_MATERIAL	(0,8)		Manuell

Tabelle 6 Ziel des Schritts 0 im Informationsmodell

7.5 Schritt 1 - External Analysis

Der erste Schritt bei der Analyse des Domäneninhalts ist die Analyse der relevanten externen Spezifikationen. Im schulischen Bereich sind dies Curricula, Prüfungsanforderungen, Verordnungen und anderes Material, das den relevanten Domäneninhalt beschreibt.

Wesentlich ist, diese Materialien einschließlich ihrer Relevanz, ihres Geltungsbereiches und des Geltungszeitraums zu dokumentieren, da sie letztlich die Grundlage für die Auswahl des relevanten Inhalts darstellen. Dabei entstehen die Grundzüge der Kompetenzdimensionen, die durch diese Materialien definiert werden. Die Wurzel der Dimensionshierarchien entspricht einer extern vorgegebenen Kompetenz (**COMPETENCE SPECIFICATION**).

Außerdem können aus den externen Materialien die Ziele abgeleitet werden. Bei der Definition der Ziele sei beispielhaft auf die SMART-Logik für Ziele hingewiesen, wie sie beispielsweise auch im Projektmanagement gängig ist, also

- spezifisch (das heißt so eindeutig wie möglich)
- messbar (Quantifizierbarkeit)
- akzeptiert¹⁰ (mit den externen Vorgaben und der internen Umsetzung im Einklang)
- realistisch (mit den Ressourcen erreichbar)
- terminiert (Einordnung beispielsweise der Klassenstufe und Unterrichtswoche)

¹⁰ manchmal auch „anspruchsvoll“.

Ziele können sich an verschiedenen Kriterien, beispielsweise an Kompetenzen, an Qualitätsanforderungen oder anderen Maßstäben orientieren. Die Ziele werden mit den externen Spezifikationen verbunden, um deren Ursache ableitbar zu dokumentieren.

Außerdem können den externen Spezifikationen Operatoren entnommen werden, sofern dort spezifische Operatoren vorhanden sind. Anderenfalls sind allgemeine im Vorfeld definierte Operatoren zu verwenden.

DIMENSION SPECIFICATIONS (CSP)			
EXTERNAL_GOAL	(1)		Manuell
CURRICULUM_STANDARD	(1)		Manuell
DEFINE_GOAL	(1)		Manuell
COMPETENCE DIMENSIONS (CSD)			
COMPETENCE_DIMENSION	(1.2)		Manuell
SKILL (SKI)			
OPERATOR	(0,1,4)		Manuell

Tabelle 7 Ziel des Schritts 1 im Informationsmodell

7.6 Schritt 2 - Competence Modeling

Die externe Analyse liefert Vorstellungen von Zielen und Kompetenzen auf einer relativ abstrakten Ebene. Diese müssen daher weiter verfeinert werden. Hierzu bieten sich hierarchische Strukturen an, die auf Einzelkompetenzebene enden. Entscheidend ist, auf welcher feinsten Granularitätsstufe Kompetenzen gemessen werden sollen und können.

Als Ausgangspunkt können zwei Bereiche dienen. Zum einen die **ABSTRACT DIMENSION**, die unabhängig von einer konkreten Spezifikation die grundlegende Vorstellung der Struktur der zu modellierenden Kompetenzen beinhalten kann. Diese können, müssen aber nicht mit den aus der extern vorgegebenen Kompetenz (**COMPETENCE SPECIFICATION**) modellierten Kompetenzdimensionen verbunden werden.

Die eigentliche Gliederung der Kompetenzdimensionen in Kompetenzen kann beispielsweise eine Gliederung aus der externen Spezifikation sein, so in Anlehnung an (Kultusministerium 2009) für die Qualifikationsphase des Fachgymnasiums in Mathematik hier in verfeinerter Form:

Methodenkompetenzen (**ABSTRACT DIMENSION**)

1. Prozessbezogene Kompetenzen für das Berufliche Gymnasium M 11 (**COMPETENCE DIMENSION**)
 - 1.1. Mathematisch argumentieren (**COMPETENCE**)
 - 1.1.1. Die Schülerinnen und Schüler erkennen in inner-und außermathematischen Situationen Strukturen und Zusammenhänge und stellen darüber Vermutungen auf. (**COMPETENCE**)
 - 1.1.2. Die Schülerinnen und Schüler begründen oder widerlegen Aussagen in angemessener Fachsprache mit mathematischen Mitteln und reflektieren ihre Vorgehensweise(**COMPETENCE**)
 - 1.1.3. ...(**COMPETENCE**)
 - 1.2. Probleme mathematisch lösen(**COMPETENCE**)
 - 1.2.1. ...(**COMPETENCE**)
 - 1.3. Mathematisch modellieren(**COMPETENCE**)
 - 1.4. ...(**COMPETENCE**)

Fachkompetenzen (**ABSTRACT DIMENSION**)

2. Inhaltsbezogene Kompetenzen für das Berufliches Gymnasium M 11 (**COMPETENCE DIMENSION**)
 - 2.1. Leitidee: Funktionaler Zusammenhang als Zuordnung zwischen Zahlen und Größen (**COMPETENCE**)
 - 2.2. Leitidee: Verschiedene symbolische Darstellungsformen für Funktionen(**COMPETENCE**)
 - 2.3. ... (**COMPETENCE**)
 - 2.4. Leitidee: Die SuS grenzen in Zusammenhängen lineares und exponentielles Wachstum gegeneinander ab. (**COMPETENCE C110, siehe unten**)
 - 2.5. ...(**COMPETENCE**)
 - 2.6. Leitidee: Die SuS lösen mit der Ableitung von ganzrationalen Funktionen 4. Grades Optimierungsprobleme auch unter Verwendung der eingeführten Technologie. (**COMPETENCE**)

Kompetenzen können auch aus dem Fachwissen der handelnden Personen strukturiert werden. Dies kann von Beginn an oder als Untergliederung der curricularen Gliederung geschehen. Die weitere Unterteilung der Kompetenzen kann unterschiedlich definiert und gegliedert werden. Dass geht soweit, dass man sogar sagen kann, dass die Unterteilung des Kompetenzkonstrukts teilweise in der Praxis willkürlich anmutet (Beyen 2013, 282). Tatsächlich können die durch die entsprechenden Spezifikationen vorgegebenen Kompetenzen in der Praxis zumeist weiter verfeinert werden. Das folgende Beispiel zeigt dies wiederum in Fortführung des Beispiels für die dem Kerncurriculum Mathematik (Kultusministerium 2009, 13) entnommene Kompetenz (C110) (siehe Leitidee 2.4 oben):

(C110) Die Schülerinnen und Schüler grenzen in Sachzusammenhängen lineares und exponentielles Wachstum gegeneinander ab, auch unter Verwendung der eingeführten Technologie.

Dies ist bereits als **COMPETENCE** abgelegt und mit der Dimension verknüpft. Daraus ließe sich beispielsweise die folgende unterhalb der Kompetenz C110 angesiedelte Hierarchie ableiten:

- 2.4 Die SuS grenzen in Zusammenhängen lineares und exponentielles Wachstum gegeneinander ab (C110)
 - 2.4.1. Zuwachs/Abnahme beinhaltet dieselbe Logik
 - 2.4.2. Wachstum wird durch Veränderung gegenüber einem Anfangswert, einem Bestand oder einem noch fehlenden Bestand beschrieben.
 - 2.4.2.1. Lineares Wachstum bezieht sich auf einen Anfangsbestand
 - 2.4.2.2. Exponentielles Wachstum bezieht sich auf den aktuellen Bestand
 - 2.4.2.3. Lineares Wachstum wird durch konstante Veränderung pro Zeiteinheit beschrieben
 - 2.4.2.4. Exponentielles Wachstum wird durch gleiche relative (prozentuale) Veränderung beschrieben
 - 2.4.3. Wachstum kann in verschiedenen mathematischen Funktionstypen beschrieben werden
 - 2.4.3.1. Die Funktionstypen beinhalten Aussagen wozu und wie (relativ/absolut) die Veränderung gemessen wird

2.4.3.2. Proportionalität kann in einer Differenzialgleichung verdeutlicht werden

2.4.3.3. Die Parameter der Funktionen haben logische Bedeutungen für das Wachstum

Es existiert also ein Verfeinerungsbaum auf Ebene der **COMPETENCE** bis zu den detailliertesten auszumodellierenden Einzelkompetenzen.

Die **COMPETENCE LEVELS** entsprechen den intendierten Gliederungsebenen. Dabei können die Kompetenzen derselben Ebene einer Kompetenzdimension dieser Ebene zugeordnet werden. Im oben angegebenen Beispiel gibt es im Bereich der Prozesskompetenzen laut Curriculum zwei Ebenen, im inhaltlichen Bereich nur eine Ebene. Hier sind zwei weitere interne Ebenen ergänzt worden. Kompetenzebenen sollten modelliert werden, wenn die Gleichrangigkeit nebeneinanderliegender Kompetenzen verdeutlicht werden soll. Durch die Zuordnung zu derselben Ebene werden die Kompetenzen auf dieselbe Stufe bezüglich ihrer Dimension gestellt. Kompetenzebenen sind selbst hierarchisch und dürfen nur innerhalb derselben Dimension angelegt werden. Es dürfen auch nur Kompetenzen derselben Dimension einer Ebene derselben Dimension zugeordnet werden. Es ist nicht zwingend, dass jede Ebene in jeder Verfeinerung von Kompetenzen belegt wird. So können unterschiedlich tiefe Verfeinerungen erfolgen und es können auch Ebenen übersprungen werden.

Einer einzelnen zu messenden Kompetenz sollten Kompetenzniveaus zugeordnet werden können, so dass Aussagen über das verfügbare Niveau pro Kompetenz möglich werden. Außerdem muss festgelegt werden, wie das Vorhandensein eines bestimmten Kompetenzniveaus festgestellt werden soll. Bei der Wahl der Kompetenzniveaus kann auf verschiedene Taxonomien wie die von Bloom oder die SOLO-Taxonomie zurückgegriffen werden. Dabei sind die adressierten Bereiche (beispielsweise kognitiv) ebenso zu beachten wie die Allgemeinheit oder Spezifität für bestimmte Fächer. Zudem sind ein-, zwei- und mehrdimensionale Systeme verfügbar. Verschiedene Ansätze sind in Abschnitt 4.2.2.2 beschrieben.

Die Taxonomien können als **COMPETENCE NIVEAU SCALES** dokumentiert werden. Jede Dimension einer Taxonomie bildet eine Niveauskala. Die Stufen der einzelnen Skalen können als **COMPETENCE NIVEAUS** beschrieben werden. Die Anwendung der Niveauskalen auf die einzelnen Kompetenzen wird im **SPECIFIC COMPETENCE NIVEAU** beschrieben. Hier wird für jede zu messende Kompetenz und jede dafür verwendete Niveauskala für jedes Niveau die Bedeutung beschrieben.

Beispiele:

Kompetenz C110 und Anwenden nach Bloom

Lineares und exponentielles Wachstum werden in ihrer Anwendbarkeit auf in der Umwelt vorkommende konkrete Prozesse abgegrenzt.

oder

Kompetenz C110 und Anforderungsniveau I

Die Unterschiede zwischen linearem und exponentiellem Wachstum werden beschrieben.

oder

Kompetenz C110 und Multi-structural nach SOLO

Lineares und exponentielles Wachstum werden hinsichtlich ihrer Anfangswerte, Wachstumsraten, Wachstum und Zerfall und der verwendeten mathematischen Funktionstypen dargestellt.

Das bedeutet, dass für jede einzelne Kompetenz entschieden werden kann, auf welchen Skalen sie gemessen werden soll und festgelegt wird, wie die Erreichung des Kompetenzniveaus definiert ist.

Schließlich kann auch eine Beziehung zu den definierten externen Zielen hergestellt werden.

COMPETENCE DIMENSIONS (CSD)			
COMPETENCE_DIMENSION	(1,2)		Manuell
COMPETENCE_LEVEL	(2)		Manuell
COMPETENCE	(2)		Manuell
COMPETENCE_NIVEAU_SCALE	(2)		Manuell
COMPETENCE_NIVEAU	(2)		Manuell
SPECIFIC_COMPETENCE_NIVEAU	(2)		Manuell

Tabelle 8 Ziel des Schritts 2 im Informationsmodell

7.7 Schritt 3 - Content Modeling

Die **COMPETENCES** der untersten Detaillierungsstufe werden durch konkrete **FACTS** repräsentiert, deren Beherrschung als Indikator für die verfügbare Kompetenz angesehen wird. Die Einbettung in den Gesamtzusammenhang sieht vor, dass die Fakten, die den Inhalt beschreiben, aus den Kompetenzen abgeleitet werden. Dies stellt einen wesentlichen Schritt bei der Operationalisierung dar. Mit den **FACTS** müssen gleichzeitig die **VALENCES**, die **CONSTRAINTS** und die sie ergänzenden Informationen modelliert werden. Letztlich werden Aussagen formuliert, die Zusammenhänge zwischen konkreten oder abstrakten Entitäten (Begriffen) darstellen und deren Wahrheitsgehalt bestimmbar ist. Sind später die Assessment-Teilnehmer in der Lage in konkreten Situationen diesen Wahrheitsgehalt richtig zu ermitteln, wird davon ausgegangen, dass sie die Kompetenzen beherrschen.

In Abbildung 27 sind unter anderem die Fakten F1104 und F1107 modelliert. Diese erlauben folgende Sätze zu bilden:

Alle linearen Prozesse werden beschrieben durch gleiche Veränderung pro Periode (F1104)

und

Alle exponentiellen Prozesse haben in der Exponentialfunktion $f(x) = a \cdot b^x$ den Parameter a als Anfangswert und den Parameter b als Basis- „Wachstumsfaktor“. (F1107)

Die Sätze sind grammatikalisch nicht ganz richtig, aber die Aussagen sind erkennbar. Jeder Fakt ist ein Aussagesatz. Abgeleitet wird F1104 von der Kompetenz C110

Die Schülerinnen und Schüler grenzen in Sachzusammenhängen lineares und exponentielles Wachstum gegeneinander ab, auch unter Verwendung der eingeführten Technologie.

Dies ist ein Aspekt dieser Kompetenz. Es lässt sich daraus auch

Wachstum kann in verschiedenen mathematischen Funktionstypen beschrieben werden.

und schließlich

Die Parameter der Funktionen haben logische Bedeutungen für das Wachstum.

ableiten, was als Basis für die Aussage F1107 dient. Die Gliederung ist letztlich individuell festzulegen.

Bei der Gestaltung eines FACT kann es durchaus auf die Reihenfolge der Valenzen ankommen. Nimmt man das Beispiel aus Absatz 6.2.2.5 so gibt es mehrere Möglichkeiten dieselbe Aussage darzustellen.

(1) Einstein zog im Jahr 1933 in die USA.

Diese Aussage lässt sich syntaktisch umformulieren

(2) Im Jahr 1933 zog Einstein in die USA oder

(3) In die USA zog Einstein im Jahr 1933.

Es ist offensichtlich, dass der Wahrheitsgehalt aller drei Aussagen identisch ist. Wenn eine der drei Aussagen richtig ist, sind es auch die beiden anderen und im Falle einer falschen Aussage würde dasselbe gelten.

Trotzdem scheinen die drei Sätze nicht inhaltsgleich zu sein. Wäre dieser Satz Teil eines Curriculums so scheint die erste Formulierung eher die Betonung auf Einstein zu legen, ist eher eine biographische Aussage, die Teil einer größeren Biografie sein könnte. Die zweite Formulierung betont eher das Jahr 1933, das Besondere dieses Jahres, Sie wäre im Geschichtsunterricht ein prominentes Beispiel für eine Problematik Deutschlands. Der dritte Sachverhalt scheint eher Teil einer Folge von Umzügen in verschiedene Länder zu sein. Die Frage ist also, inwiefern sich der Inhalt des Satzes durch die syntaktische Umstellung geändert hat.

Hilfreich ist hier die Unterscheidung, die Freundlich macht, der den Inhalt in die Bedeutung und den Sinn aufteilt.¹¹ (Freundlich, Sprachtheorie 1970, 155f.). Die Bedeutung eines Satzes ist genau der Satzinhalt, von dem die Wahrheit eines Satzes abhängt (Freundlich 1972, 43). Bereits Schott hat diese Unterscheidung in den pädagogischen Kontext übernommen und die Bedeutung von Aussagen weiter analysiert (Schott 1975, 37f.).

Die Bedeutung der Aussage (1), also die Untersuchung auf deren Wahrheitsgehalt, die im Übrigen nach dieser Definition identisch mit der Bedeutung der Aussagen (2) und (3) ist, erfordert zumindest folgendes Wissen

- 1) wer mit dem Namen „Einstein“ bezeichnet wird,
- 2) was mit dem Jahr „1933“ bezeichnet wird,
- 3) was mit dem Namen „USA“ bezeichnet wird,
- 4) ob eine Relation $_1$ zog im $_2$ in $_3$ mit den entsprechenden Instanziierungen besteht und
- 5) ob die Relation bereits in der Vergangenheit besteht (und nicht erst in der Zukunft bestehen könnte).

¹¹ Freundlich weist darauf hin, dass die Verwendung der Begriffe gegenüber der von Frege (Frege 1892) eingeführten, weit verbreiteten Bedeutung leicht abweicht.

Dies wurde bereits in Abschnitt 6.2.3 zur Motivation der **SNIPPETs** angesprochen. Schott nennt dies Wahrheitsbedingungen (Schott 1975, 38), da ihre Beantwortung die Wahrheit der zugrundeliegenden Aussage beschreibt. Man könnte auch sagen, dass durch sie die Bedeutung der Aussage operationalisiert wird. Sie beschreiben jeweils bestimmte Einzelaspekte beziehungsweise Vorwissen hinsichtlich der Aussage. Sie bilden einzelne Aspekte ab, die zur Beurteilung der Aussage relevant sind und wurden als **SNIPPETs** in das CSP-Informationsmodell aufgenommen.

Auch sollen **SNIPPETs** verwendet werden können, die der Abgrenzung der Aussage zu anderen Aussagen dienen, da auch Gegenbeispiele ein wichtiges Instrument der Pädagogik darstellen und gerade die Fehleranalyse und –klassifikation auf bestimmten „falschen“ **SNIPPETs** aufbauen kann.

Bevor weiter auf die **SNIPPETs** eingegangen wird, soll noch einmal zum zweiten Bestandteil des Inhaltes, dem Sinn zurückgekehrt werden. Er macht den Unterschied zwischen den oben angegebenen bedeutungsgleichen Aussagen aus. Soll in einem Domäneninhalt ein besonderer Sinn herausgehoben werden, der in der Bedeutung der Aussage unterzugehen droht, so kann dies nicht nur durch eine Umstellung, sondern auch durch eine Ergänzung wie hier beispielhaft für die Aussage in der Form (1) geschehen.

- (1) Einstein zog im Jahr 1933 in die USA und er war ein großer Wissenschaftler.

oder

- (1) Einstein zog im Jahr 1933 in die USA und er zog aus Deutschland in die USA und in Deutschland ist die Machtergreifung der Nationalsozialisten.

Alle diese Ergänzungen der ursprünglichen Aussage lassen sich als weitere Valenzen oder Fakten modellieren, die in das System aufgenommen werden können. Der Sinn kann also durch weitere Valenzen mit ihren Bedeutungen modelliert werden. Dies hat auch den Vorteil einer genaueren Modellierung. Folglich kann bei der Modellierung auch eine Reihenfolge wie „Am Anfang steht das Prädikat“ vereinbart werden, ohne dadurch die möglichen Inhalte einzuschränken.

Andererseits verkompliziert es das Modell. Eine weitere Möglichkeit den Sinn genauer zu modellieren, stellt wiederum die Aufnahme zusätzlicher **SNIPPETs** dar, die im nächsten Schritt ergänzt werden. **SNIPPETs** beruhen auch auf der Nutzung von Vorwissen. So wäre es mühsam und nicht zielführend zunächst den Begriff „USA“ modellieren zu müssen, bevor die angegebene Aussage genutzt werden kann. Auch die Verwendung des Jahresbegriffs „1933“ setzt voraus, dass es sich um die christliche Zeitrechnung handelt und dass das aktuelle Jahr bekannt ist oder zumindest bekannt ist, dass es nach 1933 liegt.

Aus dem Vorwissen können sich Probleme ergeben, wenn es in einem Kontext vorhanden ist. Dies kann sogar zu unterschiedlichen Bedeutungen führen. Das Vorwissen kann als Verbindung mit Aussagen angesehen werden, die nicht ausdrücklich genannt werden, aber beim Aufnehmenden vorhanden sind. So bringt Schott ein Beispiel (Schott 1975, 45) zu der Aussage

- **Kolumbus entdeckte Amerika**

indem er darauf hinweist, dass bei entsprechendem Vorwissen diese Aussage als falsch angesehen werden kann, weil bereits vorher dort Menschen gelebt haben und der Kontinent daher nicht zu

entdecken war. Auch Vorwissen über die Fahrten der Wikinger kann Zweifel an der Richtigkeit der Aussage aufkommen lassen.

Daher muss darauf hingewiesen werden, dass es sich bei dem modellierten Domäneninhalt um bewusste pädagogische Entscheidungen handelt. Die Wahrheit der Bedeutung der Aussagen wird als im gegebenen Zusammenhang pädagogisch prädeterniert angesehen, wobei allerdings auch unterschiedliche sich teilweise widersprechende Aussagen erlaubt sind.

Allerdings können die SNIPPETs genutzt werden, um vermeintliche Fehler, die aus derartigen Situationen entstehen, zu lokalisieren und Widersprüche kenntlich zu machen. So ließen sich zum Beispiel zu Kolumbus SNIPPETs formulieren wie

- Kolumbus war Europäer.
- Amerika wurde von Europäern entdeckt, obwohl vorher andere Menschen dort lebten.
- Eine Entdeckung ist etwas für alle Menschen Neues.
- Entdeckungen kann jeder Mensch und jede Gruppe für sich machen.
- In Amerika lebten vor den Europäern Menschen.
- Die Wikinger waren Europäer.
- Die Fahrten der Wikinger nach Amerika sind nicht nachgewiesen.
- Die Fahrten der Wikinger nach Amerika sind nachgewiesen.

Beispielhaft könnte auch eine Unterteilung für FACT 1107 erfolgen.

Die SNIPPETs hierzu beinhalten unter anderem

(SWE2) b als Basis ist der „Basis“-Wachstumsfaktor,

(SWE3) a ist der Anfangswert für das Wachstum,

(SWE4) Der Wachstumsprozentwert/100 + 1 ergibt den Wachstumsfaktor,

(SWE9) a kann am Funktionsgraphen als Schnittpunkt mit der Ordinate abgelesen werden,

(SWE25) der Anfangswert a ist Funktionswert an der Stelle 0.

FACTS (FCT)			
FACT	(3)		Manuell
FACT_TYPE	(3)		Manuell
VALENCE	(3)		Manuell
VALENCE_TYPE	(3)		Manuell
CONSTRAINT	(3)		Manuell
CONSTRAINT_TYPE	(3)		Manuell
ENTITY	(3)		Manuell
LABEL_TYPE	(3)		Manuell
ITEMS (ITM)			
SNIPPET	(3,4)		Manuell

Tabelle 9 Ziel des Schritts 3 im Informationsmodell

7.8 Schritt 4 - Abstract Assessment Modeling

Der Content stellt den Umfang der in einem Assessment prinzipiell zu prüfenden Inhalte dar. Für jede Art von Assessment werden bestimmte Teile davon ausgewählt. Dabei können die zeitliche Nähe zum abgelaufenen Bildungsprozess, die übergreifende Relevanz oder andere Aspekte die Auswahl beeinflussen.

7.8.1 Operationalisierung zu Items

Im Abstract Assessment Modeling werden die relevanten Inhalte soweit operationalisiert, dass eine Reihe zu prüfender Einzelaspekte formuliert werden kann. Dieses Ergebnis in Form einer Menge von theoretisch zu überprüfenden Einzelaspekten bildet die Basis für die Formulierung eines konkreten Assessments. Dies stellt damit eine Art Standardisierung der zu prüfenden Inhalte dar, die bei wiederholten gleichartigen Bildungsprozessen für die Erzeugung neuer, aber vergleichbarer Assessments verwendet werden können.

Bei der Operationalisierung zu **SNIPPET**s und **ABSTRACT ITEM**s werden die Zusammenhänge der Ableitung dokumentiert, so dass die so entstehenden kausal abgeleiteten abstrakten Testitems letztlich zum Content und zu den Kompetenzen zurückverfolgt werden können. In Abbildung 44 ist beispielhaft und in Ausschnitten die Operationalisierung des Fakt 1107 zu **SNIPPET**s und **ABSTRACT ITEM**s zu sehen. Dabei sind die Beziehungen in Fremdschlüssel aufgelöst, die jeweils am Ende der Beispieldatensätze stehen.

7.8.2 Verwendung von Skills

Ziel bei der Operationalisierung ist die Definition von Items, die getrennt genutzt und zu den Content-Bereichen zurückverfolgt werden können. Damit stehen als Beschreibung des Domäneninhaltes **FACT**s, **SNIPPET**s und **ABSTRACT ITEM**s zur Verfügung. Alle drei Typen werden zu **CONTEXT ITEM**s abstrahiert, die jeweils als Repräsentanten dieser Inhaltstypen und Anknüpfungspunkt für **SKILL**s aber auch für andere Teile des CSP dienen.

Diese **SKILL**s werden unter Verwendung der bereits definierten **OPERATOR**en formuliert. Die oben verwendeten Beispiele sollen noch einmal aufgegriffen werden.

BESCHREIBE: Alle linearen Prozesse werden beschrieben durch gleiche Veränderung pro Periode (F1104)

INTERPRETIERE: Alle linearen Prozesse werden beschrieben durch gleiche Veränderung pro Periode (F1104)

Die Fähigkeit(SKILL) wird als Operator(Verb) angegeben: Dieses Verb ist von dem Verb zu unterscheiden, das als Prädikat im Fakt selbst verwendet wird. Durch das Hinzufügen des Operators, hier „BESCHREIBE“ beziehungsweise „INTERPRETIERE“ entsteht aus dem Fakt eine Fähigkeitsbeschreibung, die beobachtbar ist. Es ist erkennbar, ob ein Assessment-Teilnehmer eine Beschreibung, beispielsweise in Funktionsform, erstellen kann. Es ist auch beobachtbar, ob er beispielsweise interpretieren kann, dass, wenn die Zeit auf der Abszisse und der Bestand auf der Ordinate angetragen werden, eine lineare Funktion eine Gerade als Graphen hat, dessen Steigung die Veränderung des Bestandes pro Periode ist und somit wie bei allen linearen Funktionen unveränderlich ist.

Über **SKILL INDICATING** kann für die **SKILL**s festgelegt werden, wie die Erreichung eines der bereits definierten **SPECIFIC COMPETENCE NIVEAU**s bestimmt werden kann. Dabei wird angegeben, wie aus den später ermittelten Scores der Assessment-Teilnehmer auf die verfügbaren Skills (oder **CONTEXT ELEMENT**s) und Kompetenzniveaus geschlossen werden kann. Diese Regeln werden allgemein in den **COMPETENCE NIVEAU METRIC**s beschrieben, die somit die Kompetenzniveaus mit den Skills oder Fakten in algorithmischer Form zusammenführen. Sie

ordnen den Abstufungen der Kompetenzniveaus (die zumeist auf Ordinalskalenebene liegen) die höchste Kompetenzstufe zu, die diesem Niveau entspricht.

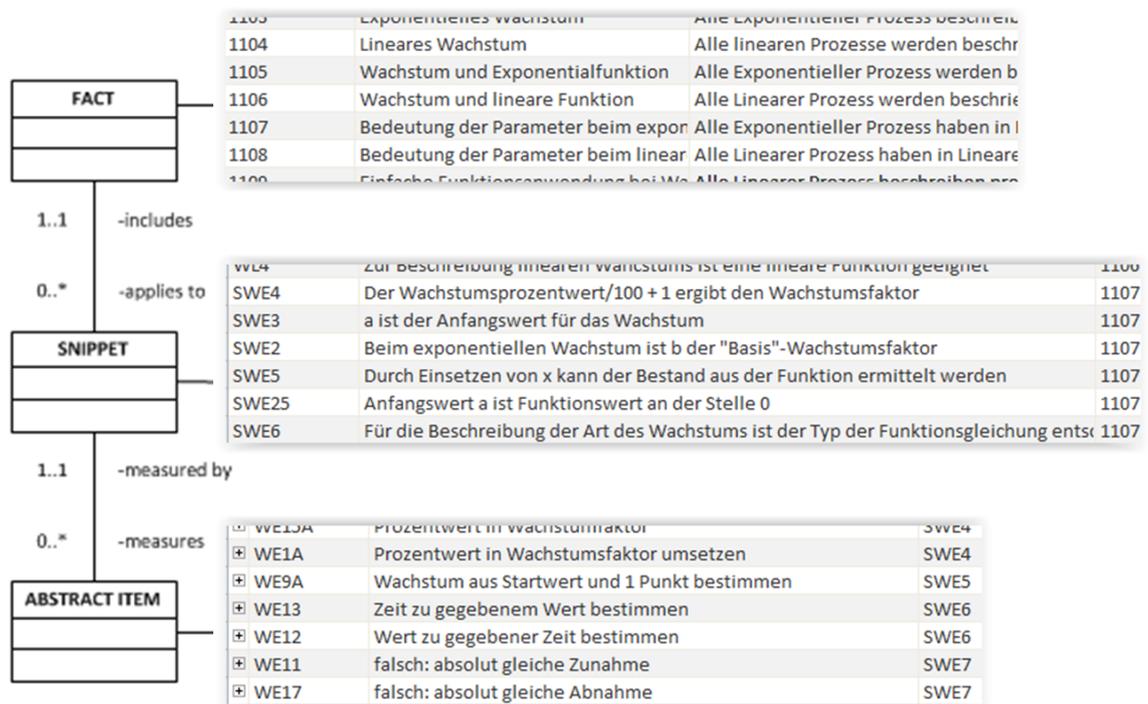


Abbildung 44 Zusammenhang FACT - SNIPPET - ITEM (Ausschnitt)

Ein Verfahren kann dabei die Verwendung von Schwellenwerten sein, die jeweils pro Niveau festgelegt werden und deren Überschreiten durch den Score das Erreichen des entsprechenden Levels anzeigt. Die Stufen beziehen sich auf die Ausprägung hinsichtlich einer konkreten Kompetenz. Ein solches Verfahren in Form der „Competence Niveau Maps“ wird hier vorgestellt.

So kann für einen **SKILL** ein **SPECIFIC COMPETENCE NIVEAU** ermittelt werden. Umgekehrt wird für das Erreichen eines solchen Niveaus bezüglich einer bestimmten Kompetenz im Normalfall nicht ein einziger **SKILL** ausreichen. Für ein Niveau könnte zum Beispiel definiert sein: „Wenn bei Skill 1 mindestens 80% und bei Skill 2 mindestens 60% und erreicht sind, dann wird für die Kompetenz xyz die Stufe 3 erreicht.“ Geht man davon aus, dass die angegebenen Prozentwerte jeweils wie beschrieben in ein Niveau umgesetzt werden können, muss noch die Kombination der beiden **SKILLS** und eventuell weiterer **SKILLS** berücksichtigt werden. Dazu muss eine Regel angegeben werden, die festlegt, welche Stufen zur Erreichung der übergeordneten Stufe führen.

An Stelle der **SKILLS** können vereinfachend auch direkt beispielsweise **FACTs** treten, die die **SKILLS** zu diesen **FACTs** „aggregieren“.

7.8.2.1 Competence Niveau Maps

Im Rahmen des CSP wird hier ein Verfahren vorgeschlagen, bei dem Mapping-Tabellen verwendet werden, die wiederum einzelne Ordinalstufen der Kompetenzen, also die Kompetenzniveaus aufgreifen. Durch Rückgriff auf die bereits in **COMPETENCE NIVEAU** definierten Stufen kann eine Map wie in Abbildung 45 dargestellt erzeugt werden. Eine solche Map beschreibt jeweils eine **SKILL INDICATION**, die auf die zu beschreibende **COMPETENCE** bezogen ist. Insofern

beschreibt sie zumindest einen Aspekt dieses Zusammenhangs. Dies geschieht mittels einer **SPECIFIC COMPETENCE METRIC**.



Abbildung 45 Muster Competence Niveau Map (CNM)

Diese Stufen müssen in den Kompetenzniveaus definiert werden. Bereits die Definition der Bildungsstandards enthält Hinweise wie „Die konkrete Ausgestaltung ist in der Praxis nicht einfach. (Bildungsstandards der Kultusministerkonferenz 2005, 17). Dort findet man auch Hinweise auf die Anforderungsniveaus, die als empirische Orientierungshilfe dienen können. Das IQB hat erste Vorschläge für einige Fächer erarbeitet. So sieht ein Vorschlag für den Sekundarbereich I in Mathematik die Abstufungen, wie in Abbildung 46 angegeben, vor. Die Abstufung wird für Haupt- und Realschulen um eine Stufe versetzt auf den Niveaus 1 bis 5 verwendet [IQ12].

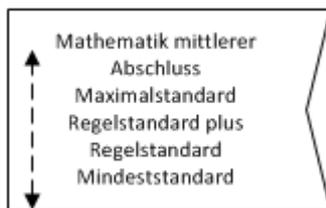


Abbildung 46 CNM Mathematik mittlerer Abschluss

Das CSP-Modell kann diese „Competence Niveau Maps“ (CNM) als graphische Darstellung auf Basis des konzeptionellen Modells verwenden. Dabei werden Schwellenwerte, die in die CNM eingefügt werden, als ein Verfahren vorgeschlagen, das als **COMPETENCE NIVEAU METRIC** dokumentiert und zur Bestimmung der Kompetenzniveaus genutzt werden kann.

Dabei kann die **COMPETENCE NIVEAU METRIC** für die CNM die **LEVELINDICATION** verwenden, in der die Schwellenwerte beschrieben sind. Das Überschreiten eines Schwellenwertes bestimmt das Erreichen einer Kompetenzstufe. Ähnliche auf Schwellenwerte bezogene Vorgehensweisen findet man beispielsweise unter Mitarbeit des IQB zum Fach Deutsch (IQB 2015).

Soweit wird jede Kompetenz durch mindestens eine CNM (**SKILL INDICATION**) mit dem **SKILL** verbunden. Sie erlaubt entsprechend dem Score eines Assessments-Teilnehmers die Festlegung eines Kompetenzniveaus. Werden die **SKILLS** für einen **FACT** aggregiert, kann das Verfahren analog ablaufen. Der **FACT** tritt an die Stelle des **SKILL**.

Bisher wurde von jeweils einem Einflussfaktor für die Kompetenz ausgegangen. Durch Berücksichtigung verschiedener Fähigkeiten (**SKILL** ohne Aggregation zu Fakten) oder verschiedener Fakten bei der Bestimmung eines Kompetenzniveaus für eine Kompetenz können auch verschiedene Aspekte einer Kompetenz berücksichtigt werden. Dies wird eher der Normalfall sein. Dann müssen alle beteiligten **SKILLS** beziehungsweise Fakten in die Bestimmung des Kompetenzniveaus einbezogen werden.

Da es sich bei den Scores um kardinal messbare Werte handelt, kann die konkrete Berechnung des relevanten Scores mittels Gewichtung der Scores der einzelnen **SKILLS** erfolgen. Diese Gewichtung kann in **SKILL INDICATING** also der CNM hinterlegt werden. Werden die **SKILLS** (oder **FACTs**) in eine Grafik mit der CNM aufgenommen, kann, wie in Abbildung 47 gezeigt, die Gewichtung an den Kanten zwischen CNM und **SKILLS** beziehungsweise **FACTs** erfolgen.

Die Zuordnung zwischen **SKILL** und **SPECIFIC COMPETENCE NIVEAU** kann auch mehr als einen Aspekt haben, es können Kompetenzniveaus hinsichtlich verschiedener Fähigkeits- oder Faktenbündel festgelegt werden. Hierzu sei der oben angegebene Satz „Wenn bei Skill 1 mindestens 80% und bei Skill 2 mindestens 60% und ... erreicht sind, dann wird für die Kompetenz xyz die Stufe 3 erreicht.“ noch einmal als Beispiel zitiert. Hier werden verschiedene Skills gewichtet, zwei Kompetenzniveaus bestimmt, also zwei CNMs verwendet, die dann mittels eines UND-Operators zusammengeführt werden.

Diese Zusammenführung kann ihrerseits nicht auf Gewichtungen beruhen, da es sich um ordinal messbare Niveaus handelt. Dementsprechend kann beispielsweise mit einem Operator AND erreicht werden, dass beide CNMs mindestens ein bestimmtes Niveau erreicht haben müssen, damit dieses angenommen wird. Es wird also bei unterschiedlichen Niveaus der beteiligten CNMs das niedrigere Niveau verwendet. Dementsprechend wird bei einem OR das höhere Niveau der beteiligten CNMs gewählt. Der Operator kann entsprechend visualisiert werden.

So könnte beispielsweise die Kompetenz C110 in zwei Teilkompetenzen „Proportionalität“ und „Funktionale Beschreibung“ untergliedert werden.

Die Abbildung 47 zeigt beispielhaft Instanzen für einen Modellausschnitt mit CNMs. Es wird wieder die Kompetenz (C110) „Lineares und exponentielles Wachstum“ betrachtet, bei der es um die Unterscheidungsfähigkeit zwischen beiden Wachstumsarten geht. Zwei CNMs (**SKILL INDICATING**) sind über einen AND-Operator verbunden (**CMN AGGREGATE**). Die beiden CNMs mit denen diese Kompetenz gemessen werden soll, werden als „Proportionalität“ und „Funktionale Beschreibung“ mit ihren CNMs bezeichnet. Beide enthalten eine Liste von vorgeschlagenen Kompetenzniveaus mit Schwellenwerten. Den einzelnen CNMs sind jeweils mit Gewichtungen Fakten zugeordnet.

So kann beispielsweise bei der Beurteilung der Unterscheidungsfähigkeit zwischen linearem und exponentiellem Wachstum einerseits das Verständnis des Proportionalitätsgedankens auf verschiedenen Stufen gemessen werden, andererseits das Verständnis der Abbildung in entsprechende mathematische Funktionstypen. Beide Fähigkeiten sind in sich abgestuft, sind einander jedoch nicht über- oder untergeordnet. Jede CNM wird getrennt bestimmt. Anschließend werden beide durch einen UND-Operator verbunden, so dass bei Unterschieden nur das niedrigere Niveau als erreicht angesehen wird.

Im CSP können so Assessment-Teilnehmern Kompetenzniveaus bezogen auf Kompetenzen, Fähigkeiten und Fakten zugeordnet werden, indem Scores aus den einzelnen Items gebildet werden. Die Reliabilität dieser Zuordnung kann durch eine psychometrische Analyse mittels des RASCH-Modells unterstützt werden.

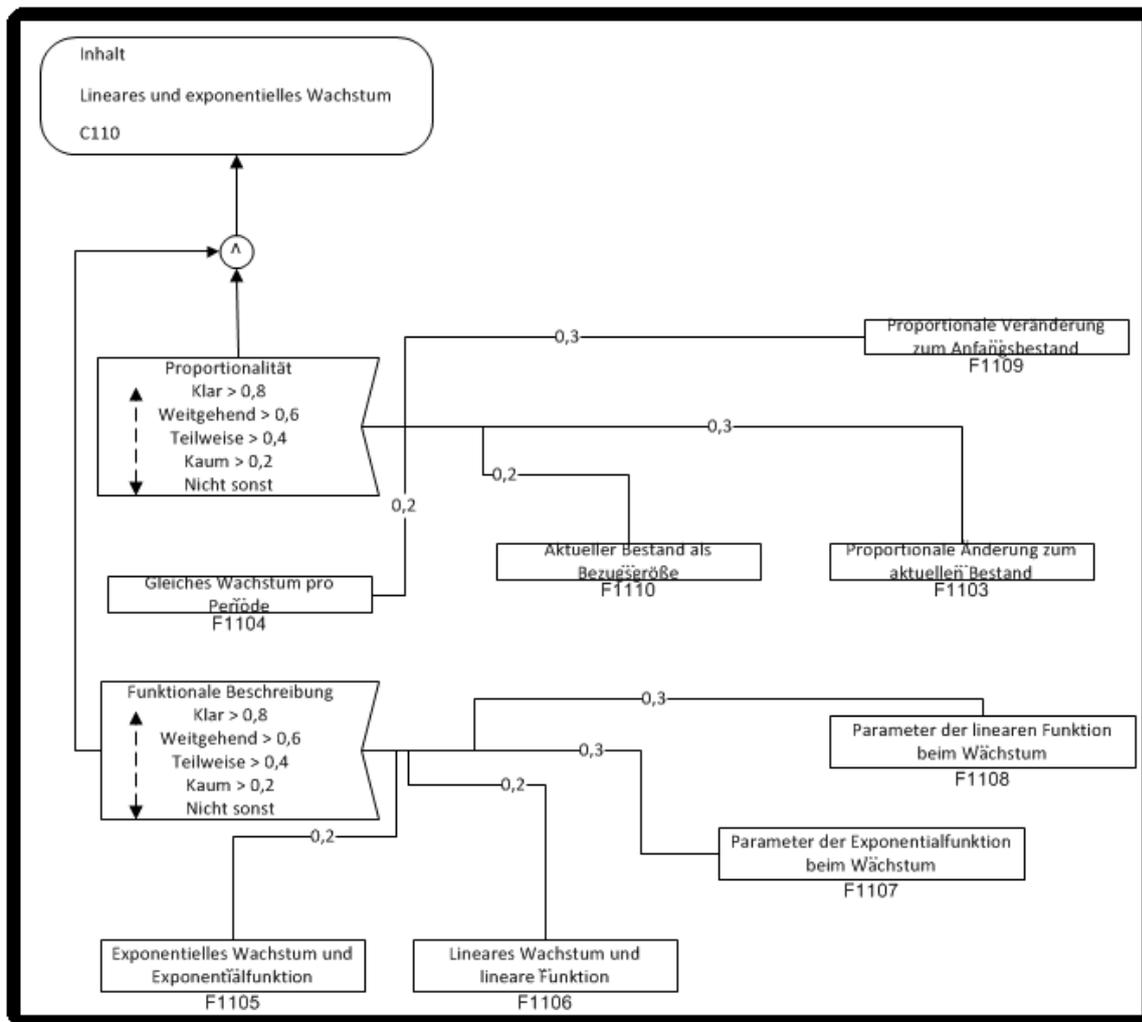


Abbildung 47 CNMs zur Visualisierung des Zusammenhangs zwischen Kompetenz und Fakten

7.8.2.2 Operatoren und Kompetenzniveaus

Eine etwas ähnliche Vorgehensweise, die ebenfalls die CNMs nutzt, lässt sich in diesem Zusammenhang für die Operatoren ableiten, die den Fähigkeiten (Skills) zugeordnet sind. Insbesondere im schulischen Bereich wird gern auf den Begriff der Operatoren zurückgegriffen. Operatoren werden in Standards definiert oder zumindest durch sie eingeführt, wie bereits in Abschnitt 7.4 beschrieben. So orientieren sich kompetenzorientierte Aufgabenstellungen an Operatoren, die wegen ihrer Bedeutung für zentrale Prüfungsaufgaben auch für die entsprechenden Aufgaben im Unterricht von zentraler Bedeutung sind (Kultusministerium 2009, 54ff.) (Hermes und Vaßen 2012, 16ff.). Oft findet man in diesem Zusammenhang auch eine Zuordnung der Operatoren zu den Anforderungsbereichen I bis III.

Aus diesem Grund scheint es sinnvoll, die Operatoren getrennt im Modell zu berücksichtigen und als Standardisierung der Prädikate mit Bezug zu entsprechenden Kompetenzniveaus zu modellieren. Die Operatoren sind somit mit den Kompetenzniveaus verbunden. Daher können die Operatoren statt der Kompetenzniveaus in die CNMs übernommen werden. Da sie ebenfalls mit den entsprechenden Skills verbunden sind, lässt sich die Verdichtung auch nach Operatoren gruppieren, um zu ermitteln welches „Operatoreniveau“ von einem Assessment-Teilnehmer erreicht wurde.

Trägt man anschließend wieder Schwellenwerte in die entsprechende **LEVEL_INDICATIONs** ein, so lässt sich aus dem Grad des Erreichens einer Fähigkeit über den Operator ein **COMPETENCE NIVEAU** ermitteln. Damit könnte aus der **COMPETENCE NIVEAU METRIC** und den Schwellenwerten hinsichtlich der Operatoren eine CNM angegeben werden, die wie in Abbildung 48 aussehen könnte.

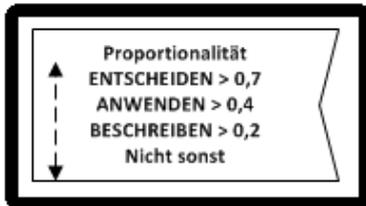


Abbildung 48 **COMPETENCE NIVEAU MAP** mit Operatoren

Selbstverständlich müssen sich die Operatoren auf zugeordnete Fähigkeiten und deren Fakten beziehen. Eine Darstellung wie hier setzt eine Zuordnung zu ordinal messbaren, hierarchisch gegliederten Niveaus voraus. Außerdem sollte ein Assessment sicherstellen, dass die Anforderungen an die untergeordneten Operatoren ebenfalls erfüllt werden.

Ein weiterer Aspekt ist, dass die Operatoren nicht immer mit den Begriffen im Assessment wörtlich übereinstimmen müssen. Deshalb ist bei der Abbildung des Assessments auf eine inhaltlich sinnvolle Zuordnung, nicht unbedingt eine wörtliche Zuordnung zu achten.

ITEMS (ITM)			
SNIPPET	(3,4)		Manuell
ABSTRACT ITEM	(4)		Manuell
CONTEXT_ELEMENT	(4)		Datenbank
SKILL (SKI)			
SKILL	(4)		Manuell
OPERATOR	(0,1,4)		Manuell
COMPETENCE_NIVEAU_METRICS	(4)		Manuell
SKILL_INDICATING	(4)		Manuell
CNM_AGGREGATE	(4)		Manuell
LEVELINICATOR	(4)		Manuell

Tabelle 10 Ziel des Schritts 4 im Informationsmodell

7.9 Schritt 5 – Assessment Modeling

Jim Folgenden wird innerhalb des CSP-Modells der Übergang vom CORE-Modell zum SITUATION-Modell beschrieben.

7.9.1 Assessmentbeschreibung

Bei der Festlegung des Assessments sind zunächst grundlegende Fragen zum Format des Assessments zu klären. Dazu gehört beispielsweise, ob es sich um ein e-Assessment oder eine paper-and-pencil-Klausur handeln soll. Die Dauer, der Ort und die Art der erlaubten Hilfen sind festzulegen und zu dokumentieren, um nur ein paar Beispiele zu nennen. Je exakter hier vorgegangen wird, umso objektiver und reliabler ist das Assessment letztlich.

Die Dokumentation erfolgt innerhalb des Assessment-Modells in der **SETTING EXECUTION** und den **EXECUTION PARAMETERS**. Hier liegt auch die höchste Ebene möglicher Wiederverwendung in dem die bereits vorhandenen **SETTINGS** mit ihren **QUESTIONNAIRES** geprüft werden, inwieweit sie unmittelbar für eine weitere Durchführung verwendet werden können.

Im **EXECUTION PROBLEM** können bekannte Typen von Problemen während der Durchführung beschrieben und empfohlene Aktionen dokumentiert werden. Dies kann beispielhaft der Ausfall eines Computers vor oder während des Assessments sein. Dementsprechend können erwartete Fehler dokumentiert werden, die nicht primär aus dem Domäneninhalt resultieren, beispielsweise die Verwechslung des Dezimalzeichens Komma mit dem Punkt. Zusätzlich kann auch der Umgang damit dokumentiert werden. Dies geschieht mittels des Typs **RAW ERROR TYPE**.

7.9.2 Statistisches Modell

Anschließend ist zu klären, ob ausschließlich mit manifesten Variablen gearbeitet werden oder auch auf latente, mit statistischen Methoden zu analysierende Variablen zurückgegriffen werden soll. Sind solche Variablen vorgesehen, muss ein geeignetes stochastisches Modell gewählt werden. Hierfür stehen grundsätzlich die Methoden der klassischen Schätz- und Testtheorie (CTT) und psychometrische Verfahren der Item Response Theory (IRT) zur Verfügung.

Die Dokumentation erfolgt im Statistik-Modell. Der **STATISTICAL MODEL TYPE** und die **STATISTICAL QUALITY PARAMETERS** sind gegebenenfalls einzurichten, sofern diese noch nicht vorhanden sind. Dazu gehört auch die Pflege der **STATISTICAL QUALITY TESTS** als Verbindung dieser beiden, die dokumentiert, welche Tests auf welche Modelle anwendbar sind. Generell sind dies Dokumentationen, die übergreifend einzurichten sind und für Assessments zur Verfügung stehen sollten. In jedem Fall muss ein erstes **STATISTICAL MODEL** definiert werden. Außerdem sind die latenten Variablen festzulegen, sofern dies nicht schon geschehen ist.

7.9.3 Item-Design

Als Nächstes muss das eigentliche Item-Design erfolgen, also die Ausformulierung der Aufgaben. Die Basis stellen die Vorgaben des Abstract Assessment Modeling dar. Zu den **ABSTRACT ITEMS** müssen ein oder mehrere konkrete Items, die **ASSESSMENT ITEMS**, formuliert werden, die mit den abstrakten Items über **CAUSES_ITEMS** verbunden werden. Das schließt auch die Möglichkeit ein, dass ein Item mehrere Aspekte des Contents messen kann. Die Items können also für verschiedene Wege in der Aggregation genutzt werden (Wilson 2004b, 5). Daher ist es wichtig, dass sie über **CAUSES_ITEM** nicht nur mit den **ABSTRACT ITEMS** verbunden werden, die ihre Entwicklung verursacht haben, sondern auch mit einem **SKILL**, der bereits die Fähigkeit beinhaltet, die hinsichtlich des **ABSTRACT ITEMS** mit dem gerade definierten **ASSESSMENT ITEM** gemessen werden soll.

An dieser Stelle ist ebenfalls die Auswahl aus einem Pool vorhandener Aufgaben aus anderen Assessments denkbar, die im **ASSESSMENT POOL** mit den **ASSESSMENT ELEMENTS** dokumentiert sind. Dadurch kann ein wachsender wiederverwendbarer Pool angelegt werden. Hinzu kommt, dass zu den wiederverwendeten **ASSESSMENT ELEMENTS** bereits Ergebnisse hinsichtlich deren möglichen Antworten aus früheren Assessments verfügbar sein können. So kann das eigentliche Item-Design schrittweise automatisiert und die bisherigen Modellierungsaufwände können genutzt werden. Es erscheint auch wünschenswert, neben den reinen Items bestimmte Lösungsstrategien und Lösungswege als Prozesse zu dokumentieren, was aber nicht Teil dieser Beschreibung ist.

7.9.4 Item-Gruppierung

Beim Entwurf des Assessments werden im allgemeinen Gruppen gebildet, die sich inhaltlich oder technisch ergeben. Die oberste Gruppierungsebene stellt ein **QUESTIONNAIRE** dar. Er enthält alle **EXERCISEs** eines Assessments. Ein Assessment als **SETTING** kann mehrere **QUESTIONNAIREs** enthalten, um unterschiedliche Aufgaben (A/B) innerhalb eines **SETTINGs** zu erlauben. Im Regelfall dürfte ein **QUESTIONNAIRE** genügen. Die Aufgaben stellen als **EXERCISE** die oberste logische Ebene dar. Sie können beliebig in **ASSESSMENT ITEM SETs** hierarchisch strukturiert werden. Jede derartige Gruppierung stellt eine Gruppierung für die Zwecke des Assessments dar. Diese Gruppierung kann unter verschiedenen Aspekten zusammengestellt werden, die beispielsweise mit dem technischen Aufbau des Assessments zu tun haben. Sie kann aber auch der Struktur der **ABSTRACT ITEMs** und der dahinterliegenden **SNIPPETs** entsprechen. In jedem Fall muss es keinen Bezug zu dem Scoring, also zur Bewertung geben und am Ende müssen die **ASSESSMENT ITEMs** stehen. In Abbildung 49 ist ein Beispiel eines **ASSESSMENT ITEM SET** dargestellt, das aus sieben einzelnen **ASSESSMENT ITEMs** besteht und zugleich eine **EXERCISE** sein kann.

6. Entscheiden Sie welche der folgenden Aussagen richtig sind:		
	ja	nein
Beim exponentiellen Wachstum kommt in jeder Zeiteinheit gleich viel hinzu.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gilt eine Funktionsgleichung der Form $f(x)=ax+b$.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gilt eine Funktionsgleichung der Form $f(x)=a*b^x$	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gibt in $f(x)=ax+b$ das b an, wie viel am Anfang vorhanden ist.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gibt in $f(x) = a*b^x$ das a an, wie viel am Anfang vorhanden ist.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum kommt prozentual je Zeiteinheit gleich viel dazu.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gilt, je mehr da ist, umso mehr kommt je Zeiteinheit dazu.	<input type="radio"/>	<input type="radio"/>

Abbildung 49 Beispiel für ein **ASSESSMENT ITEM SET**

Auf jeder Ebene von der **EXERCISE** bis zum **ASSESSMENT ITEM** besteht eine Abstraktion zum **ASSESSMENT ELEMENT**. Diese Elemente können Teile eines Pools sein und ihnen kann zusätzliches erläuterndes **MATERIAL**, beispielsweise in Form von Diagrammen, Testbausteinen, Zeitungsausschnitten oder anderen ergänzenden Materialien zugeordnet werden.

7.9.5 Response-Modeling

Zu den **ASSESSMENT ITEMs** können und sollen mögliche Lösungen dokumentiert werden. Diese können als Vorgaben direkt Bestandteil des Assessments in Form geschlossener Aufgaben werden. Sie können aber auch erst den tatsächlichen Lösungen zugeordnet werden. Zudem sind eine Fehlerklassifikation und die Verbindung mit Feedbackmaterial sinnvoll. Des Weiteren können bereits Bewertungsentscheidungen getroffen und dokumentiert werden. Dies soll im Rahmen des Response-Modeling geschehen.

Bereits vor der Modellierung der **ASSESSMENT ITEMs** ist es hilfreich, sich zu verdeutlichen, dass jede Art von Faktenerhebung oder Prüfung letztlich eine Entscheidung darüber beinhaltet, wie detailliert die Inhalte und die Ergebnisse vorher spezifiziert werden oder inwiefern sie auf Basis der Ergebnisse festgelegt werden. Hilfreich ist hier eine Übersicht, wie sie Tabelle 11 zeigt.

Die Übersicht zeigt sehr unterschiedliche Ansätze bei der Ausführlichkeit und Genauigkeit der Vorgabe für ein Assessment. Mit jeder weiteren Zeile wird der Umfang der bereits vor der Durchführung des Assessments definierten Bestandteile größer. Häufig hat man gerade bei e-Assessments nur die letzte Zeile mit einer Vordefinition aller Items einschließlich fester Antworten im Rahmen des Assessment-Designs im Blick. Tatsächlich sind aber durchaus hier höhere Freiheitsgrade möglich.

Geht man von den unteren Zeilen in Richtung der oberen Zeilen so findet man schrittweise offenere Formate. Selbst die teilnehmende Beobachtung hat ihren Platz, so beispielsweise Gespräche mit Schülern, wie sie in der Praxis auch erwähnt werden (Beyen 2013, 285). Der kleinste gemeinsame Nenner ist zumeist die Festlegung des zu untersuchenden Themas, wobei selbst das in manchen Bereichen nicht vorher erfolgt. Anthropologen gehen im Rahmen der beobachtenden Teilnahme soweit, selbst dieses teilweise erst später festzulegen und die untersuchten Personen sind sich der Beobachtung oft gar nicht bewusst.

Vordefinition	Zu messendes Element	Allgemeine Komponenten	Spezifische Komponenten	Items ohne Bewertung	Items mit Bewertung	Items mit Antworten
Beobachtung	Vorher oder Nachher	Nachher	Nachher	Nachher	Nachher	Nachher
Allgemeine Themenliste	Vorher	Vorher	Nachher	Nachher	Nachher	Nachher
Spezifische Themenliste	Vorher	Vorher	Vorher	Nachher	Nachher	Nachher
Offene Fragen	Vorher	Vorher	Vorher	Vorher	Nachher	Nachher
Offene Fragen mit Bewertungshinweisen	Vorher	Vorher	Vorher	Vorher	Vorher	Nachher
Feste Antworten	Vorher	Vorher	Vorher	Vorher	Vorher	Vorher

Tabelle 11 Konkretisierung des Assessments (Wilson 2004b, 51)

Klassische paper-and-pencil-Klausuren bewegen sich heute auf der Ebene der offenen Fragen mit oder ohne vorherige Festlegung eines Erwartungshorizonts. Dieser existiert zumeist nicht in so detaillierter Form, dass alle möglichen Antworten oder Lösungswege vorher bewertet worden sind.

Tatsächlich lassen sich Assessments mit einer vorherigen Festlegung der Items deutlich effizienter analysieren, so dass die letzten drei Zeilen in Tabelle 11 im Interesse einer effizienten Analyse zu bevorzugen sind.

Wilson empfiehlt, unabhängig davon, wie die einzelne Entscheidung ausfällt, in jedem Fall die Zeilen der Matrix von oben nach unten konsequent Schritt für Schritt zu gehen. Das Auslassen einzelner Schritte führt leicht zu Validitätsproblemen. Das Design der Items für das Assessment geschieht also in mehreren Stufen. Wichtig ist dabei möglichst gut die Vorgehensweise, die Entscheidungsgrundlagen und die Rahmenbedingungen zu dokumentieren. Bei allen Versuchen dieses Vorgehen zu standardisieren bleibt letztlich immer ein Rest an Intuition und Erfahrung, die nicht zu dokumentieren ist, „partly science partly art“ (Wilson 2004b, 53). Andere Ansätze zur

Item-Generierung bestätigen diesen Einfluss des Expertenwissens, so empfiehlt Bayrhuber Items durch Experten der Fachdidaktik auf Basis von

- Schulbuchaufgaben wegen der curricularen Validität und
- Schülerinterviews

zu erstellen (Bayrhuber, et al. 2010, 31).

Das CSP unterstützt das oben beschriebene Vorgehen indem es genau ein solches Schritt-für-Schritt Verfahren ermöglicht. Der Assessment-Entwurf wird zugleich entlang der Typen des Modells dokumentiert. Somit wird beginnend mit der Content-Modellierung ein schrittweiser Prozess bis zum Item-Design und Response-Modeling durchlaufen. Die ersten Schritte in Form des Content Modeling und auch des Abstract Assessment Modeling sind nicht für jedes Assessment zu wiederholen, da sie allgemein im CORE-Modell hinterlegt sind.

Der Übergang zwischen CORE-Modell und SITUATION-Modell liegt auf der Ebene der „Offenen Fragen“ gemäß Tabelle 11. Die Beobachtung entspricht der Kompetenzmodellierung, die allgemeine und spezifische Themenliste der Content-Modellierung¹². Die Items ohne Bewertung sind die **ABSTRACT ITEMS** des CORE-Modells, es können aber auch im Assessment-Modell in diesem Stadium des Vorgehens noch **ASSESSMENT ITEMS** ohne Bewertung existieren. Die Festlegung der **RESPONSE TYPEs** erfolgt frühestens hier und spätestens beim Outcome-Modeling.

Es werden in der Literatur auch teilautomatisierte Ansätze zur Assessment-Modellierung vorgeschlagen, so beispielsweise von Zeuch als regelgeleitete Item-Entwicklung (Zeuch, et al. 2010), auch Item-Cloning genannt, oder die adaptive Item-Entwicklung zur Optimierung der Fragenanzahl bei Kompetenzclustern, insbesondere Minimierung der benötigten Items (Frey und Seitz 2010). Die Klärung beispielsweise, inwieweit im CSP zusätzlich Informationen für derartige Automatismen abgelegt werden müssen, übersteigt den Rahmen dieser Arbeit. Grundsätzlich ist das Modell für derartige Erweiterungen offen.

1. Sie haben ein Kapital von 20000 Euro. Geben Sie jeweils den Faktor (die Zahl) an mit dem das Kapital multipliziert werden muss, wenn der angegebene Zinssatz gilt, um das Kapital im nächsten Jahr auszurechnen.

bei 4%

bei 10%

bei -4%

bei 0,5%

Abbildung 50 ASSESSMENT ITEM SET EL01, mit EL01.1 bis EL01.4 als Items

¹² Unterschied zwischen den **FACTs** und den **SNIPPETs**.

correct	definition	description	outcoming_	belongs_to	corresponds_to
F	-9	<empty>	ex-post	EL01.1	unbestimmt
F	other	sonstige	predefined	EL01.1	unbestimmt
F	45		ex-post	EL01.1	unbestimmt
F	0,04	Bestehendes r	predefined	EL01.1	Prozent1
F	0,4		ex-post	EL01.1	Prozent3
F	800		ex-post	EL01.1	unbestimmt
F	50	<empty>	ex-post	EL01.1	unbestimmt
T	1,04	<empty>	predefined	EL01.1	
F	0,96		predefined	EL01.1	Prozent2
F	41365	<empty>	ex-post	EL01.1	unbestimmt

Tabelle 12 Predefined und Ex-post Response Types für Item EL01.1

Sollen die Antworten für die **ASSESSMENT ITEMS** vor Durchführung des Assessments definiert werden, kann dies mittels der **RESPONSE TYPEs** geschehen, bei denen es sich um „predefined“ **RESPONSE TYPEs** handelt. Es sollten die Antwort sowie die Information hinterlegt werden, ob diese als korrekt oder falsch eingestuft wird. Im einfachsten Fall, wie bei den **ITEMs** in Abbildung 49, handelt es sich je Item nur um die Antwort „Ja“ und die Antwort „Nein“. Nimmt man beispielsweise das in Abbildung 50 dargestellte **ASSESSMENT ITEM SET** „EL01“ und daraus das erste **ASSESSMENT ITEM** mit der internen Bezeichnung „EL01.1“ so könnten sich **RESPONSE TYPEs** entsprechend der Tabelle 12 ergeben.

Die mit „predefined“ gekennzeichneten Antworten können in diesem Vorgehensschritt erzeugt werden. So ist der typische Fehler, dass nur die Zinsen berechnet werden (0,04) vorhersehbar, ebenso wie die Verwechslung von Zu- und Abnahme (0,96). Auch generelle Fehler (other) sollten vorgesehen werden. Die anderen Ergebnisse werden erst später im Rahmen des „Outcome Modeling“ ergänzt. Zusätzlich können Fehler bereits klassifiziert (corresponds_to) und als **ERRORTYPE** und **FEEDBACK** dokumentiert werden. Dies wird beim „Feedback Modeling“ genauer betrachtet.

Schließlich kann auch ein Scoring bereits vorbereitet werden. Dazu wird zunächst die **MEASUREMENT SCALE** festgelegt oder eine vorhandene Skala ausgewählt. Die einzelnen **SCOREs** sind innerhalb dieser Skala festgelegt. Generell ist dies aber erst im Schritt 9 - Performance Analysis - notwendig.

ASSESSMENT MODEL (ASS)			
SETTING	(0,5)		Manuell
SETTING_EXECUTION	(5)		Manuell
EXECUTION_PARAMETERS	(5)		Manuell
EXECUTION_PROBLEM	(0,5,14)		Manuell
RAW_ERROR_TYPE	(0,5,14)		Manuell
QUESTIONNAIRE	(0,5)		Manuell, zur Generierung des Assessments
CAUSES_ITEM	(5)		
ASSESSMENT_ITEM	(5)	Inhaltliche Nachträge	Manuell, zur Generierung des Assessments

ASSESSMENT_ITEM_SET	(5)	Inhaltliche Nachträge	Manuell, zur Generierung des Assessments
EXERCISE	(5)		Manuell, zur Generierung des Assessments
ASSESSMENT_ELEMENT	(5,14)		Manuell
MATERIAL	(5,14)		Manuell
STATISTICS MODEL (STA)			
STATISTICAL_MODEL_TYPE	(0,5)		Manuell
STATISTICAL_QUALITY_TEST	(0,5)		Manuell
APPLICABLE_TEST	(0,5)		Manuell
STATISTICAL_MODEL	(5,11,12)		Manuell
FEEDBACK (FBCK)			
ERRORTYPE	(5,7,8)		Manuell
FEEDBACK	(5,7,8)		Manuell
FEEDBACK_SCN	(5,7,8)		Manuell
SCORE MODEL (SCO)			
MEASUREMENT_SCALE	(0,5,9)		Manuell
SCORE	(0,5,7,9)		Manuell
RESPONSE_TYPE	(5,7)		Manuell und aus den Antworten
MANIFEST_VARIABLE	(5,9,10)		Aus Items
SCORING	(5,7,9)	X	Datenbank

Tabelle 13 Ziel des Schritts 5 im Informationsmodell

7.10 Schritt 6 - Assessment Execution

Auch die Durchführung des Assessments selbst ist zu dokumentieren. Räumliche und zeitliche Durchführung, Anzahl der Teilnehmer, fehlende Teilnehmer und besondere Vorkommnisse müssen festgehalten werden.

Die möglichen Probleme bei der Durchführung des Assessments sind bereits im Typ **EXECUTION PROBLEM** dokumentiert worden. Dort sind die im Vorhinein erkannten Probleme und die Regeln zum Umgang mit diesen Problemen hinterlegt worden. Jetzt müssen die bei der Durchführung tatsächlich aufgetretenen Probleme und die tatsächlich ergriffenen Maßnahmen beschrieben werden. Dafür ist der Typ **EXECUTION PROBLEM ACTION** vorgesehen, wobei nur rein technisch bedingte Korrekturen vorgenommen werden dürfen. Dies können beispielsweise zu rekonstruierende Datensätze sein. Inhaltlich dürfen die Daten nicht verändert werden.

Die Antworten und alle anderen Informationen müssen für jeden Assessment-Teilnehmer im Original erfasst und gesichert werden, sowohl gegen Verlust als auch nachträgliche Manipulation.

Das Ergebnis sind die Rohdaten für alle Assessment-Teilnehmer. Diese beinhalten den Eintrag in **PERSON** für jeden Teilnehmer als zentralen Anker, sowie der **PERSONAL_DATA**, soweit diese verfügbar sind, oder den Eintrag eines entsprechenden Verweises. Außerdem kann die Gruppierung der Assessment-Teilnehmer in **GROUPs** erfolgen. Die tatsächlich ermittelten Daten werden in **PERSONAL ITEM RESPONSE** gespeichert.

ASSESSMENT MODEL (ASS)			
EXECUTION_PROBLEM_ACTION	(6)		Manuell
PERSONAL MODEL (PERS)			
GROUP	(6)		Manuell
PERSON	(6)		Individuell
PERSONAL_DATA	(6)	Externe Daten	Linkaufbau
PERSONAL_ITEM_RESPONSE	(6,7)		Import

Tabelle 14 Ziel des Schritts 6 im Informationsmodell

7.11 Schritt 7 – Outcome Modeling

Das Ergebnis des Outcome Modeling ist der Outcome Space, die Summe aller Einzelbeobachtungen für alle Assessment-Teilnehmer, entsprechend den von technischen Fehlern bereinigten Antwortklassen (RESPONSE TYPES). Das Ergebnis ist somit eine Beobachtung und die Basis für eine Bewertung, ist selbst aber noch keine Bewertung.

7.11.1 Raw-Error-Korrektur

Die Rohdaten werden im Outcome Modeling nach exakt definierten Vorschriften zu Beobachtungen strukturiert, die später einer Bewertung unterzogen werden können. Im Fall eines e-Assessments geht es insbesondere um Korrekturen bei offenen Fragestellungen, beispielsweise im Falle eines Mathematik-Assessments um mögliche Rundungsfehler, um Verwechslungen der Ziffer Null mit dem Großbuchstaben "O", um Missverständnisse über das Eingabeformat eines Funktionsterms oder ähnliche Probleme.

Wichtig im Sinne der Objektivität sind hier klare Regeln für die Vorgehensweise. Diese möglichen Probleme und Korrekturregeln wurden bereits allgemein in den **RAW ERROR TYPES** dokumentiert. Jetzt wird in **RAW ERROR TYPE ACTION** eine Dokumentation der tatsächlich aufgetretenen Fehler und der tatsächlich angewendeten Korrekturregeln vorgenommen. Dabei sollte, soweit möglich, auf die standardisierten Regeln zurückgegriffen werden.

Die Bearbeitung umfasst beispielsweise die Korrektur typischer denkbare Fehler wie

- falsches Dezimaltrennzeichen,
- Datentypprobleme,
- O (Großbuchstabe) und 0 (Null) verwechselt,
- falsche Zeichen oder zusätzliche Zeichen eingegeben oder
- falsche Rundung.

Die Frage ist, inwieweit eine solche Korrektur zulässig ist, da sie die Ergebnisse beeinflusst. Jede Korrektur stellt letztlich eine Manipulation der tatsächlichen Antwort und in Folgeschritten unter Umständen auch der Bewertung dar. Daher hier ist es sinnvoll

- die anzuwendenden Regeln eindeutig zu definieren (Reliabilität).
- die Regeln möglichst automatisiert anzuwenden (Objektivität).
- die Regeln mit dem Ziel der zu überprüfenden Fähigkeiten anzuwenden (Validität) und
- alle Regeln zu dokumentieren.

Abweichende Regeln für einzelne Items können sinnvoll sein.

Beispiel: Es sollen allgemein Rundungsfehler unbewertet bleiben. Im Spezialfall eines Items sollen die Grenzen eines Konfidenzintervalls gerundet werden. Zugrunde liegt eine Ungleichung für die

Wahrscheinlichkeit, die eine Mindestwahrscheinlichkeit für das Konfidenzintervall erfordert. Daher darf das Intervall nur vergrößert werden. Dementsprechend ist Teil des Items die Untergrenze beziehungsweise Obergrenze unterschiedlich zu „runden“. Es kann auch sein, dass genau dieser Aspekt überprüft werden soll. Dann darf hier keine Fehlerkorrektur erfolgen.

Auch unterschiedliche Regeln für die Größe der Rundungsdifferenz können sinnvoll sein, wenn die Zahlen inhaltlich unterschiedlich zu bewerten sind.

Beispiel: Durch falsche Rundung bei der Berechnung der Stelle einer Funktion ergeben sich größere Abweichungen bei der Berechnung des Funktionswertes. Hier muss eine größere Spanne für den Funktionswert erlaubt werden.

Insgesamt scheint es bei der Rundung hilfreich zu sein, den Aspekt der erlaubten Spanne als Teil der Indikatorbestimmung zu nutzen. Hier können dann Ergebnisspannen verwendet werden. Das erlaubt zudem das Problem der zu starken Rundung zu beheben, also fehlende Stellen virtuell durch Nullen zu ergänzen, was bei einer tatsächlichen Ergänzung und einem Vergleich auf Gleichheit schwierig ist.

Man sieht, dass die Regeln sorgfältig und oft im Einzelfall festzulegen sind. Entscheidend ist letztlich, dass durch die Korrekturen die zu messenden Kompetenzen beziehungsweise der relevante Domäneninhalt nicht oder nur marginal betroffen sind.

Hilfreich kann bei der Festlegung der Regeln sein, zu überprüfen, inwieweit man dies bei einer klassischen Klausur ähnlich getan hätte.

7.11.2 Outcome Space

Nachdem die Fehler, die nicht die zu messenden Kompetenzen unmittelbar betreffen, korrigiert worden sind, müssen die vorliegenden Antworten mit den im Vorfeld definierten Antworten abgeglichen werden. Je nach Vorgehen sind hier in mehr oder weniger großem Umfang „Predefined“ **RESPONSE TYPE**s dokumentiert worden (siehe Abschnitt 7.9.5).

Entsprechen alle tatsächlichen Antworten diesen vordefinierten Antworten im **RESPONSE TYPE** sind für ein solches Item keine weiteren Schritte notwendig, da bereits alle für eine spätere Bewertung notwendigen Schritte im Assessment Modeling erfolgt sind. Anderenfalls hängt das weitere Vorgehen vom Umgang mit dem Outcome Space ab. So sind als Lösungsmöglichkeiten beispielsweise die Zuordnung bisher nicht antizipierter Antworten zu einer „Sonstiges“-Kategorie („other“), die direkte Übernahme neuer Antworten als eigener **RESPONSE TYPE** oder die Zuordnung zu einer anderen **RESPONSE TYPE** Kategorie denkbar. In jedem Fall muss die Antwort einem **RESPONSE TYPE** zugeordnet werden. Neue **RESPONSE TYPE**s werden als „EX-POST“ gekennzeichnet.

Bei anderen, offeneren Formen des Assessments ist an dieser Stelle unter Umständen erheblich mehr Arbeit notwendig. So muss beispielsweise bei einer verbalen Interpretation ein umfangreicher Textanalyseprozess erfolgen, der die relevanten **RESPONSE TYPE**s erst ermittelt. Dies ist aber nicht Gegenstand dieser Arbeit und wird in Abschnitt 10.1 nur kurz thematisiert.

Die Antworten der Assessment-Teilnehmer werden nach erfolgter Zuordnung des **RESPONSE TYPE**s mit der Standardantwort dieses Typs im PERSONAL-Modell in **PERSONAL ITEM**

RESPONSE TYPE gespeichert und bilden die Grundlage für die Bewertung. Es erfolgt also hier der Übergang von **PERSONAL ITEM RESPONSE** zu **PERSONAL ITEM RESPONSE TYPE**.

Nachgelagert ist der Zuordnung der **RESPONSE TYPEs** die Zuordnung von Fehlerklassen **ERROR TYPEs** und in der Folge entsprechender **FEEDBACK** Angaben. Außerdem sollten für die neuen **RESPONSE TYPEs** die **SCOREs** festgelegt werden. Dies entspricht dem Vorgehen, wie es ebenfalls bereits in Absatz 7.9.5 beschrieben wurde.

ASSESSMENT MODEL (ASS)			
RAW_ERROR_TYPE_ACTION	(7)		Manuell
SCORE MODEL (SCO)			
SCORE	(0,5,7,9)		Manuell
RESPONSE_TYPE	(5,7)		Manuell und mit R aus den Antworten
SCORING	(5,7,9)		INSERT_SCO_SCORING_T (F)
FEEDBACK (FBCK)			
ERROR_TYPE	(5,7,8)		Manuell
FEEDBACK	(5,7,8)		Manuell
PERSONAL MODEL (PERS)			
PERSONAL_ITEM_RESPONSE	(6,7)		Datentransformator
PERSONAL_ITEM_RESPONSE_TYPE	(7)		Datentransformator

Tabelle 15 Ziel des Schritts 7 im Informationsmodell

7.12 Schritt 8 - Feedback Modeling

Sofern noch nicht geschehen, kann jetzt die allerdings optionale Modellierung des Bildungsprozesses bis hin zum **EDUCATION MODULE** erfolgen. Diese Elemente werden an die zu modellierenden **FEEDBACKs** angebunden.

Zudem ist in diesem Schritt die Anbindung der relevanten Bereiche des Messkonstrukts an die Typen des Bereichs Feedback erforderlich. Das sind **SPECIFIC COMPETENCE NIVEAU, SKILL** und **CONTEXT ELEMENT**, die später im Rahmen des Schritts 12 (Competence Estimation) analysiert werden und für die Kompetenzschätzungen und **SCOREs** pro Assessment-Teilnehmer bestimmt werden können. Damit insbesondere auf die Defizite gezielt reagiert werden kann, ist deren Zuordnung zu einem **FEEDBACK** notwendig. Im **FEEDBACK** werden mögliche Ursachen der Defizite dokumentiert. Jedes **FEEDBACK** kann sich auf ein **EDUCATION MODULE** beziehen. Außerdem sind sinnvolle Maßnahmen für eine Intervention in den **ACTIONs** und dem entsprechenden **FEEDBACK MATERIAL** beschrieben. Damit stehen die entsprechenden Dokumente für eine Reaktion auf Defizite zur Verfügung.

Eine noch genauere Analyse und genauere Reaktion lassen sich mittels der Fehleranalyse erreichen. Hierzu ist eine Fehlerklassifikation erforderlich. Ein inhaltlich leicht zugängliches Beispiel einer Fehlerklassifikation findet sich bei Kreschnak (Kreschnak 1985, 61f.). Er beschreibt das schriftliche Subtrahieren am Beispiel:

5031
~~-2945~~
 2086

Die für möglich gehaltenen Ergebnisse werden aufgelistet, wie sie in Tabelle 16 wiedergegeben sind. Unabhängig vom Rang, der die Häufigkeit des Ergebnisses ausdrückt, erkennt man, dass die Fälle 2, 3, 5, 7, 9 und 10 verwandt erscheinen, wobei auch hier keineswegs eine deterministische Regel verwendet werden darf, wie sich in den weiteren Erläuterungen des Beispiels zeigt (Kreschnak 1985, 61ff.). Hier soll dies als Beispiel reichen, was die verschiedenen Fehlertypen repräsentieren können. Im Sinne des CSP-Modells handelt es sich bei den möglichen Ergebnissen um **RESPONSE TYPES**. Diesen **RESPONSE TYPES** werden **ERROR TYPES** zugeordnet, die der Spalte „Hinweis“ in Tabelle 16 entsprechen können. Dabei können durchaus mehrere **RESPONSE TYPES** einem **ERROR TYPE** zugeordnet werden, wenn sie als verwandt angesehen werden.

Rang	Ergebnis	Hinweis
1	2086	Richtiges Ergebnis
2	3086	Linker Übertrag fehlt
3	2186	Mittlerer Übertrag fehlt
4	3914	Spaltenweise kleinere von größerer Zahl subtrahiert
5	3196	Alle Überträge fehlen
6	-	Kein sinnvoller Erklärungsversuch möglich
7	2096	Rechter Übertrag fehlt
8	7976	Richtiges Ergebnis beim Addieren
9	3186	Linker und mittlerer Übertrag fehlen
10	2196	Mittlerer und rechter Übertrag fehlen

Tabelle 16 Fehlertypen für die schriftliche Subtraktion

Ein oder mehrere **ERROR TYPES** werden einer möglichen Ursache, einem **FEEDBACK** zugeordnet. Auch die Verbindung zu weiterführenden Materialien kann genutzt werden.

Die Modellierung der möglichen Fehler kann bereits auf Basis der **FACTs** oder **SNIPPETs** erfolgen. Sie wird zumeist aber nicht vollständig sein und kann beim konkreten Assessment verfeinert oder im Nachhinein ergänzt werden. Der Schwerpunkt sollte auf der Anbindung der Kompetenzen liegen, die detaillierte Fehleranalyse ist hier nur unterstützend und auch nur punktuell zu sehen.

FEEDBACK (FBCK)			
EDUCATION_COURSE	(0,8)		Manuell
EDUCATION_MODULE	(0,8)		Manuell
ERROR_TYPE	(5,7,8)		Manuell
EDUCATION_MATERIAL	(0,8)		Manuell
FEEDBACK_MATERIAL	(8)		Manuell
FEEDBACK	(5,7,8)		Manuell
ACTION	(8)		Manuell

Tabelle 17 Ziel des Schritts 8 im Informationsmodell

7.13 Schritt 9 – Performance Analysis

In diesem Schritt geht es um die Bewertung der Ergebnisse für die einzelnen Items. Die manifesten Variablen je **ASSESSMENT ITEM** und die Bewertung für die **RESPONSE TYPES** dieser Variablen sind jetzt festzulegen.

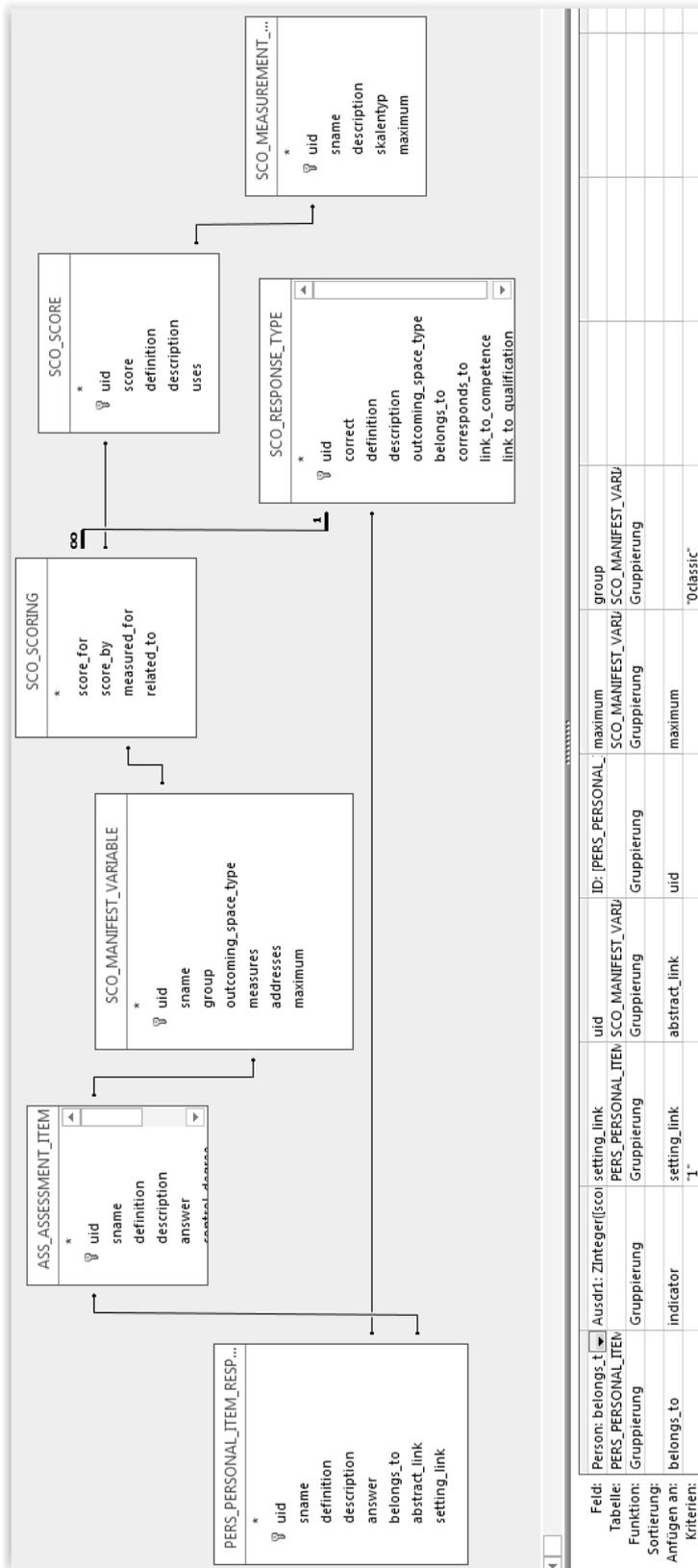


Abbildung 51 Scoring mit PRO_PERS_PERSONAL_ITEM

Insbesondere sind jetzt die Messskalen, also die **MEASUREMENT SCALES** und die zugehörigen **SCOREs** auszuwählen oder neu zu definieren, sofern dies nicht schon geschehen ist. Ferner muss für jedes **ASSESSMENT ITEM** eine **MANIFEST VARIABLE** angelegt werden, die der Bewertung des Items dient. Alle **RESPONSE TYPEs** müssen über **SCORING** mit dem entsprechenden Score

und der manifesten Variable verbunden werden, damit eine Bewertung der **MANIFEST VARIABLE** pro **PERSON** möglich ist.

Dadurch kann in **PERSONAL ITEM SCORE** eine Bewertung hinterlegt werden. Die Berechnung ist in Abbildung 51 strukturell dargestellt. Sie enthält im Ergebnis als Indicator die Bewertung eines Items in einer manifesten Variable (abstract_link) in einer Ausführung eines Assessments (setting_link). Die Spalte belongs_to gibt die Kennung des Assessment-Teilnehmers an, während das Maximum die für dieses Item gemäß der verwendeten Messskala maximale Punktzahl enthält, wie beispielhaft in Abbildung 52 dargestellt. Die so berechnete Bewertung kann in **PERSONAL ITEM SCORE** manifestiert werden, was aus Gründen der Vereinfachung weiterer Auswertungen sinnvoll erscheint. Die Bewertung kann zudem dynamisch in weitere Auswertungen eingebunden werden. Erweiterungen wie die Aufnahme der Kennung der Messskala oder eines Minimums scheinen möglich, um noch größere Flexibilität bei der Verwendung verschiedener Messskalen zu erhalten.

uid	indicator	maximum	belongs_to	abstract_lin	setting_link
100_EL01.1_0	0	1 100	100	EL01.1	1
100_EL01.2_0	0	1 100	100	EL01.2	1
100_EL01.3_1	1	1 100	100	EL01.3	1
100_EL01.4_0	0	1 100	100	EL01.4	1
100_EX01.1_1	1	1 100	100	EX01.1	1
100_EX01.2_1	1	1 100	100	EX01.2	1
100_EX02.1_1	1	1 100	100	EX02.1	1
100_EX02.2_1	1	1 100	100	EX02.2	1
100_EX02.3_0	0	1 100	100	EX02.3	1
100_EX02.4_0	0	1 100	100	EX02.4	1
100_EX03.11_0	0	1 100	100	EX03.11	1
100_EX03.12_1	1	1 100	100	EX03.12	1
100_EX03.3_1	1	1 100	100	EX03.3	1
100_EX03.4_1	1	1 100	100	EX03.4	1
100_EX03.5_1	1	1 100	100	EX03.5	1
100_EX04.1_1	1	1 100	100	EX04.1	1
100_EX04.2_0	0	1 100	100	EX04.2	1
100_EX05.1_0	0	1 100	100	EX05.1	1
100_EX05.2_0	0	1 100	100	EX05.2	1

Abbildung 52 Bewertung in **PERSONAL ITEM SCORE**

Mit der Durchführung dieses Schritts ist die Basis für alle weiteren Bewertungen geschaffen.

SCORE MODEL (SCO)			
MEASUREMENT_SCALE	(0,5,9)		Manuell
SCORE	(0,5,7,9)		Manuell
MANIFEST_VARIABLE	(5,9,10)		Aus Items
SCORING	(5,7,9)		Datenbank
PERSONAL MODEL (PERS)			
PERSONAL_ITEM_SCORE	(9)		Datenbank

Tabelle 18 Ziel des Schritts 9 im Informationsmodell

7.14 Schritt 10 - Manifest Score

Bisher ist die Bewertung der Ergebnisse auf Itemebene erfolgt. Das Ziel dieses Schritts ist eine kumulierte Bewertung bis zum Gesamtergebnis für einen Assessment-Teilnehmer, wobei Gesamtergebnisse auf verschiedenen Ebenen ermittelt werden sollen:

- **SKILLS**,
- **FACTs** (wahlweise auch **SNIPPETs** und **ABSTRACT ITEMs**),
- **COMPETENCEs** und
- insgesamt.

Vier wichtige Aspekte für die Berechnung der kumulierten Scores sind

- die Entscheidung für eine dynamische Berechnung oder eine Manifestierung der Scores in Tabellen,
- die Gruppierung der bewerteten **MANIFEST VARIABLEs**,
- die Gewichtung der Variablen bei der Gruppierung und
- die Aggregationsfunktion.

Bei der Entscheidung zwischen dynamischer Berechnung der Scores und Speicherung der Scores in eigenen Tabellen stehen sich insbesondere die Dynamik bei Veränderungen und die Performanz und Simplizität der Folgeanalysen gegenüber. Daher trifft das Modell keine Entscheidung für eine der beiden Varianten, sondern erlaubt die Nutzung beider Varianten.

Die Gruppierung der manifesten Variablen sollte zielorientiert erfolgen. Es sind sowohl rein hierarchische als auch netzwerkartige Formen möglich. Die Gruppen können jeweils über eine eigene Group-ID identifiziert werden. Um die Aggregation flexibler zu gestalten, können Gewichtungen verwendet werden. So kann beispielsweise festgelegt werden, dass bestimmte Items in unterschiedlicher Gewichtung zur Analyse eines Skills beitragen. Noch flexibler ist die Nutzung unterschiedlicher Aggregationsfunktionen bis zur Festlegung von freien Algorithmen. Hier wird in den Beispielen jedoch durchgehend eine Summation als Aggregationsfunktion verwendet.

7.14.1 Gesamtscore pro Person

Der Gesamtscore pro Person und Setting kann aus den Basisscores in **PRO_PERS_PERSON_SCORE** erzeugt werden.

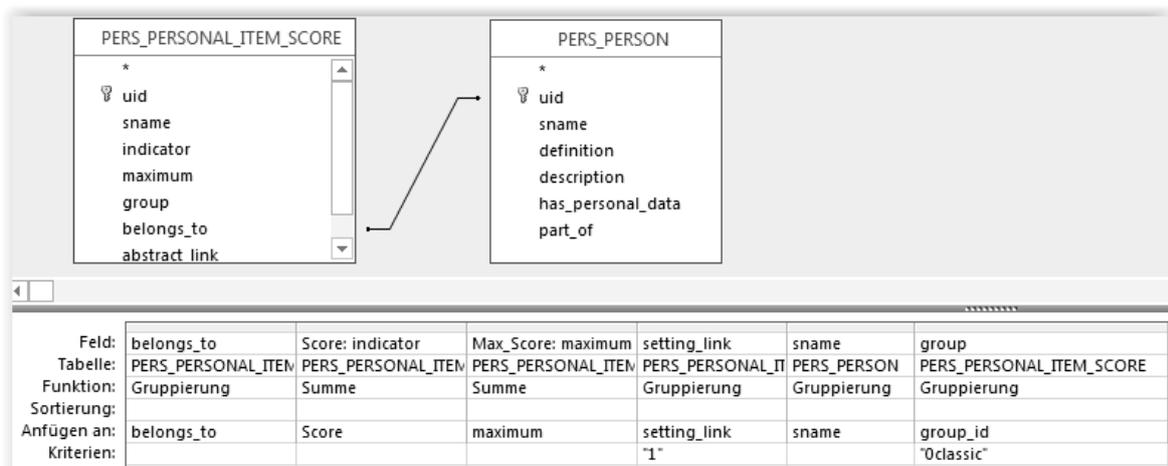


Abbildung 53 Abfrage zur Erzeugung der Gesamtscores **PRO_PERS_PERSON_SCORE**

Dies geschieht beispielhaft mit der Abfrage PRO_PERS_PERSON_SCORE, wobei die Gruppe auf „0classic“ gesetzt wird, da es sich um Basisvariablen handelt.

Zusätzlich kann jede Aggregation oder Teilmenge als Basis anderer Scores verwendet werden, wenn diese benötigt werden. Beispielsweise kann das Ergebnis in PERSONAL MANIFEST SCORE wie in Abbildung 54 aussehen. Die Scores entsprechen Punktzahlen in klassischen Klausuren, über das Maximum kann eine Prozentzahl ermittelt werden. Die Angabe „belongs_to“ verweist auf den Assessment-Teilnehmer.

uid	sname	group_id	score	maximum	belongs_to	setting_link
1		0	63	92	100	1
2		0	63	92	101	1
3		0	64	92	102	1
4		0	63	92	103	1
5		0	56	92	104	1
6		0	65	92	105	1
7		0	75	92	106	1
8		0	52	92	107	1
9		0	63	92	108	1
10		0	53	92	109	1
11		0	61	92	110	1

Abbildung 54 Beispielscores in PERSONAL MANIFEST SCORE

7.14.2 Analyse für Skills

Die Analyse soll nicht nur einen Gesamtwert liefern, sondern auch Ergebnisse für einzelne SKILLS. In PERSONAL ITEM SCORE findet man die manifestierte Bewertung pro PERSON und ASSESSMENT ITEM, beziehungsweise der dieses Item repräsentierenden manifesten Variable. Dies sind die Basisscores für die manifesten Variablen der untersten Stufe. Geht man davon aus, dass auch weitere aggregierte Variable gebildet und Scores für sie ermittelt werden, so sind diese Basisscores doch als relativ konstant anzusehen. Sie bilden eine eigene große Gruppe innerhalb vieler denkbarer Scores. Außerdem können sie für die statistische Analyse genutzt werden und bilden daher einen eigenen Typ im Modell.

belongs_to	consists_of	controlled_	weight
Fact1103b	W107.2	1	100
Fact1103b	W108.1	1	100
Fact1103e	W102.3	1	100
Fact1103e	W103.3	1	100
Fact1103e	W109.1	1	100
Fact1103e	W109.3	1	100
Fact1103e	W109.4	1	100

Abbildung 55 Beispieldaten aus AGGREGATE

Die Aggregationen werden in PERSONAL AGGREGATED SCORE manifestiert. Für die Aggregation muss deren Struktur dokumentiert werden. Dazu werden weitere MANIFEST VARIABLES beschrieben, von denen jede eine Gruppierung repräsentiert. Die zugehörigen Variablen werden über AGGREGATE verbunden. Die Aggregationsregel wird in VARIABLE AGGREGATION RULE ergänzt. In Abbildung 55 sieht man unter „belongs_to“ die Angabe eines Skills (Fact plus Buchstabe für den Operator, also beispielsweise b für berechnen), der gleichzeitig der Verweis auf die MANIFEST VARIABLE gleichen Namens ist. Unter „consists_of“ findet man

die untergeordneten Variablen, hier handelt es sich um **MANIFEST VARIABLES**, die direkte Repräsentanten von **ASSESSMENT ITEMS** sind, also Variablen der untersten Ebene.

Mit „controlled_by“ wird auf die Aggregationsregel verwiesen, unter „weight“ kann ein Aggregationsgewicht angegeben werden. Mittels dieser Beschreibung kann die Aggregation zu Scores für Skills erfolgen.

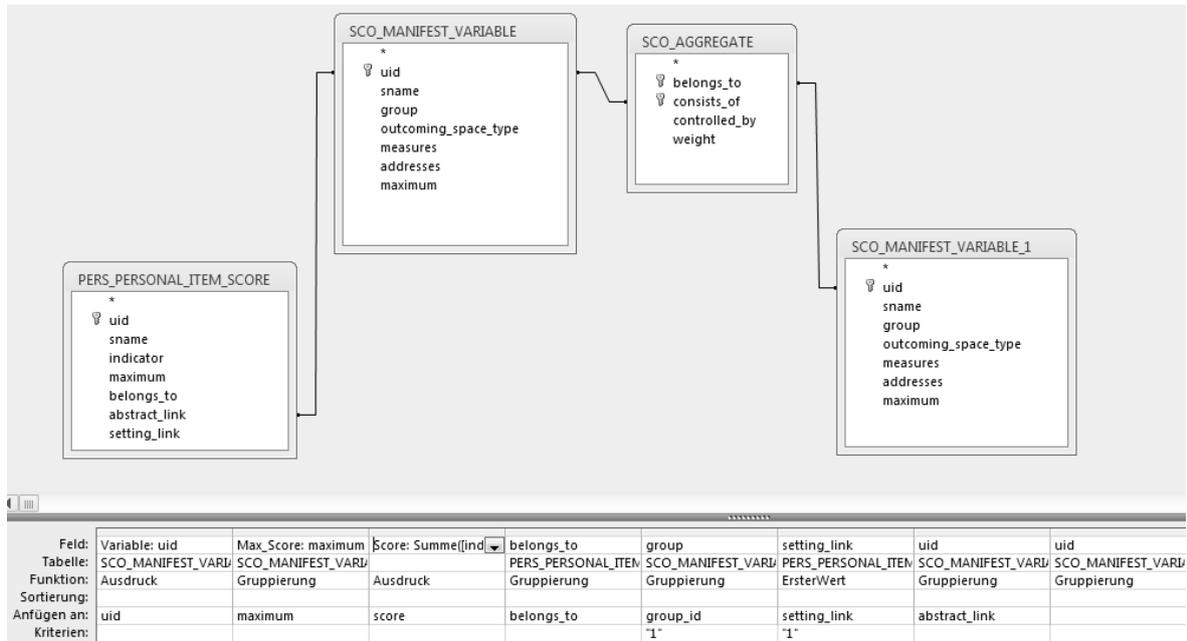


Abbildung 56 Aggregation mit PRO_PERS_AGGR_SKILL

Die Aggregation zu den Scores ist in Abbildung 56 beispielhaft mit PRO_PERS_AGGR_SKILL erfolgt. Hierbei ist bei der Aggregation über Summenbildung auch die Gewichtung in **AGGREGATE** berücksichtigt worden. Somit sind in **PERSONAL AGGREGATED SCORE** die aggregierten Ergebnisse pro Person und Variablengruppe abgelegt worden. Diese Gruppen sind bereits an je einem **SKILL** orientiert, so dass sie als Basis eines Scores je **SKILL** dienen können.

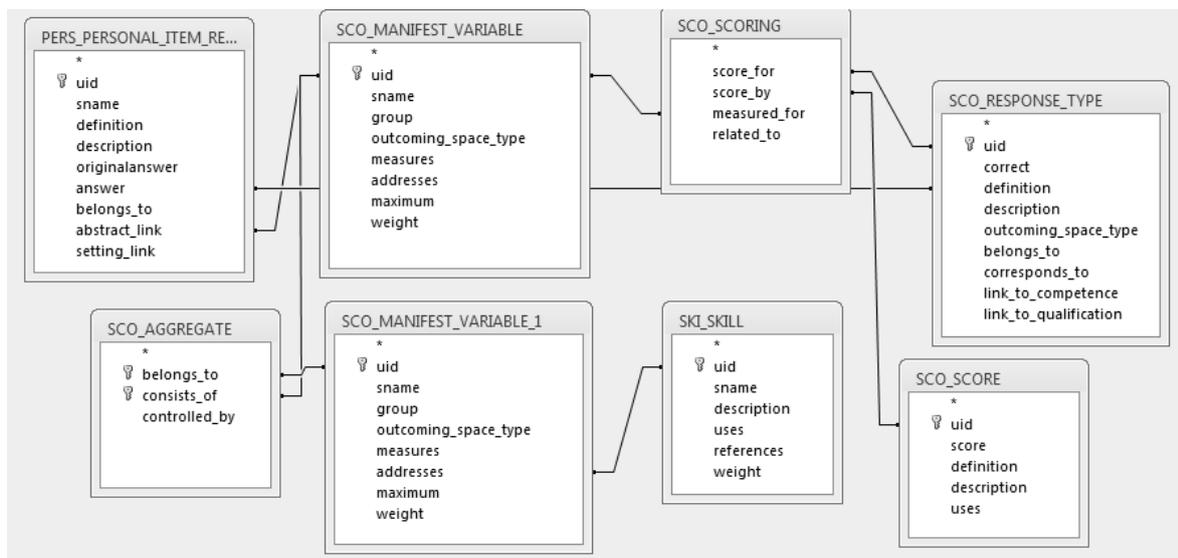


Abbildung 57 Dynamische Ermittlung des Scores Variablengruppe

Alternativ kann das Scoring dynamisch ermittelt werden. Die gewählte Gruppierung in Abbildung 57 ist dieselbe, also an **SKILLS** orientiert. In Abbildung 57 ist jedoch die Grundstruktur einer dynamischen Ermittlung des Scores pro Skill zu sehen. Es wird dabei immer auf das Setting und die Variablengruppe gefiltert.

Person	Variable	Skill	Maximum	Score
100	Fact1103b	bestimmen/ermitteln: Exponentielles Wachstum	2	1
100	Fact1103e	entscheiden: Exponentielles Wachstum	10	8
100	Fact1104e	entscheiden: Lineares Wachstum	3	3
100	Fact1105b	berechnen: Wachstum und Exponentialfunktion	1	1
100	Fact1105e	entscheiden: Wachstum und Exponentialfunktion	8	5
100	Fact1106e	entscheiden: Wachstum und lineare Funktion	4	4
100	Fact1107b	bestimmen/ermitteln: Bedeutung der Parameter beim exponentiellen Wachstum	9	7
100	Fact1107e	entscheiden: Bedeutung der Parameter beim exponentiellen Wachstum	15	12
101	Fact1103b	bestimmen/ermitteln: Exponentielles Wachstum	2	2
101	Fact1103e	entscheiden: Exponentielles Wachstum	10	8
101	Fact1104e	entscheiden: Lineares Wachstum	3	3
101	Fact1105b	berechnen: Wachstum und Exponentialfunktion	1	1
101	Fact1105e	entscheiden: Wachstum und Exponentialfunktion	8	6
101	Fact1106e	entscheiden: Wachstum und lineare Funktion	4	4
101	Fact1107b	bestimmen/ermitteln: Bedeutung der Parameter beim exponentiellen Wachstum	9	4
101	Fact1107e	entscheiden: Bedeutung der Parameter beim exponentiellen Wachstum	15	13
102	Fact1103b	bestimmen/ermitteln: Exponentielles Wachstum	2	2
102	Fact1103e	entscheiden: Exponentielles Wachstum	10	8

Abbildung 58 Ergebnis einer manifesten Analyse pro Person und Skill

Diese Variante ist dynamischer, da das Scoring noch verändert werden kann. Es ist allerdings auch komplexer, darauf in weiteren Analysen zuzugreifen. Daher sollte, wenn die Gruppierung und das Scoring als stabil angesehen werden, die oben angegebene Version genutzt und die Daten in **PERSONAL AGGREGATED SCORE** abgelegt werden.

uid	sname	group_id	score	maximum	belongs_to
Skill: 39	bestimmen/ermitteln: Bedeutung der Parameter beim exponentiellen Wachstum	Skill	7	9	100
Skill: 10	entscheiden: Wachstum und lineare Funktion	Skill	4	4	100
Skill: 1	entscheiden: Wachstum und Exponentialfunktion	Skill	5	8	100
Skill: 12	entscheiden: Lineares Wachstum	Skill	3	3	100
Skill: 46	bestimmen/ermitteln: Exponentielles Wachstum	Skill	1	2	100
Fact:1103	Exponentielles Wachstum	1103	9	12	100
Skill: 18	entscheiden: Bedeutung der Parameter beim exponentiellen Wachstum	Skill	13	15	100

Abbildung 59 Beispieldaten aus PERSONAL AGGREGATED SCORE

In **PERSONAL AGGREGATED SKILL** findet man Scores für Gruppierungen, wie sie beispielhaft in Abbildung 59 dargestellt sind. Die „uid“ ist der Schlüssel und weist hier inhaltlich auch auf den Hintergrund hin, die Gruppe ist beispielsweise Skill für alle **MANIFEST VARIABLES**, die für Skills gebildet wurden, „score“ und „maximum“ haben den bekannten Bezug ebenso wie „belongs_to“ als Schlüssel des Assessment-Teilnehmers. Darauf basierend kann eine Auswertung der erreichten Scores pro **SKILL** und **PERSON** erstellt werden, wie sie in Abbildung 60 angegeben ist.

In Abbildung 61 ist das Ergebnis zu sehen, wobei zusätzlich die prozentual erreichten Anteile pro Person und Skill ergänzt worden sind.

Die jeweiligen Ergebnisse pro **PERSON** und **SKILL** können ebenfalls in **PERSONAL AGGREGATED SCORE** gespeichert werden. Dies ist hier beispielhaft mit **PRO_PERS_SCORE_SKILL_PERSON** für die Skills pro Person erfolgt.

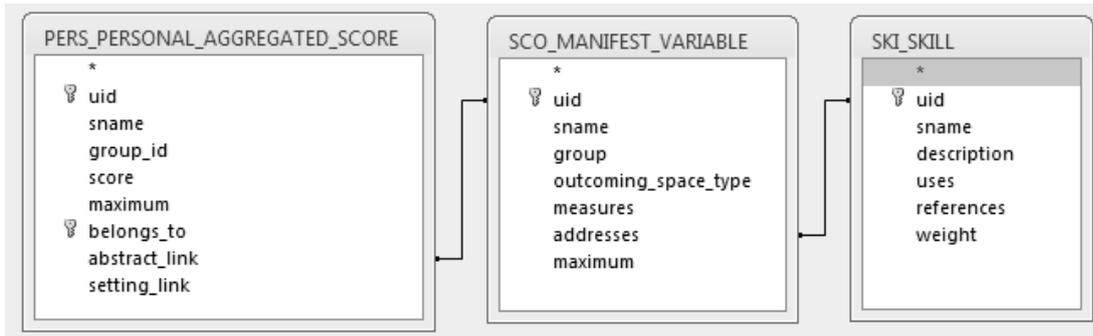


Abbildung 60 Analyse der Scores pro Skill auf materialisierten Scores

Person	Variable	Skill	Maximum	Score	Prozent
100	Fact1103b	bestimmen/ermitteln: Exponentielles Wachstum	2	1	50
100	Fact1103e	entscheiden: Exponentielles Wachstum	10	8	80
100	Fact1104e	entscheiden: Lineares Wachstum	3	3	100
100	Fact1105b	berechnen: Wachstum und Exponentialfunktion	1	1	100
100	Fact1105e	entscheiden: Wachstum und Exponentialfunktion	8	5	63
100	Fact1106e	entscheiden: Wachstum und lineare Funktion	4	4	100
100	Fact1107b	bestimmen/ermitteln: Bedeutung der Parameter beim exponentiellen Wachstum	9	7	78
100	Fact1107e	entscheiden: Bedeutung der Parameter beim exponentiellen Wachstum	15	12	80
101	Fact1103b	bestimmen/ermitteln: Exponentielles Wachstum	2	2	100
101	Fact1103e	entscheiden: Exponentielles Wachstum	10	8	80
101	Fact1104e	entscheiden: Lineares Wachstum	3	3	100
101	Fact1105b	berechnen: Wachstum und Exponentialfunktion	1	1	100
101	Fact1105e	entscheiden: Wachstum und Exponentialfunktion	8	6	75
101	Fact1106e	entscheiden: Wachstum und lineare Funktion	4	4	100
101	Fact1107b	bestimmen/ermitteln: Bedeutung der Parameter beim exponentiellen Wachstum	9	4	44
101	Fact1107e	entscheiden: Bedeutung der Parameter beim exponentiellen Wachstum	15	13	87
102	Fact1103b	bestimmen/ermitteln: Exponentielles Wachstum	2	2	100
102	Fact1103e	entscheiden: Exponentielles Wachstum	10	8	80

Abbildung 61 Ergebnis derselben Analyse auf Basis materialisierter Daten

7.14.3 Analyse für Facts

Die Analyse der Bewertung kann von den Skills auf die Grundfakten übertragen werden, indem über alle benötigten Skills eines Fakts aggregiert wird, mit oder ohne Gewichtung.

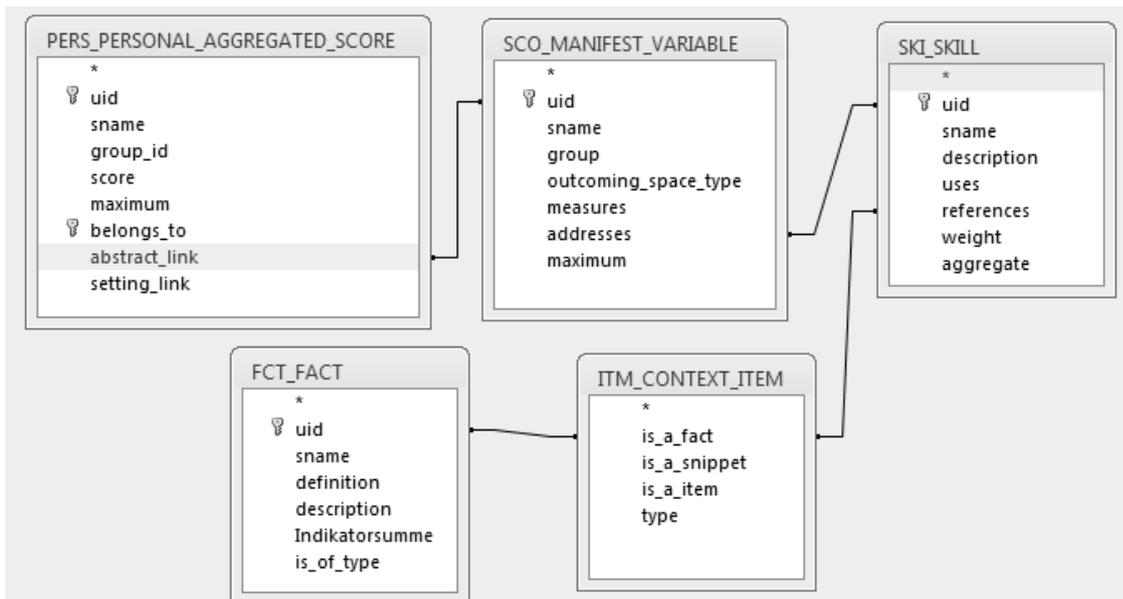


Abbildung 62 Analyse pro Person und Fakt

So zeigt Abbildung 62 die Struktur einer Auswertung mit dem entsprechenden Ergebnis in Abbildung 63.

Person	Fakt	Kurzbeschreibung	Score	Maximum
100	1103	Exponentielles Wachstum	9	12
100	1104	Lineares Wachstum	3	3
100	1105	Wachstum und Exponentialfunktion	6	9
100	1106	Wachstum und lineare Funktion	4	4
100	1107	Bedeutung der Parameter beim exponentiellen Wachstum	19	24
101	1103	Exponentielles Wachstum	10	12
101	1104	Lineares Wachstum	3	3
101	1105	Wachstum und Exponentialfunktion	7	9
101	1106	Wachstum und lineare Funktion	4	4
101	1107	Bedeutung der Parameter beim exponentiellen Wachstum	17	24
102	1103	Exponentielles Wachstum	10	12
102	1104	Lineares Wachstum	3	3

Abbildung 63 Ergebnis der Analyse pro Fakt

Die jeweiligen Ergebnisse pro **PERSON** und **FACT** können ebenfalls in **PERSONAL AGGREGATED SCORE** gespeichert werden. Dies ist mit **PRO_PERS_SCORE_FACT_PERSON** für die Fakten erfolgt. Weitere Aggregationen können analog durchgeführt werden.

SCORE MODEL (SCO)			
AGGREGATE	(10,12)		Manuell
MANIFEST_VARIABLE	(5,9,10,12)		Aus Items
VARIABLE_AGGREGATION_RULE	(10,12)		Manuell
PERSONAL MODEL (PERS)			
PERSONAL_AGGREGATED_SCORE	(10)		Datenbank
PERSONAL_MANIFEST_SCORE	(10)		Datenbank

Abbildung 64 Ziel des Schritts 10 im Informationsmodell

7.15 Schritt 11 – Statistical Analysis

Im Anschluss an die manifeste Analyse mit dem klassischen Scoring kann eine zusätzliche statistische Analyse erfolgen. Dazu wird das RASCH-Modell als ein Modell der Item Response Theory (IRT) bereits in Abschnitt 4.2.5 beschrieben, hier vorgeschlagen und beispielhaft verwendet.

Der Modelltyp RASCH-Modell wird zusammen mit einer kurzen Beschreibung im **STATISTICAL MODEL TYPE** hinterlegt. Die verwendeten Testverfahren, hier der grafische Test, der LR-Test nach Andersen und der Wald-Test werden im **STATISTICAL QUALITY TEST** hinterlegt und mit dem Modelltyp verbunden. Für die Parameter der Tests werden jeweils eigene Tabellen angelegt.

Die Analyse nach dem RASCH-Modell läuft iterativ ab. Die Basis bilden immer **STATISTICAL VARIABLES**, die ihrerseits auf **MANIFEST VARIABLES** beruhen und zusätzlich statistische Informationen beinhalten. Die direkt beobachtbaren **MANIFEST VARIABLES** und die sie repräsentierenden **STATISTICAL VARIABLES** bilden die Basis für die Schätzung der **LATENT VARIABLES**. Grundsätzlich wird pro Assessment-Teilnehmer und **LATENT VARIABLE** eine Kompetenzschätzung vorgenommen. Jede Analyse mit einem statistischen Modell betrifft also eine **LATENT VARIABLE**. Dabei werden zunächst alle für eine latente Variable inhaltlich relevanten Items und die zugehörigen manifesten Variablen ausgewählt. Eine latente Variable beruht wie geschildert zumeist auf einer Gruppe von manifesten Variablen, eventuell sogar auf dem kompletten Assessment. Beschreibt die Gruppe nur alle Items, die eine **COMPETENCE** betreffen, wird die entsprechende Kompetenz mit der latenten Variablen geschätzt. Dies sind dann genau die

MANIFEST VARIABLES, die im CSP entweder über das **ASSESSMENT ITEM**, **ABSTRACT ITEM** und **SNIPPET** mit dem **FACT** und schließlich der **COMPETENCE** verbunden sind (cause) oder die über den **SKILL**, **SKILL_INDICATION** und das **SPECIFIC COMPETENCE NIVEAU** damit verbunden sind (INFER).

Für die ausgewählten **MANIFEST VARIABLES** müssen die entsprechenden **STATISTICAL VARIABLES** angelegt werden. Die meisten statistischen Modelle oder Tests zu deren Validierung beinhalten auch Aussagen über diese einzelnen Variablen. Zur Entkopplung des statistischen Modells vom Score-Modell ist es daher sinnvoll, zusätzlich eine statistische Variable für jede manifeste Variable, die verwendet wird anzulegen, um die zugehörigen statistischen Informationen beschreiben zu können.

Die Bewertungen, also Scores der ausgewählten **MANIFEST VARIABLES**, müssen dahingehend überprüft werden, ob sie den Voraussetzungen des statistischen Modells entsprechen. Wird das RASCH-Modell gewählt sind nur dichotome Bewertungen zulässig. Da die Bewertungen im **PERSONAL ITEM SCORE** hinterlegt sind, dürfen dort dementsprechend nur 0 und 1-Werte auftreten. Sollte dies der manifesten Analyse widersprechen, können über eigene **MANIFEST VARIABLES** für dieselben Items und deren Bewertung mit einer eigenen dichotomen **MEASUREMENT SCALE** im Modellbereich SCORE die entsprechenden Bewertungen erzeugt und zusätzlich im **PERSONAL ITEM SCORE** hinterlegt werden.

Anschließend wird der erste Zyklus der Analyse des Modells gestartet. Das Ergebnis sind Parameterschätzungen und Testergebnisse aus den Validierungstests. Dabei können bereits statistische Variablen entfallen, die Items entsprechen, die beispielsweise von allen oder keinem Assessment-Teilnehmer gelöst werden konnten. Außer über die modellspezifischen Parameter können auch über die Tests auch Items isoliert werden, die den Modellvoraussetzungen nicht entsprechen. Dies können im grafischen Test Items sein deren Signifikanzregion außerhalb der Mittelhalbierenden liegt oder im Wald-Test Items deren p-Werte zu klein sind.

So entsteht eine verkleinerte Menge von statistischen Variablen, die ein neues Modell bilden. Mit diesen wird ein neues RASCH-Modell erstellt und das Ergebnis wie oben behandelt. Die Modelle können als Vorgänger-Nachfolger-Modell im **STATISTICAL MODEL** miteinander verbunden werden. Für den neuen Durchlauf wird eine neue **LATENT VARIABLE** angelegt, die mit dem neuen Modell verbunden wird. So können alle Einzelschritte protokolliert werden. Ob alle Einzelschätzungen und Testergebnisse tatsächlich protokolliert werden, muss im Einzelfall entschieden werden. Dieses Verfahren kann wiederholt werden, bis alle Parameter den gewünschten Signifikanzniveaus entsprechen.

Ist dieser Zustand erreicht, können die ermittelten Kompetenzwerte θ_i und Schwierigkeitswerte β_j im abschließenden Modell dokumentiert werden. Die Kompetenzen werden je Person für die **SETTING EXECUTION** in **PERSONAL TRAIT SCORE** eingetragen. Dort wird eine Verbindung zur **LATENT VARIABLE** des abschließenden Modells hergestellt. Die Schwierigkeitswerte werden in den statistischen Variablen eingetragen, die auf Ebene der Einzelitems angelegt worden sind.

Zusätzlich können statistische Parameter, die sich entweder auf das Gesamtmodell oder einzelne latente Variable beziehen in den **STATISTICAL QUALITY PARAMETERS** dokumentiert werden.

STATISTICS MODEL (STA)			
STATISTICAL_MODEL	(5,11,12)		Manuell
STATISTICAL_QUALITY_PARAMETERS	(11,12)	LR Wald Graphic	Aus R Aus R Aus R
LATENT_VARIABLE	(11,12)		Verknüpfung über Datenbank, Werte über R-Import
STATISTICAL_VARIABLE	(11,12)		Manuell/ R
AGGREGATE	(11,12)		Manuell, Import
PERSONAL MODEL (PERS)			
PERSONAL_TRAIT_SCORE	(11,12)		Import aus R

Tabelle 19 Ziel des Schritts 11 im Informationsmodell

7.16 Schritt 12 – Competence Estimation

In diesem Schritt geht es um die detaillierte Kompetenzschätzung. Es werden einzelne Aussagen zu Kompetenzen aus dem Assessment abgeleitet, die dazu dienen sollen, die Aussagen des manifesten Scorings zu bestätigen und in Form eines Traits zu verallgemeinern.

Es geht also darum Aussagen über die im CSP-Modell modellierten Kompetenzen zu treffen und jedem Assessment-Teilnehmer ein Kompetenzniveau hinsichtlich dieser Kompetenzen zuzuordnen. Im CSP-Modell ist die Definition von Kompetenzniveaus zunächst eine Modellierungsentscheidung. In Schritt 2 sind die **SPECIFIC COMPETENCE NIVEAUS** festgelegt worden. In Schritt 4 sind die **SKILLS** definiert und deren Nutzung zur Bestimmung der Kompetenzniveaus definiert und in den **COMPETENCE NIVEAU METRICs** algorithmisch hinterlegt worden. Auch die Schwellenwerte für die Erreichung der einzelnen Kompetenzstufen sind definiert worden.

Nachdem jetzt die Scores je Fähigkeit und Assessment-Teilnehmer bekannt sind, können diese vorbereiteten Strukturen zur Ermittlung der Kompetenzniveaus verwendet werden. Dazu müssen gegebenenfalls die in den **COMPETENCE NIVEAU METRICs** definierten Algorithmen ausgestaltet und für die ausgewählten Items konkretisiert werden. Damit lässt sich für jede Fähigkeit und jede Person eine Aussage über deren Kompetenzniveau hinsichtlich einer bestimmten Kompetenz bestimmen. Es lassen sich zudem statistische Parameter wie Mittelwerte und Standardabweichung für die Fähigkeiten und Analysen über die Abweichung von Assessment-Teilnehmern von diesen Werten ermitteln. Im Ergebnis entsteht somit über die statistischen Parameter ein Feedback für die Lehrperson und über die Individualanalyse eines für jeden Assessment-Teilnehmer.

Gruppen von Items werden durch manifeste Variablen dokumentiert. Die Gruppierung erfolgt über die Aggregation in **SCO_AGGREGATE** zu einer neuen manifesten Variablen. In Schritt 10 (und Schritt 11) sind bereits derartige Gruppen zu neuen manifesten Variablen zusammengefasst worden. Diese Variablen sind jeweils einem **SKILL** oder einem **FACT** (allgemeiner **CONTENT ELEMENT**) zugeordnet, für die das Scoring der Gruppe genutzt werden kann. Das Scoring ist zumeist schon erfolgt und insbesondere sind die aggregierten Scores für die Gruppen in **PERSONAL AGGREGATED SCORE** hinterlegt worden. So lassen sich für die einzelnen Personen Aussagen über deren Score hinsichtlich des zugrunde gelegten **SKILLS** und **CONTENT ELEMENTs** machen.

Die Grundidee ist jetzt, latente Variablen von diesen gruppierten manifesten Variablen abzuleiten, die wiederum mittels des RASCH-Modells genutzt werden, um Traits für diese Variablen zu bestimmen, die bestimmten **SKILLS** oder **CONTEXT ELEMENTs** entsprechen. Damit werden diese Gruppen einer statistischen Analyse unterzogen, wobei auf die Vorgehensweise im Schritt 11 (Statistical Analysis) zurückgegriffen werden kann. Es können dieselben Modelle und Regeln auf die neuen latenten Variablen sukzessive angewendet werden. Die ermittelten statistischen Ergebnisse dienen der Verifizierung der Aussagekraft des Scores der ausgewählten Items. Lässt sich aus der statistischen Analyse eine sinnvolle Kompetenzschätzung ableiten, ist dies ein Indikator für die Reliabilität der Scores der ausgewählten Items. Die Ergebnisse werden wie im Schritt 11 dokumentiert.

Korreliert man also beispielsweise die Traits der latenten Variablen und die Scores hinsichtlich eines bestimmten Skills und es ergeben sich ausreichende Zusammenhänge, so erhält man eine weitere Kontrolle hinsichtlich des Scorings. Diese unterstützt eine Verwendung der Scores für eine Kompetenzschätzung gemäß der Modelldokumentation in den CNMs.

STATISTICS MODEL (STA)			
STATISTICAL_MODEL	(5,11,12)		Manuell
STATISTICAL_QUALITY_PARAMETERS	(11,12)	LR	Aus R
	LR, Wald	Wald	Aus R
	Graphic	Graphic	Aus R
LATENT_VARIABLE	(11,12)		Verknüpfung über Datenbank, Werte über R-Import
AGGREGATE	(11,12)		Manuell, Import
SCORE MODEL (SCO)			
AGGREGATE	(10,12)		Manuell
MANIFEST_VARIABLE	(5,9,10,12)		Aus Items
VARIABLE_AGGREGATION_RULE	(10,12)		Manuell
PERSONAL MODEL (PERS)			
PERSONAL_TRAIT_SCORE	(11,12)		Import aus R
PERSONAL_AGGREGATED_SCORE	(10,12)		Datenbank
PERSONAL_MANIFEST_SCORE	(10,12)		Datenbank

Tabelle 20 Ziel des Schritts 12 im Informationsmodell

7.17 Schritt 13 – Personal Feedback

Das Ziel dieses Schrittes ist die Bereitstellung von Ursachen- und Fehleranalysen und möglichen Maßnahmen zur Behebung von Defiziten. Das Feedback soll sich einerseits an die Assessment-Teilnehmer, andererseits an die Lehrpersonen richten, die für den Bildungsprozess verantwortlich sind.

Die Analyse der Defizite kann prinzipiell auf mindestens vier Ebenen erfolgen, wenn man die Anknüpfung an die anderen Teile des CSP-Modells betrachtet. Dies sind die Ebenen der **SPECIFIC COMPETENCE NIVEAUS**, der **CONTEXT ELEMENTs**, der **SKILLS** und der **RESPONSE TYPEs**.

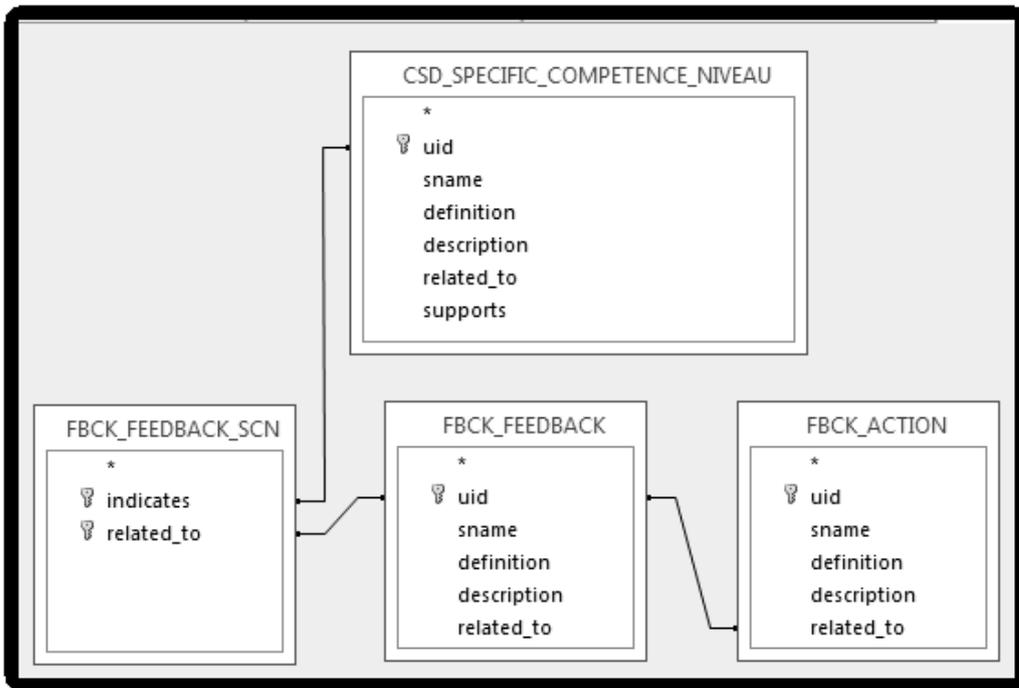


Abbildung 65 Zusammenhang zwischen Specific Competence Niveau und Feedback

In Abbildung 65 ist der grundsätzliche Zusammenhang zwischen dem ermittelten **SPECIFIC COMPETENCE NIVEAU** und dem **FEEDBACK** zu sehen (FBCK_SCN_ACTION). Für den Aufbau der Zusammenhänge können die bestehenden Verbindungen zwischen der **COMPETENCE**, den **FACTs** und **SNIPPETs** bis zu den Items des Assessments genutzt werden, um die relevanten Items zu erkennen und aus deren inhaltlicher Konstruktion auf das richtige Feedback zu schließen. Ist so für einen Assessment-Teilnehmer dessen Kompetenzniveau hinsichtlich einer bestimmten Kompetenz bekannt, kann das entsprechende Feedback ermittelt und dokumentiert werden und es kann auf vorhandene Materialien verwiesen werden.

Die nächste Ebene stellen die **CONTEXT ELEMENTs** dar, die ebenfalls direkt mit den entsprechenden im **FEEDBACK** dokumentierten Ursachen verbunden werden können. In Abbildung 66 ist beispielhaft eine Analyse mittels der Abfrage QRY_FBCK_CONTEXT_ELEMENT zu sehen, die diese für Fakten realisiert. Ein mögliches Ergebnis ist in Abbildung 93 zu sehen.

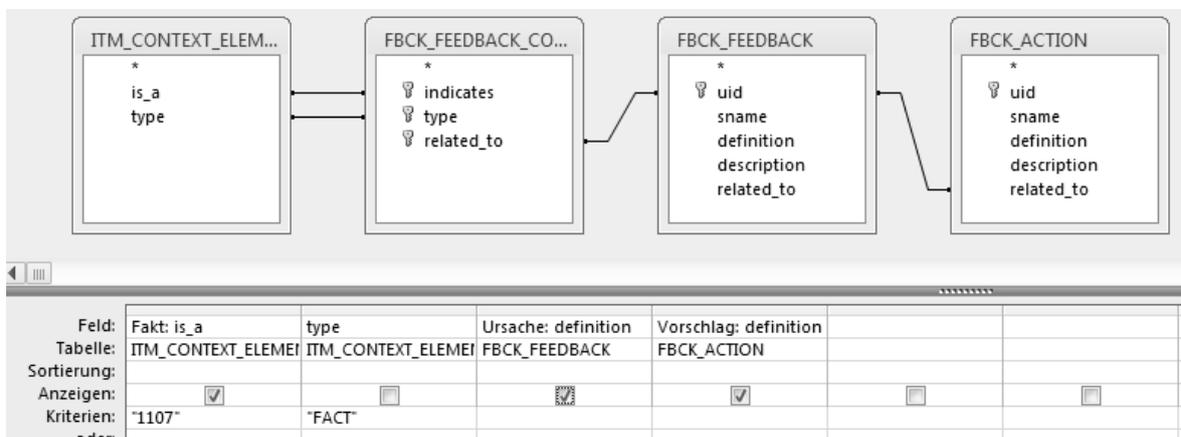


Abbildung 66 Feedback-Analyse auf Faktenebene

Damit lassen sich für die einzelnen Fakten mögliche Fehlerursachen darstellen. Ist für eine Person bekannt, dass sie besondere Defizite hinsichtlich eines Fakts hat, wie sie die vorausgegangenen Analysen aufgezeigt haben, so können hier die möglichen Ursachen direkt angegeben und durch die weitere Nutzung des Feedback-Modells bis hin zu den Aktionen und eventuellen Materialien unmittelbare Hinweise zu deren Behebung gegeben werden.

Die Analyse auf Ebene der **SKILLS** läuft prinzipiell gleichartig ab. Die **SKILLS** stellen im CSP eine Verfeinerung der **CONTEXT ELEMENTs** dar, indem die Operatoren hinzugefügt und so die Fähigkeiten modelliert werden. Für die hier durchzuführende Analyse bedeutet dies lediglich, dass das **CONTEXT ELEMENT** durch die entsprechenden einzelnen **SKILLS** ersetzt wird. Strukturell ändert sich nichts.

Die vierte Ebene ist wesentlich detaillierter, wird hier doch auf die Fehleranalyse aus Schritt 8 direkt zugegriffen. So kann jetzt für jede Antwort eines einzelnen Items punktuell eine Beschreibung des Fehlertyps, eine Ursachenanalyse und ein Vorschlag für mögliche Maßnahmen ermittelt werden.

Die entsprechende Abfragestruktur ist in Abbildung 67 zu sehen. Für ein mögliches Ergebnis sei auf Abbildung 93 verwiesen. Beispielhaft wurde der **RESPONSE TYPE** 110 gewählt (QRY_FCBK_RESPONSE_ERROR).

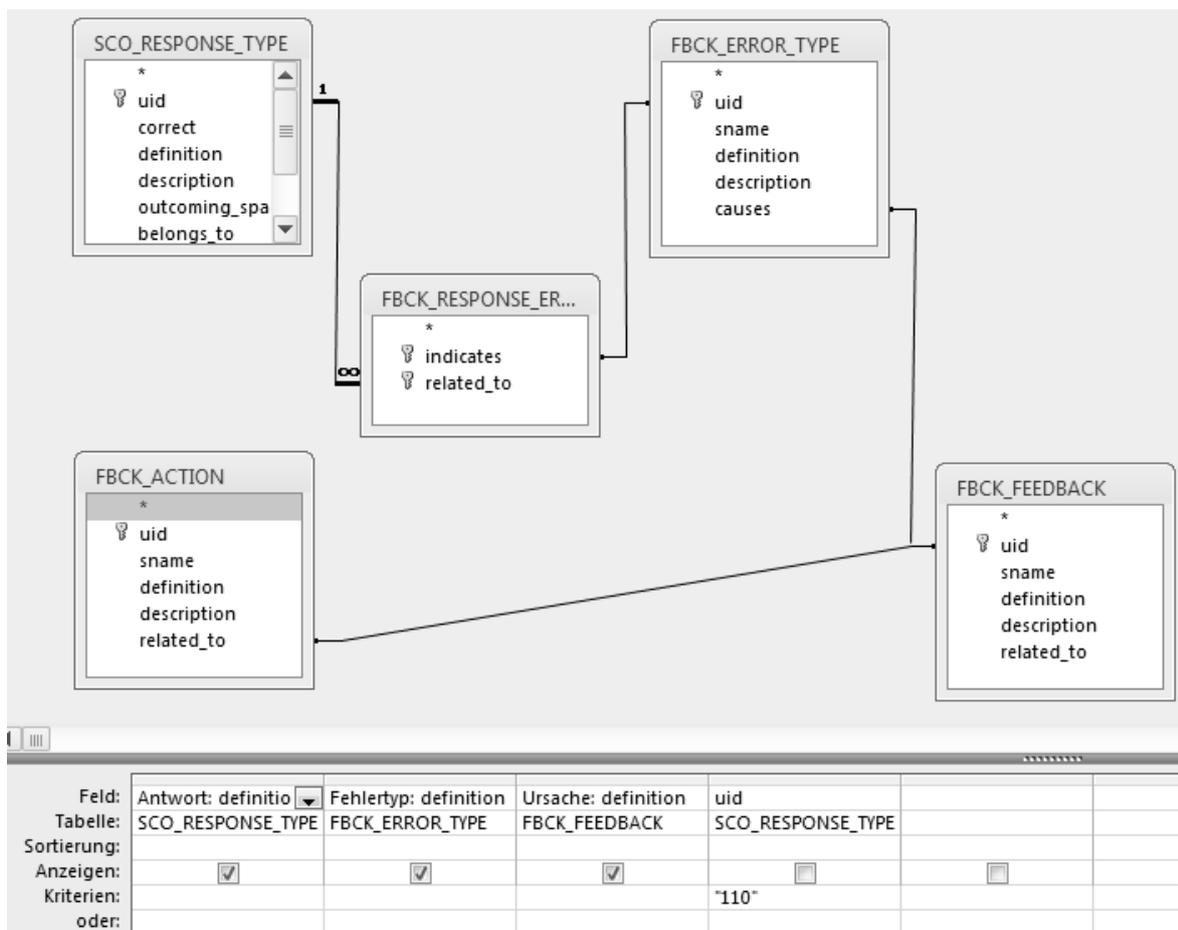


Abbildung 67 Detaillierte Feedback-Analyse für einen RESPONSE TYPE

Erweitert man die Struktur um die entsprechenden Ergebnisse der einzelnen Personen wie in Abbildung 68 dargestellt, lassen sich alle Ursachen und Vorschläge für einen Assessment-Teilnehmer auf Grund der Antworten in den einzelnen Items ermitteln (QRY_FCBK_RESPONSE_PERSON). Ein Ergebnis ist in Tabelle 39 aufbereitet.

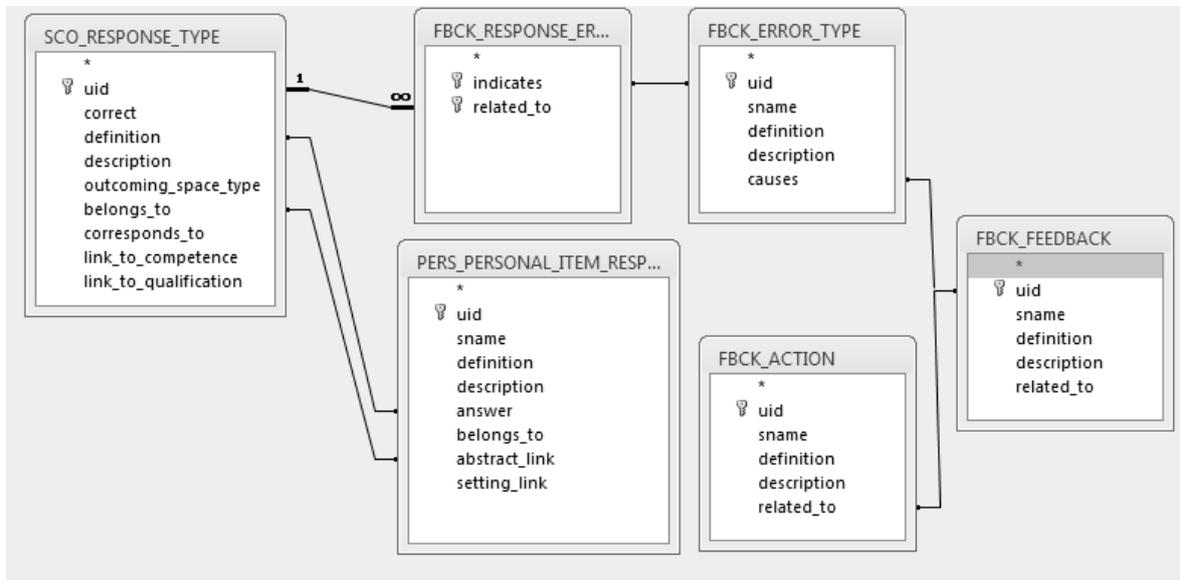


Abbildung 68 Struktur für die Ermittlung des Feedbacks für einen Assessment-Teilnehmer

Damit ergibt sich ein detailliertes Feedback für jeden einzelnen Assessment-Teilnehmer. Die Detaillierung lässt sich nach Fehlertypen, nach Fehlerursachen (**FEEDBACK**), nach Vorschlägen und Maßnahmen (**ACTION**) bis hin zu konkretem Material (**FEEDBACK MATERIAL**) variieren. Die Möglichkeiten hängen von dem Aufwand ab, der im Rahmen des Schritts 8 investiert worden ist. Außerdem ist mit jeder Wiederverwendung von Items hier eine bessere Basis vorhanden. Allerdings sollte dies ergänzend gesehen werden, nicht als Möglichkeit einer vollständigen Analyse.

Bei dem Aspekt des Feedbacks für die Lehrperson stehen die Fragen im Vordergrund, welche Fähigkeiten und welche Defizite viele Assessment-Teilnehmer aufweisen und bei welchen es sich eher um individuelle Einzelfehler handelt. Dies stellt eine wichtige Steuerungsmöglichkeit für den Bildungsprozess dar.

Eine Analyse eines einzelnen Items kann bereits Aufschluss über die aufgetretenen Defizite erbringen. Die Struktur für eine solche Analyse ist in Abbildung 69 dargestellt, ein mögliches Ergebnis in Tabelle 40 (QRY_FBCK_ERROR_TYPE_COUNT).

Diese Übersichten lassen sich jetzt auf größere Bereiche nach Bedarf übertragen. Ein weiteres Beispiel sei die in Abbildung 70 angegebene Auswertung (QRY_FBCK_FEEDBACK_COUNT). Hier wird eine Übersicht über alle Fehlerursachen erstellt, die sich bei den Assessment-Teilnehmern ergeben haben. Wiederum findet sich ein beispielhaftes Ergebnis als Ausschnitt in Abbildung 97.

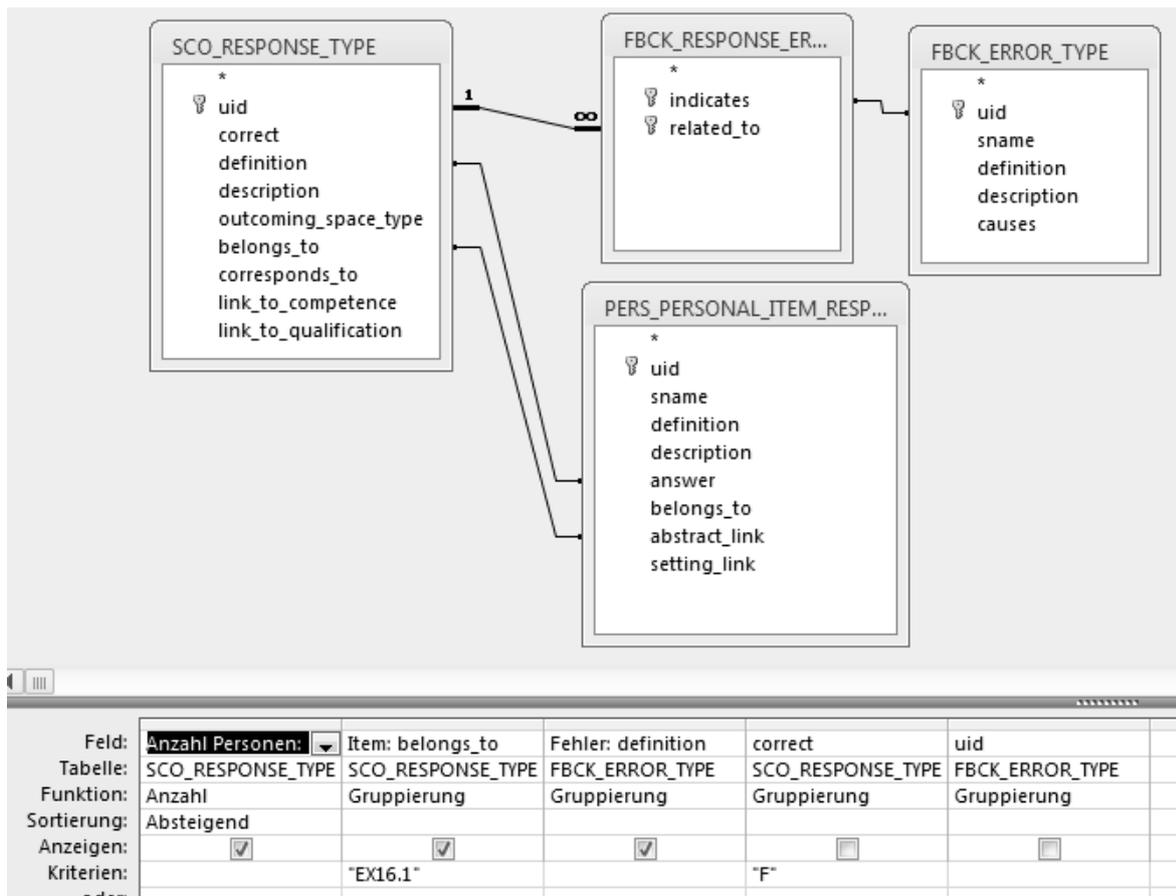


Abbildung 69 Fehleranalyse für ein einzelnes Item

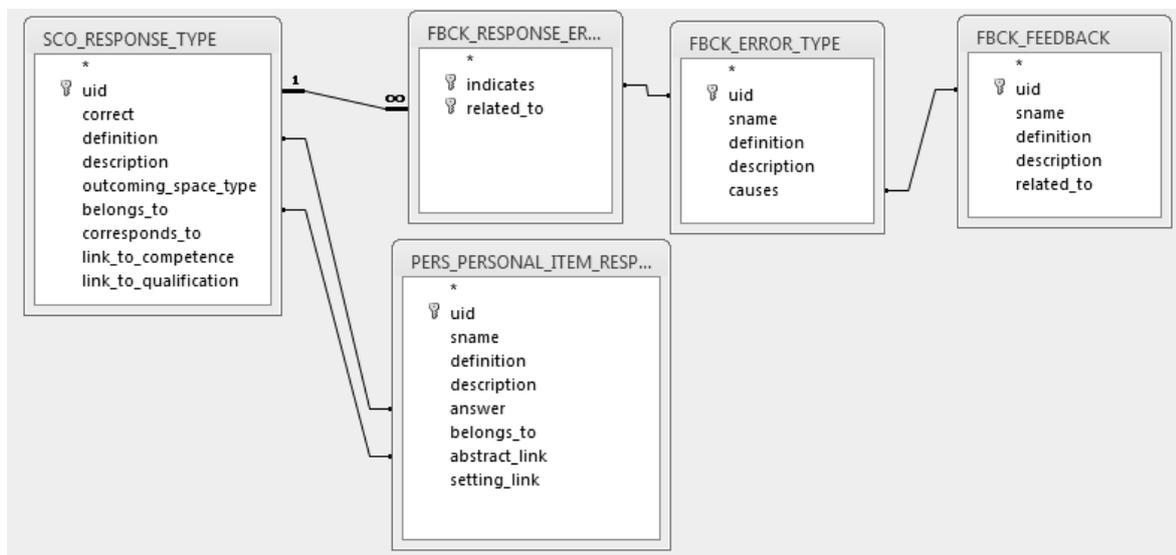


Abbildung 70 Übersicht über die aufgetretenen Fehlerursachen (Feedback)

In diesem Schritt werden keine Ergebnisse unmittelbar im Informationsmodell abgelegt. Das Ergebnis dieses Schrittes besteht aus den einzelnen Analysen. Weitere Details sind im Anhang beschrieben.

7.18 Schritt 14 – Integration

Das Vorgehensmodell kann in allen Schritten, oder auch nur in Teilbereichen durchlaufen werden. Ist beispielsweise der Domäneninhalt bereits definiert, kann die Entwicklung eines neuen Assessments genügen. Ist auch das Assessment bereits vorhanden und soll es wiederverwendet werden, reicht die Dokumentation der Ausführung und der Ergebnisse.

In jedem Fall scheint es sinnvoll zu sein, nach einem Durchlauf oder auch Teildurchlauf zu überprüfen, ob Teile der Dokumentation für spätere Schritte gezielt gesichert werden sollen. Infrage kommen hierfür insbesondere die im Assessment entwickelten Items und deren Gruppierungen. So kann jede Ebene vom **ASSESSMENT ITEM** über das **ASSESSMENT ITEMSET** bis zur **EXERCISE** zu einem **ASSESSMENT ELEMENT** abstrahiert und einem **ASSESSMENT POOL** zugeordnet werden. Das dafür benötigte **MATERIAL** bleibt entsprechend zugeordnet und kann gegebenenfalls verallgemeinert werden.

Daneben bietet es sich an, die aus den Erfahrungen bei der Durchführung gewonnenen Erkenntnisse in den entsprechenden Typen **EXECUTION PROBLEM** und **RAW ERROR TYPE** für weitere Durchführungen von Assessments zu dokumentieren.

ASSESSMENT MODEL (ASS)			
EXECUTION_PROBLEM	(0,5,14)		Manuell
RAW_ERROR_TYPE	(0,5,14)		Manuell
MATERIAL	(5,14)		Manuell
ASSESSMENT_ELEMENT	(5,14)		Manuell
ASSESSMENT_POOL	(14)		Manuell

Tabelle 21 Ziel des Schritts 14 im Informationsmodell

8 Anwendung und Ergebnisse

In diesem Abschnitt geht es um die testweise Umsetzung des CSP-Informationsmodells und der dazugehörigen Vorgehensweise. Der erste Schritt zur Umsetzung des Modells in eine praktische Anwendung ist die Implementierung des Modells in einem lauffähigen Datenhaltungssystem. Hierfür bieten sich aufgrund der weiten Verbreitung in der Praxis relationale Systeme an. Wegen der einfachen Handhabung wird zunächst Access gewählt. Das Modell ist in der Datenbank CSP.accdb hinterlegt.

8.1 Theoretische Ableitung und Rahmenbedingungen

8.1.1 Domäneninhalt

Die erste umfangreichere Validierung erfolgte mittels eines an einem Beruflichen Gymnasium Wirtschaft durchgeführten e-Assessments. Es werden vier Lerngruppen der Klassenstufe 11 (Einführungsphase) gegen Ende des Schuljahres in einem Assessment im Fach Mathematik getestet. Das im Rahmen des normalen Unterrichts durchgeführte und bewertete Assessment wurde in anonymisierter Form zusätzlich hier im CSP-Modell dokumentiert und analysiert. Das Assessment ist im Anhang beigefügt.

Zunächst wurde der Inhalt des Kerncurriculums Mathematik für die gymnasiale Oberstufe in Niedersachsen (Kultusministerium 2009) auszugsweise in das CSP umgesetzt. Es wurde als externe Spezifikation dokumentiert und die dort für die Klassenstufe 11 hinterlegten Kompetenzbeschreibungen (Kultusministerium 2009, 12f.) werden im Modellbereich Competence Design hinterlegt. Dabei geht es um prozessbezogene und inhaltliche Kompetenzen. Die Beispiele im Abschnitt 7.6 zum Competence Modeling sind dieser Umsetzung entnommen.

Es wird der Schwerpunkt im Bereich der linearen und exponentiellen Wachstumsprozesse gesetzt, die zum Testzeitpunkt auch im Mittelpunkt des Unterrichtsprozesses gestanden haben. Diese Kompetenzen werden zu **FACTs**, **SNIPPETs** und **ITEMs** verfeinert.

In Abbildung 27 ist ein relevanter Ausschnitt graphisch modelliert. Die Umsetzung in die ABSTRACT ITEMs ist in Abbildung 44 beispielhaft dargestellt.

Die so gewonnenen Anforderungen sind anschließend mit einem Assessment verbunden worden. Für die Erstellung und die Durchführung des kompletten Assessments wurde die Software SoSciSurvey (phpBB 2012) genutzt. Die Daten wurden auf dem Server von SoSciSurvey gespeichert und anschließend im CSV-Format heruntergeladen. Die so gewonnenen Daten sind hier übernommen. Dies sind die Rohdaten (in Excel Original20130611.xlsx). Die Nummerierung (CASE) stammt aus dem Download und wird von SoSciSurvey vergeben. Lücken ergeben sich aus Fehlern in der Anmeldung, lehrerseitige Testanmeldungen und zwei PC-Abstürze.

Die vier Assessmentgruppen waren vier Gruppen, die hier als A (23 SuS), die B (18 SuS), die C (19 SuS) und die D (19 SuS) bezeichnet werden. Die vier Gruppen sind von unterschiedlichen Lehrkräften unterrichtet worden. Von allen Schülerinnen und Schülern waren 76 anwesend. Die Dauer des Assessments betrug 75 Minuten. Die Schülerinnen und Schüler waren durchgehend unter Aufsicht. Jeder Schüler hatte einen eigenen Computer.

Ein Referenzdatensatz mit den entsprechenden Lösungen zum späteren Abgleich (Referenz.xlsx) ist erstellt worden.

8.1.2 Korrektur der Durchführungsprobleme(Probleme)

Am gravierendsten waren zwei Computer-Abstürze, einer vor, einer während der Durchführung. Dabei erfolgten jeweils eine neue Anmeldung und eine entsprechende zeitliche Verlängerung der Bearbeitungsdauer. Als Korrektur für den PC während der Durchführung wurden in der Originaldatei zwei Datensätze zusammengeführt, die als Folge des PC-Absturzes in zwei Sessions erstellt wurden. (Bereinigt3FehlendeEingaben20130611.xlsx). Dabei wurden, wenn im zweiten Datensatz eine Bearbeitung stattgefunden hatte, deren Ergebnisse in den ersten Datensatz übernommen. Außerdem wurden für die Items PS06_01 bis PS06_05 alle Nichteingaben -9 (der Schieberegler wurde auf seiner mittleren Position belassen) durch 50,5 als Mittelposition ersetzt. Diese Items dienen dem Feedback der SuS, wie das e-Assessment empfunden wurde.

Die Items zur Funktionssynthese sind aus der Analyse herausgenommen worden, da sich kurz vor dem Test zeigte, dass diese in der verfügbaren Zeit nicht zu bewältigen waren und zudem nicht in allen Klassen im Unterricht bis zum Zeitpunkt des Assessments behandelt worden waren.

8.1.3 Korrektur der RAW ERROR TYPES

Die weiteren Analysen sind außer in der Datenbank auch in M11V2.xlsx dokumentiert. Das Datenblatt „Daten“ stellt den Einstieg mit den bisher überarbeiteten Daten dar. Ausgewertet werden nur die Spalten Q bis GB. Die Spalten davor und danach beinhalten administrative Daten und Metadaten.

Das Datenblatt „Indikator exakt“ beschreibt einen ersten Richtig/Falsch-Indikator bezogen auf den Referenzdatensatz. Eine „1“ steht immer für eine richtige, eine „0“ für eine falsche Lösung. Dasselbe gilt für alle weiteren Indikatorarbeitsblätter. In einem ersten Bearbeitungsschritt wurden offensichtliche Bedienungsfehler und andere elementare Fehler bearbeitet, die nicht Teil der im Assessment zu überprüfenden Kompetenzen sind.

Die Daten wurden in die entsprechende Tabelle des CSP übernommen, also in **PERSONAL ITEM RESPONSE**. Die einzelnen Schritte im Rahmen der Korrektur der RAW ERRORS wurden zur detaillierten Nachvollziehbarkeit und zur Überprüfung des Umfangs der Auswirkungen für diese erste Analyse in Excel getrennt dokumentiert.

Der erste Bearbeitungsschritt ist das Runden der Daten. Hier geht es um die Korrektur von offensichtlichen Rundungsfehlern, beispielsweise 0,98 statt 0,97. Außerdem sind zu viele oder zu wenige Nachkommastellen korrigiert worden. Die Entscheidung war, Rundungsfehler nicht als fehlerhaft einzustufen, wenn es sich um eine Abweichung von +/-1 in der letzten Stelle handelt. Das Blatt „Indikator gerundet“ erstellt wiederum den Richtig/Falsch-Indikator auf den so korrigierten Daten. Im nächsten Schritt sind erkennbare Fehleingaben, die auf Tippfehler, falsches Verständnis des Antwortformates oder offensichtlich falsche Bedienung zurückzuführen sind, korrigiert worden.

Typische Beispiele hierfür sind:

- falsches Dezimaltrennzeichen,
- zusätzliche Zeichen führen zu Textfeld statt numerischer Angabe,
- O (Großbuchstabe) und 0 (Null) verwechselt,
- Exponentialschreibweise statt Dezimalschreibweise,
- Leerzeichen als Zifferntrennung,
- + 1 statt 1,
- Komma vergessen (beispielsweise 1005 statt 1,005),
- Funktionen falsch eingegeben.

In einem letzten Schritt sind die Items bearbeitet worden, die keine eindeutige Eingabe bieten. Dies sind insbesondere die Abfragen zu den Prozessschritten zur Berechnung lokaler Extrema und von Wendepunkten. Hier sind alle Antworten auf eine richtige Reihenfolge geprüft und bis zum ersten falschen Schritt als korrekt gewertet worden. Abweichende Reihenfolgen sind dabei zulässig. Bei der späteren Indikatorvergabe wird die Länge der richtigen Antwort als Anzahl der Punkte vergeben. Das Blatt „Indikatorgerundetfehleingabe“ erstellt wiederum den Richtig/Falsch-Indikator auf den so korrigierten Daten. Das entspricht dem Inhalt von **PERSONAL ITEM RESPONSE SCORE**.

Insgesamt zeigen sich in den Indikatoren geringe Unterschiede. Die Auswirkungen dieser Korrekturen sind also gering.

8.1.4 Analyse der Antworten (Response Modeling)

Anschließend wurden die Antworten der einzelnen Items unter dem Gesichtspunkt analysiert, ob sie bereits in den **RESPONSE TYPEs** als „predefined“ dokumentiert waren. Das Ergebnis dieser Schritte ist die Ergänzung um die nicht vorher berücksichtigten, tatsächlich aber aufgetretenen Antworten. Diese Antworten wurden im **PERSONAL ITEM RESPONSE TYPE** als „ex-post“ dokumentiert. Naturgemäß ist hier die Anzahl nachzutragender Antworten umso größer, je offener die Fragestellung ist. Bei reinen Ja/Nein-Fragen ist keine Nachbearbeitung notwendig, während die Eintragung eines zu berechnenden Wertes deutlich mehr überraschende Antworten zulässt.

Bei diesen auftretenden Antworten sind beispielhaft einige Fehlertypen analysiert worden, die eine bestimmte Antwort nahelegen. Dies ist beispielsweise im Rahmen der Prozentrechnung geschehen, wo bei einer erwarteten Antwort von „1,04“ als Multiplikator basierend auf 4% Zuwachs, eine Antwort „0,04“ auf das Vergessen des Bestands hinweist. Antworten wie „-0,04“ oder „0,96“ deuten auf eine Verwechslung von Zuwachs und Abnahme hin. Die analysierten Fehlertypen sind im **ERROR TYPE** dokumentiert worden. Sie wurden auf mögliche Ursachen analysiert, die anschließend mit **FEEDBACK** verbunden worden sind. Die weitere Dokumentation möglicher Maßnahmen und Vorschläge in **ACTIONs** ist nur beispielhaft beschrieben. Auf ergänzende Arbeitsmaterialien wurde verzichtet.

8.2 Ergebnisse im Vergleich zu klassischen Klausuren

8.2.1 Überblick

Eine erste Analyse soll einen Eindruck vom Ergebnis des e-Assessments mit dem CSP-Modell im Vergleich zu einer parallel geschriebenen Klausur geben. Diese Klausur ist eine 80-minütige paper-and-pencil-Klausur. Das Ziel dieses Vergleichs ist eine Absicherung der Ergebnisse insofern, als dass ähnliche Ergebnisse wie bei konventionellen Assessments erzielt werden.

Ein paar konkrete Probleme sollen verbalisiert werden:

- Alle Klassen sind von verschiedenen Lehrkräften unterrichtet worden. Krankheitsbedingt waren auch Lehrerwechsel erforderlich. Dadurch waren verschiedene Unterrichtsprozesse erforderlich und alle Klassen mussten voneinander variierende Klausuren zu unterschiedlichen Zeitpunkten schreiben. Dies entspricht einerseits durchaus in der Praxis immer wieder anzutreffenden Situationen, erschwert aber andererseits die Vergleichbarkeit.

- Durch unterschiedliche Prozesse sind auch leicht variierende inhaltliche Fähigkeiten im Unterricht thematisiert worden. Die Klausuren haben also verschiedene Domänenaspekte abgedeckt.
- So waren auch in den verschiedenen Lerngruppen inhaltliche Themen unterschiedlich stark ausgeprägt. Auch die klassischen Assessments in den vier Klassen wurden jeweils daran ausgerichtet. Die Inhalte umfassten in unterschiedlicher Gewichtung:
 - a. Betriebsoptimum/Betriebsminimum,
 - b. Anwendungen von Extrem- und Wendepunkten,
 - c. Exponentialfunktionen,
 - d. Logarithmierung,
 - e. Lineares und exponentielles Wachstum
 - f. Funktionssynthese (diese Items wurden nicht gewertet).

Es wurden also unterschiedliche paper-and-pencil-Klausuren geschrieben, was nicht wünschenswert ist. Tatsache bleibt aber in der schulischen Realität, dass mittels derartiger Klausuren Leistungsbewertungen der Schülerinnen und Schüler vorgenommen werden und als objektiv, reliabel und valide im Sinne einer „best practice“ erklärt werden. Daher sollten sie auch mit den Ergebnissen des hier durchgeführten e-Assessments verglichen werden. Allerdings wird nur die Klasse mit der inhaltlich am besten vergleichbaren Klausur gewertet, da sonst die inhaltlichen Überlappungen als zu gering erscheinen.

Zunächst sollen die erzielten Ergebnisse des kompletten e-Assessments mit dem kompletten paper-and-pencil-Assessment verglichen werden. Für dieses Vorgehen spricht, dass dies der schulischen Realität entspricht, in der auch das komplette Assessment gewertet wird.

Ein zweiter Aspekt ist die Bewertung, also die **MEASUREMENT SCALE**. In der klassischen Bewertung der paper-and-pencil-Assessments werden die einzelnen Aufgaben unterschiedlich gewichtet. Dies entspricht in der Praxis einer zumeist nicht klar definierten Mischung aus Aufwand für die Bearbeitung, Schwierigkeitsgrad und Wichtigkeit für die Thematik. So schwierig das zu bewerten ist, so real und praktikabel ist diese Vorgehensweise.

Damit ist für das e-Assessment auf Basis des CSP eine Entscheidung über die Vorgehensweise bei der Bewertung der Item Responses erforderlich. Einerseits erscheint eine entsprechende Gewichtung der Items sinnvoll, da sie sich als praktikabel erwiesen hat und bei fehlender Gewichtung die Gefahr besteht, dass allein die Anzahl der Items einen Einfluss gewinnt. Andererseits ist für die statistische Analyse eine rein dichotome Bewertung erforderlich, also eine reine 0/1-Logik. Außerdem könnte eine Punktgewichtung auch als „willkürlich“ empfunden werden.

Trotz der Gefahr einer Verzerrung ist eine aus der Erfahrung in der Praxis entstandene Gewichtung vorgenommen worden. Zusätzlich ist der Vergleich des e-Assessments mit der konventionellen Klausur mit einer ungewichteten Bewertung des e-Assessments erfolgt. Da außerdem die zu Vergleichszwecken herangezogene paper-and-pencil-Klausur inhaltlich nicht hundertprozentig deckungsgleich mit dem e-Assessment war, sind einmal die Ergebnisse der kompletten Klausur und einmal die ausgewählten, inhaltlich deckungsgleichen Bereiche verglichen worden.

Umfang/Skala	Konventionell	Dichotom
Komplettes Assessment	(1)	(2)
Ausgewählte Kompetenzbereiche	(3)	(4)

Tabelle 22 Vier verschiedene Vergleichsvarianten für die Assessments

In Tabelle 22 sind vier verschiedene Vergleichsmöglichkeiten in Abhängigkeit von der verwendeten Gewichtung und des Umfangs der Assessments angegeben, die jetzt betrachtet werden.

8.2.2 Vergleich der kompletten Assessments mit konventioneller Gewichtung

Die Analyse erfolgt in der Excel-Datei M11V1.xlsx unter Berücksichtigung aller Items bis auf die Funktionssynthese.

Das Vorgehen in Excel besteht aus mehreren Schritten

- Die Punkte je Person sind als $\sum_{124 \text{ Items}} \text{Indikator} \cdot \text{Itemgewicht}$ (Spalte GC) im Datenblatt Bewertung berechnet worden.
- In Zelle GC84 und GD84 ist die Summe der erreichbaren Punkte angegeben.
- In Spalte GD sind die Prozentzahlen je Person als Punktzahl/GD84 berechnet.
- In Spalte GE sind daraus Noten berechnet, wobei der Prozentschlüssel für die Noten zugrunde gelegt wird, auf den das Assessment ausgerichtet ist.

Da die konventionelle paper-and-pencil-Klausur nur in einer Lerngruppe inhaltlich sinnvoll vergleichbar ist, sind die Ergebnisse des e-Assessments den Ergebnissen in der Klausur für diese Lerngruppe gegenübergestellt worden. Die Ergebnisse werden zur besseren Übersicht in die Excel-Tabelle SCO_ASSESSMENT.xlsx übernommen und dort als Datenblatt „SCO_ASSESSMENT1 gewichtet“ hinterlegt (CSP-Spalten).

Für die Personen dieser Klasse sind die Ergebnisse des konventionellen Assessments in Form der Noten und der Prozentwerte (ASS-Spalten) aufgelistet. Sowohl für die Noten als auch für die erreichten Prozentzahlen werden Korrelationen nach Pearson und Spearman bestimmt. Die Ergebnisse der Korrelation sind in Tabelle 23 zusammengefasst.

Korrelation	Nach Noten	Nach Prozenten
Pearson	0,84	0,81
Spearman	0,80	0,80

Tabelle 23 Korrelation mit konventionellem Assessment (gewichtet)

Es ergibt sich hier ein starker Zusammenhang zwischen beiden Assessments, der im Bereich der resultierenden Noten teilweise sogar noch größer ist als im Bereich der Prozentzahlen. Somit ist die Bewertung beider Assessments als ähnlich anzusehen.

8.2.3 Vergleich der kompletten Assessments mit dichotomer Bewertung (ohne Gewichtung)

Wendet man dasselbe Verfahren an, bewertet die einzelnen Items aber nicht mit gewichteten Punkten, sondern rein schematisch mittels einer dichotomen Zuordnung, also einem oder keinem Punkt pro Item ergeben sich die Zahlen, die ebenfalls in die Excel-Tabelle SCO_ASSESSMENT.xlsx übernommen und als Datenblatt „SCO_ASSESSMENT1 dichotom“ hinterlegt werden (CSP-Spalten).

Korrelation eine Lerngruppe	Nach Noten	Nach Prozenten
Pearson	0,79	0,78
Spearman	0,80	0,75

Tabelle 24 Korrelation mit konventionellem Assessment (dichotom)

Man sieht, dass praktisch kein Unterschied erkennbar ist. Dies lässt sicherlich noch keine allgemeine Aussage über die Unwirksamkeit einer Gewichtung zu, zeigt aber im konkreten Beispiel die Unempfindlichkeit gegenüber einer solchen klassischen Punktvergabe. Außerdem sind die Unterschiede zwischen Pearson und Spearman so gering, dass im Folgenden nur noch mit dem Korrelationskoeffizienten nach Pearson gearbeitet wird.

8.2.4 Vergleich ausgewählter Items mit Gewichtung

Der Vergleich der jeweils gesamten Assessments zeigt starke Zusammenhänge. Die nächste Frage ist, was passiert, wenn Items für bestimmte Kompetenzen herausgegriffen werden. Da hier der Bereich des Wachstums und der damit verbundenen Funktionen im Fokus steht, werden alle Items des e-Assessments ausgewählt, die zu diesen Bereichen gehören. Diesen werden die Ergebnisse des Teils der konventionellen Klausur gegenübergestellt, der ebenfalls diese Bereiche betrifft. Die Gewichtung ist dieselbe wie beim gesamten Assessment. Durch die Reduzierung auf 92 Items ergibt sich eine maximale Bewertung mit 131. Dem werden 79 Punkte aus dem konventionellen Assessment gegenübergestellt. Um auch die Anwendbarkeit reiner Datenbanktechnik zu zeigen, werden jetzt SQL-Abfragen in der Datenbank für die weitere Analyse verwendet. Konkret werden die Abfragen QRY_SCORE_BASE zur Berechnung der Kovarianz und QRY_SCORE_CORREL zur Berechnung des Korrelationskoeffizienten nach Pearson verwendet. Die Basis sind die Einträge in **PERSONAL MANIFEST SCORE**, die die **MEASUREMENT SCALE „2“** (in QRY_SCORE_BASE ist die group_id „0classic“), also die gewichtete Bewertung verwenden. Der berechnete Korrelationskoeffizient liegt bei $r = 0,70$ bezogen auf die Prozentzahl. Auf die Noten wird verzichtet, da sich bereits beim gesamten Assessment nur geringe Abweichungen zeigen.

8.2.5 Vergleich ausgewählter Items (dichotom)

Auch hier werden wieder die erwähnten 92 Items verwendet, jetzt aber mit einer rein dichotomen Bewertung ohne Gewichtung. Es werden wieder QRY_SCORE_BASE und QRY_SCORE_CORREL zur Berechnung des Korrelationskoeffizienten nach Pearson verwendet. Diesmal wird aber die **MEASUREMENT SCALE „1“** (in QRY_SCORE_BASE ist die group_id „0“) für dichotome Bewertung genutzt. Der berechnete Korrelationskoeffizient liegt dann bei $r = 0,65$. Das bedeutet, dass selbst im ungünstigsten Fall der dichotomen Bewertung noch ein klarer Zusammenhang zum konventionellen Assessment besteht. Eine mögliche Erklärung der kleineren Zusammenhänge kann in einer Verzerrung des Schwierigkeitsgrades gerade in der konventionellen Klausur durch Entfernung der nicht inhaltsgleichen Aufgaben liegen.

Die Daten sind quantitativ zu klein, um gesicherte Aussagen treffen zu können. Dies bedarf einer breiteren Analyse. Allerdings zeigt sich zumindest in der analysierten Situation ein deutlicher Zusammenhang, so dass erwartet werden kann, dass detailliertere Aussagen mittels des CSP-Modells ähnlich valide wie die aus einer Klausur zu gewinnenden Aussagen sind.

8.3 Ergebnisse aus der statistischen Analyse

Entsprechend der Idee in Schritt 1 erfolgt eine statistische Analyse des e-Assessments.

8.3.1 Erste statistische Analyse mit dem RASCH-Modell

Die konkrete statistische Analyse entsprechend dem RASCH-Modell wird zweimal durchgeführt, einmal mit den kompletten Items – wie angegeben mit Ausnahme der Items zur Funktionssynthese - und einmal nur mit den die Wachstumsthematik und die entsprechenden Funktionen betreffenden Items, die inhaltlich gut mit der parallelen klassischen Klausur vergleichbar waren.

Mit dem im Anhang in A.1 beschriebenen Skript wird die Analyse iterativ durchgeführt. In jeder Iteration wird zunächst das RASCH-Modell geschätzt. Dann werden die statistischen Tests zur Überprüfung der Qualität der ermittelten Werte durchgeführt. Es werden alle Items entfernt, die entweder vom LR-Test ausgeschlossen werden oder die im Wald-Test unter $p=0,05$ liegen. Dieser Vorgang wird siebenmal wiederholt, bis sowohl der LR-Test mit 0,943 einen guten Testparameter liefert als auch weder der LR-Test noch der Wald-Test noch der graphische Test signifikante Verletzungen bei einem Item zeigen.

Nr	Items	Vorher entfernte Items	LR-Test LR/chi2/p
1	121	Alle Synthese-Items	174/111/0
2	100	Auf Grund des LR-Tests EX06_02, EX10_03, EX11_01, EX11_03, EX12_02, EX12_03, EX16_01, W102_02, W103_02 Auf Grund des Wald-Tests EX01_01(0,018), EX03_12(0,000), EX07_01(0,001), EX07_06(0,017), EX07_12(0,001), EX13_03(0,012), EX15_02(0,017), W105_01(0,001), W109_06(0,001), BO02_05(0,040), BO02_11(0,032), BO03_08(0,048)	117/97/0,082
3	92	Auf Grund des LR-Tests W108_1 und W109_4 Auf Grund des Wald-Tests EX01_02(0,044), EX03_03(0,014), EX03_04(0,018), EX13_01(0,008), BO02_08(0,018), BO03_12(0,028)	87/91/0,592
4	90	Auf Grund des Wald-Tests EX02_01(0,035), BO04(0,049)	81/89/0,722
5	89	Auf Grund des Wald-Tests EX07_11(0,048)	78/88/0,788
6	87	Auf Grund des Wald-Tests W106_01(0,044), BO02_04(0,049)	68/95/0,906
7	85	Auf Grund des LR-Tests: W109_1 Auf Grund des Wald-Tests: W105_02(0,035)	65/84/0,943

Tabelle 25 Schritte bei der RASCH-Modell Analyse

In Tabelle 25 sind für jede Iteration die Anzahl Items, die vor dem Durchlauf entfernten Items und die Kontrollparameter des LR-Tests angegeben. Hier ist neben der Testcharakteristik und dem χ^2 -Wert besonders der für die Signifikanz relevante p-Wert interessant, der dem Signifikanzniveau entspricht und möglichst nahe bei 1 liegen muss.

Die genauen Testergebnisse einschließlich der ermittelten Kompetenzschätzungen für die einzelnen Assessment-Teilnehmer sind ebenfalls im Anhang beschrieben.

Diese Ergebnisse (θ , β und sonstige Parameter) können zur Befüllung insbesondere von **PERSONAL TRAIT SCORE, STATISTICAL VARIABLE, LATENT VARIABLE** und **STATISTICAL QUALITY PARAMETERS** verwendet werden.

Somit ergibt sich aus den Testergebnissen, dass das RASCH-Modell durchaus auf die verwendeten Items anwendbar ist und trotz der geringen Anzahl von Assessment-Teilnehmern Werte für eine Kompetenzschätzung liefert.

Auch der Algorithmus der Item-Elimination mit den Kriterien des LR-Tests und des Wald-Tests ist klar nachvollziehbar, auch wenn der graphische Test nur am Ende zur Überprüfung und in den Zwischenschritten rein informativ genutzt wird.

8.3.2 Zweite statistische Analyse mit dem RASCH-Modell

Um die Auswirkungen des Einsatzes unterschiedlicher Eliminationsstrategien zu testen, wird eine zweite Analyse begonnen, die mit denselben Items erfolgt.

Auch die zweite Analyse folgt einem iterativen Verfahren. Diesmal werden die Items nicht mit dem Wald-Test sondern mit einem graphischen Test herausgefiltert. In jedem Iterationsschritt werden die Items entfernt, die signifikant von der Hauptdiagonalen abweichen. Zusätzlich werden die signifikant zu schwierigen oder zu einfachen Items mittels des graphischen Tests herausgefiltert, da diese als geringer in ihrer Aussagekraft eingeschätzt werden. In jedem Schritt erfolgt wiederum eine Kontrolle mit dem LR-Test. In der folgenden Tabelle sind die Items und die Werte des LR-Tests dokumentiert.

Nr	Items	Vorher entfernte Items	LR-Test LR/chi2/p
1	114	EX06_01, W101_3, W101_4, W102_1, W105_1, W108_1, W109_6, W110_01	168/103/0
2	106	BO02_01, BO03_05, BO_04, EX10_02, EX13_01, EX11_02, W103_04, W109_3	153/95/0
3	93	BO01_01, BO01_02, EX06_02, EX10_03, EX11_01, EX11_03, EX11_04, EX12_02, EX12_03, EX16_01, W102_02, W103_02, W109_04	132/92/0,004
4	84	BO02_05, EX01_01, EX03_12, EX07_01, EX07_06, EX07_12, EX13_04, EX15_02, W107_01	90/83/0,281
5	76	BO02_11, BO03_01, BO03_04, EX02_01, EX03_03, W102_04, W109_01, W110_03	57/75/0,942
6	72	BO01_04, EL01_01, EL01_02, W109_02	46/71/0,989

Tabelle 26 Schritte bei der RASCH-Modell Analyse

Man sieht, dass der p-Wert nach der fünften Iteration sprunghaft ansteigt und der Test dann ebenfalls sehr befriedigende Ergebnisse zeigt. Die entsprechenden Werte des LR-Tests wie auch des anschließend durchgeführten Wald-Tests für die verbliebenen Items zeigen ebenfalls keine Verletzungen der Signifikanzniveaus und sind im CSP-Modell dokumentiert. Somit ergeben sich auch hier sinnvolle Schätzungen für die Kompetenzen.

Die Werte für die geschätzten Kompetenzen θ_i der einzelnen Assessment-Teilnehmer werden ebenfalls direkt im CSP-Kompetenzmodell dokumentiert.

8.3.3 Vergleiche der statistischen mit der manifesten Analyse

Nachdem die beiden statistischen Analysen durchgeführt sind und die Tests auf eine Zuverlässigkeit der ermittelten Kompetenzen hinweisen, wird die Korrelation der ermittelten Kompetenzen (Traits), also der latenten Variablen aus beiden Tests miteinander und mit den Ergebnissen der manifesten Scores des e-Assessments bestimmt. Die Ergebnisse sind in Tabelle 27 zusammengefasst. Der Score, die mittels des CSP ermittelt wurden, werden als „e-Assessment CSP“ bezeichnet. Die geschätzten Traits für die Assessments-Teilnehmer aus den beiden statistischen Analysen sind mit Theta1 und Theta2 bezeichnet.

Inhaltliche Gegenüberstellung	Korrelationskoeffizient nach Pearson
e-Assessment CSP gewichteter Score <-> Theta1	0,89
e-Assessment CSP dichotom <-> Theta1	0,89
e-Assessment CSP gewichteter Score <-> Theta2	0,94
e-Assessment CSP dichotom <-> Theta2	0,93
Theta1 <-> Theta2	0,86

Tabelle 27 Korrelation der ermittelten Kompetenzen (Theta)

Dabei werden zunächst die in der ersten statistischen Analyse (Theta1) ermittelten Kompetenzen für alle Assessment-Teilnehmer mit den CSP-Scores aus der manifesten Analyse mittels gewichteter Bewertung und mittels dichotomer Bewertung gegenübergestellt. Ebenso wird mit den Ergebnissen der zweiten statistischen Analyse (Theta2) verfahren. Man sieht, dass der Einfluss der Messskala, also gewichtet oder dichotom, kaum erkennbar ist. In beiden Fällen ergibt sich eine hohe Übereinstimmung der gemessenen Werte. Daraus lässt sich schließen, dass die mittels der Bewertung der Ergebnisse ermittelten Aussagen über die Assessment-Teilnehmer in hohem Maße mit den statistisch aus den denselben Werten ermittelten Kompetenzen übereinstimmen. Bei einem validen RASCH-Modell bedeutet das, dass die Ergebnisse des Assessments Aussagen liefern, die über das eigentliche Assessment hinaus auch für die damit zu messenden Kompetenzen gelten.

In einem weiteren Schritt werden die Ergebnisse der paper-and-pencil-Klausur, die bereits in Abschnitt 8.2 mit den Scores aus dem CSP-Modell korreliert wurden mit den geschätzten Thetas korreliert. Dabei zeigen sich zwar immer noch deutliche, allerdings geringere Zusammenhänge, im Mittel ergibt sich ein Korrelationskoeffizient von $r = 0,64$. Daraus kann man folgern, dass entweder die Thetas der Kompetenzen nur teilweise die Ergebnisse widerspiegeln, die mit der klassischen Klausur erzielt werden, oder die Klausur nur teilweise die Kompetenzen der Assessment-Teilnehmer wiedergibt. Der hohe Zusammenhang zwischen den Scores des e-Assessments und den Thetas spricht eher für die zweite These.

Schließlich werden zur Kontrolle der Vorgehensweise bei der Ermittlung der beiden Thetas diese miteinander korreliert und weisen mit 0,86 die erwartete hohe Korrelation auf, was die Stabilität der verwendeten Algorithmen bestätigt.

8.4 Ergebnisse auf Kompetenzebene

Nach der Gesamtsicht sollen jetzt beispielhaft die Ergebnisse für bestimmte Skills und Fakten untersucht werden, also eine detaillierte Analyse auf der Ebene der Fähigkeiten und Aussagen erfolgen.

8.4.1 Ergebnisse auf Skill- und Faktenebene

Für einige Skills werden die Items selektiert, die diesen Skills entsprechen. Für die unter fachlichen Aspekten ausgewählten Items werden die **MANIFEST VARIABLES** mittels **SCO_AGGREGATE** zu neuen **MANIFEST VARIABLES** gruppiert. Für diese Gruppen wird der (dichotome) Score ermittelt. Außerdem wird eine statistische Analyse nur für diese Variablen durchgeführt und somit die Traits für diesen Skill geschätzt. Wiederum werden Korrelationen zwischen den Scores und Traits gebildet, um zu prüfen, inwieweit mit einer vergleichsweise geringen Anzahl Items ein individueller Trait ermittelbar ist.

Inhalt	Typ	Anzahl Items	Korrelation	Anmerkung
FACT1103e	Skill	12	0,95	11 Personen nicht schätzbar

Tabelle 28 Skill-basierter Zusammenhang zwischen Score und Trait

Die Tabelle 28 zeigt, dass es möglich ist, einen Zusammenhang zwischen der Bewertung (Score) und dem ermittelten Trait auf der Ebene der Skills herzustellen. Somit gibt es Anzeichen, dass in diesen Fällen der Score auch zur Kompetenzermittlung der betroffenen Kompetenzen herangezogen werden kann. Es kann aber nicht für jeden Assessment-Teilnehmer eine Schätzung vorgenommen werden, in einigen Fällen liefert das stochastische Modell kein valides Ergebnis. Vermutlich liegt dies an einer Mischung aus der geringen Anzahl Items und dem spezifischen Antwortmuster.

Führt man die Detailanalyse auf Faktenebene durch stehen mehr Items zur Verfügung als auf Ebene der Skills, weil alle Skills, die sich auf einen Fakt beziehen zusammengefasst werden. Es werden hier die Fakten ausgewählt, die laut Abbildung 47 zu einer Kompetenzaussage aggregiert werden. Für diese Fakten 1105, 1106, 1107 und 1108 werden die betreffenden Items ausgewählt und den Skills (Fakt und Operator) zugeordnet. Diese Items werden dann wieder mittels des RASCH-Modells analysiert und Traits für die einzelnen Fakten ermittelt. Die so ermittelten Traits werden den dichotomen Scores für diese Facts gegenübergestellt. Die Scores beruhen genau auf den entsprechenden Items. Für alle ausgewählten Fakten zeigt sich eine deutliche Korrelation zwischen dem geschätzten Trait und den berechneten Scores. Bei der Schätzung der Traits zeigt sich wiederum, dass bei wenigen Items die Traits nicht immer für alle Personen geschätzt werden können. Die Modellparameter p sind für die verschiedenen Fakten auch von unterschiedlicher Qualität. Dies scheint aber nicht nur von der Anzahl der Items abzuhängen. So liegt der p -Value für FACT1105 mit 15 Items nur bei 0,86, bei FACT1108 mit nur 4 Items dagegen bei 0,98.

Inhalt	Typ	Anzahl Items in dem endgültigen Modell	Korrelation	Anmerkung
FACT1105	FACT	15	0,86	48/76 Personen schätzbar
FACT1106	FACT	4	0,95	21/76 Personen schätzbar
FACT1107	FACT	21	0,91	70/76 Personen schätzbar
FACT1108	FACT	4	0,98	53/76 Personen schätzbar

Tabelle 29 Korrelation zwischen Score und Trait für ausgewählte Fakten

Trotz der unterschiedlichen Qualität der statistischen Modelle unterstützen diese doch die Aussagekraft der Scores für die Kompetenzbestimmung, da sich in allen Fällen eine starke Korrelation zwischen dem statistisch geschätzten Trait (Kompetenzwert) und dem rechnerisch im manifesten Modell ermittelten Score ergibt.

8.4.2 Scoring mit Anforderungsniveaus

Basierend auf diesen Ergebnissen wird beispielhaft mit den genannten vier Fakten eine Bestimmung des Kompetenzniveaus für die Kompetenz C110 vorgenommen. Laut Abbildung 47 werden die Ergebnisse auf fünf Kompetenzniveaus bestimmt, die durch die Schwellenwerte 0,2, 0,4, 0,6

und 0,8 unterschieden werden. Die entsprechende Dokumentation der Schwellenwerte ist in **LEVELINDICATION** hinterlegt worden. Die Auswertung ist nur mit SQL-Mitteln erfolgt und die entsprechenden Auswertungen **COMP_NIVEAU_METRIC_2**, **COMP_NIVEAU_METRIC_2_BASIS** und **COMP_NIVEAU_METRIC_2_FINAL** entsprechen den zugeordneten **COMPETENCE NIVEAU METRICS** „2“.

103	0,732142857142857	0,6	Funktionaler Zusammenhang weitgehend beherrscht
104	0,646904761904762	0,6	Funktionaler Zusammenhang weitgehend beherrscht
105	0,834285714285714	0,8	Funktionaler Zusammenhang klar beherrscht
106	0,695952380952381	0,6	Funktionaler Zusammenhang weitgehend beherrscht
107	0,567380952380952	0,4	Funktionaler Zusammenhang teilweise beherrscht
108	0,585238095238095	0,4	Funktionaler Zusammenhang teilweise beherrscht
109	0,702142857142857	0,6	Funktionaler Zusammenhang weitgehend beherrscht
110	0.604047619047619	0.6	Funktionaler Zusammenhang weitgehend beherrscht

Abbildung 71 Ergebnis der Bestimmung des Kompetenzniveaus

Die Ergebnisse der Auswertung sind ausschnittsweise in Abbildung 71 dargestellt. Die Darstellung erfolgt zeilenweise mit der Personennummer, dem erreichten gewichteten Score über alle vier Fakten, dem überschrittenen Schwellenwert und der zugeordneten Kompetenzbeschreibung aus **SPECIFIC COMPETENCE NIVEAU** bezogen auf die Kompetenz C110.

Bei der Ermittlung der Werte wird auf die dichotomen Scores der einzelnen Assessment-Teilnehmer für die vier Fakten zurückgegriffen, die auch im vorherigen Abschnitt verwendet werden. Diese Scores werden mittels der Gewichte in **SKILL INDICATING** gewichtet und anschließend mittels der Schwellenwerte in **LEVELINDICATION** zugeordnet.

Damit kann jeder Person hinsichtlich der Kompetenz C110 ein Kompetenzniveau zugeordnet werden. Man kann durchaus gegen diese Zuordnung argumentieren, indem man sagt, dass die Aussagekraft dieser Einordnung der einzelnen Assessment-Teilnehmer von vielen (auch anderen) Faktoren abhängig sein kann, die hier nicht berücksichtigt werden. Die Existenz weiterer Indikatoren anzunehmen ist allerdings immer möglich. Allein durch eine andere Wahl der Gewichte, Items und anderer Einflussfaktoren kann man zu anderen Ergebnissen kommen. Valide Assessments sollten allerdings ähnliche Ergebnisse liefern. Die geeignete Konstruktion des Assessments und die Wahl der Modellbestandteile sind eine grundlegende Eigenschaft jedes Assessments und müssen mit entsprechender fachlicher Kompetenz erfolgen.

Hier ist aber zumindest die gesamte Vorgehensweise und Grundlage von der Definition der Kompetenzen, deren Niveaus, der zugeordneten Fakten, der daraus abgeleiteten Testitems, deren Messung und die folgende Aggregation dokumentiert. Die Ergebnisse werden zusätzlich durch Kompetenzschätzungen für die einzelnen Fakten gestützt.

Somit bleibt festzuhalten, dass es möglich ist, eine plausible Schätzung für die Assessment-Teilnehmer auch für einzelne Kompetenzen mittels dieses Modells zu gewinnen.

8.5 Individuelle Analyse

Es wird jetzt die individuelle Verteilung der Fähigkeiten für einen Assessment-Teilnehmer ermittelt. Dabei erfolgt wieder eine Konzentration auf die im vorherigen Abschnitt betrachteten Fakten, also F1105, F1106, F1107 und F1108.

Zunächst wird auf Basis dieser Fakten mittels einer Abfrage (COMP_NIVEAU_METRIC_2_MEAN) der mittlere Score-Wert aller Assessment-Teilnehmer bezüglich der Kompetenz 101 als 0,71 ermittelt.

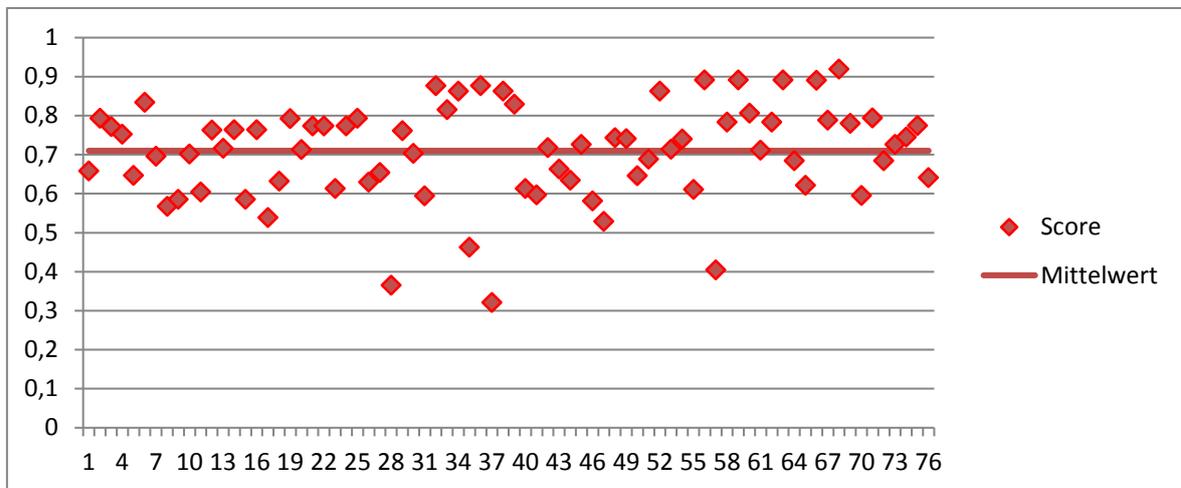


Abbildung 72 Scores (Anteile) der Assessment-Teilnehmer für Kompetenz C110

In Abbildung 72 ist der mittlere Score-Wert für die Kompetenz C110 bezogen auf die Wachstumsprozesse als Anteil zwischen 0 und 1 dargestellt. Dazu sind die jeweiligen Scores der 76 Assessment-Teilnehmer wiedergegeben. Dies entspricht genau den Werten, die in Abschnitt 8.4.2 bereits für die Assessment-Teilnehmer ermittelt wurden. Zusätzlich ist der Durchschnittswert 0,71 eingetragen.

Jetzt werden für zwei dieser Teilnehmer mit annähernd identischem Scoring die Einzelwerte für die beteiligten Fakten herausgegriffen, um Unterschiede in den persönlichen Fähigkeiten analysieren zu können. Dazu werden bewusst zwei Assessment-Teilnehmer ausgewählt, deren Score im Durchschnitt liegt. Dies sind die in Abbildung 72 als Nummer 10 (Person 109) und Nummer 42 (Person 61) dargestellten Teilnehmer. Für diese beiden Personen werden mittels Abfragen die Scores für die einzelnen Fakten als prozentualer Anteil ermittelt und mit den ebenfalls berechneten mittleren Scores pro Fakt verglichen. Für die mittleren Scores pro Fakt wird die Abfrage COMP_NIVEAU_METRIC_2_ALL_MEANS verwendet. Die einzelnen Scores pro Fakt und Person können direkt der bereits zuvor erstellten Abfrage COMP_NIVEAU_METRIC_2 entnommen werden, die die einzelnen Anteile pro Fakt und zusätzlich die gewichteten Gesamt-Scores pro Assessment-Teilnehmer bestimmt.

In Abbildung 73 sieht man die Ergebnisse beider Assessment-Teilnehmer. Da beide mit ihren Anteilen für die Kompetenz C110 im Durchschnitt liegen, sind beide in einigen diese Kompetenz bestimmenden Fakten besser, in anderen schlechter als der Durchschnitt. Man sieht aber auch, dass die individuellen Stärken und Schwächen unterschiedlich verteilt sind, insbesondere ergeben sich Unterschiede bei den Fakten F1106 und F1107. Man könnte aus dem Gesamtbild die Hypothese aufstellen, dass der Teilnehmer 61 das exponentielle Wachstum besser „verstanden“ hat, während die Stärken des Teilnehmers 109 eher im Bereich des linearen Wachstums liegen.

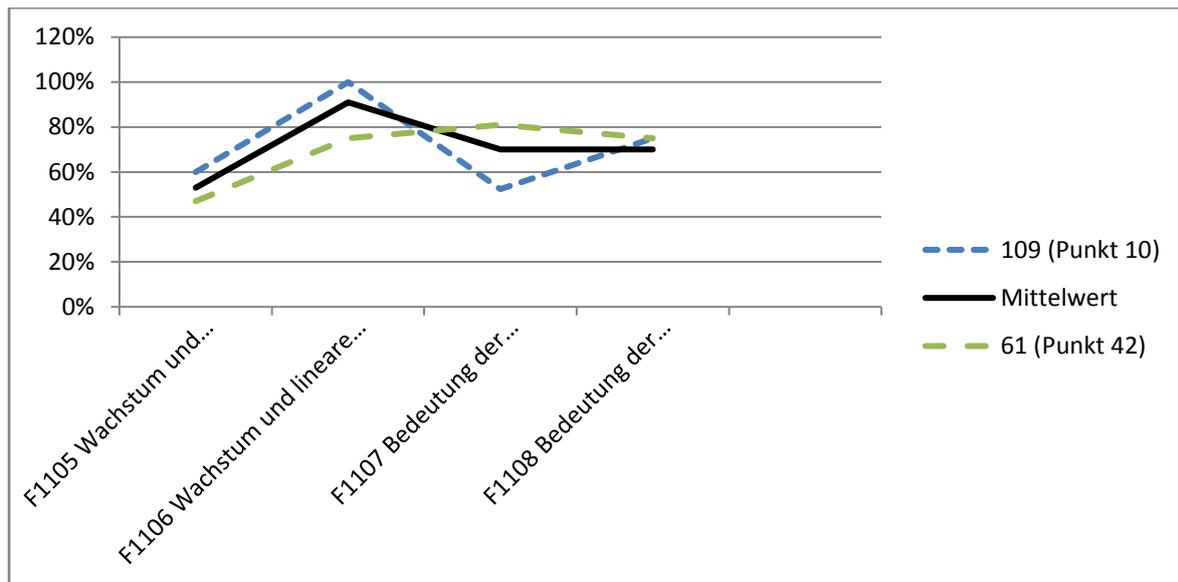


Abbildung 73 Scores pro Fact für zwei Assessment-Teilnehmer im Vergleich zum Durchschnitt

Analog lassen sich für alle Teilnehmer die Zahlen und Diagramme ermitteln. Außerdem ließe sich für alle Fakten (und bei entsprechender Detaillierung Skills) ein Mittelwert und entsprechende Standardabweichungen ermitteln. Ein weiteres Beispiel findet man in einem früheren Beitrag (Wieken und Forbrig 2014).

Damit ergeben sich zwei Möglichkeiten des Feedbacks. Zunächst kann der Lehrende auf Basis der Mittelwerte für alle Fakten (Skills) ein detailliertes Feedback der bei den Teilnehmern vorhandenen Fähigkeiten bekommen. Dabei handelt es sich um das von Hattie angesprochene besonders wirksame Feedback „[...] wenn es der Lehrperson von den Lernenden gegeben wird.“ (Hattie 2013, 206).

Der zweite Aspekt ist das Feedback für die Lernenden, indem die Assessment-Teilnehmer nicht nur eine Gesamtbewertung erhalten, sondern eine detaillierte Analyse ihrer Stärken und Schwächen bezogen auf die erwarteten Fähigkeiten. Damit entsteht eine Basis dafür, dass sich Unterricht und Feedback miteinander verweben, bis der Prozess selbst Formen des Unterrichts annimmt (Hattie 2013, 207). Allerdings sollte immer dann, wenn anschließend auch eine Bewertung erfolgt, der Unterschied zwischen Lernphase und Assessment-Phase deutlich bleiben.

8.6 Feedback

Die Analyse der spezifischen Kompetenzniveaus soll in einem Feedback münden, das einerseits individuelle Aussagen für den einzelnen Assessment-Teilnehmer als auch andererseits für die Lehrenden auf Gruppenebene Möglichkeiten des Ausgleichs von Defiziten bietet.

Die Zuordnung kann mittels der Verbindung zwischen dem **SPECIFIC COMPETENCE NIVEAU** und dem **FEEDBACK** erfolgen. Beispielhaft ist dies in Abbildung 74 dargestellt.

uid	Definition	Item	Item Definition
101	Funktionaler Zusammenhang	EX10.1	$f(x) = 5 \cdot 4^x$
101	Funktionaler Zusammenhang	EX10.3	$f(x) = 5x + 4$
101	Funktionaler Zusammenhang	EX10.4	$f(x) = 4x + 5$
101	Funktionaler Zusammenhang	EX12.2	Beim exponentiellen Wachstum gilt eine Funktionsgleichung der Form $f(x)=ax+b$
101	Funktionaler Zusammenhang	EX12.3	Beim exponentiellen Wachstum gilt eine Funktionsgleichung der Form $f(x)=a \cdot b^x$
101	Funktionaler Zusammenhang	EX12.4	Beim exponentiellen Wachstum gibt in $f(x)=ax+b$ das b an, wieviel am Anfang vorhanden ist.
101	Funktionaler Zusammenhang	EX16.1	$f(x) = \dots$
101	Funktionaler Zusammenhang	W101.3	$f(x)=150000 \cdot 1,06^x$
101	Funktionaler Zusammenhang	W101.4	$f(x)=1,06x + 150000$
101	Funktionaler Zusammenhang	W102.4	Es handelt sich um exponentielles Wachstum.

Abbildung 74 Ausschnitt der aus einer Kompetenz abgeleiteten Items

Inhaltlich kann auf die Kausalbeziehung zwischen der Kompetenz und den Items zurückgegriffen werden, die letztlich über den umfangreich dokumentierten Weg der Fakten und Snippets, ergänzt zu Skills, bis zu den einzelnen Items nachvollziehbar ist. Dies erlaubt eine inhaltliche Zuordnung des jeweils sinnvollen Feedbacks.

Soll eine genauere und inhaltlich besser abgestimmte Zuordnung des Feedbacks bezogen auf ein erreichtes Kompetenzniveau bezüglich einer bestimmten Kompetenz ermöglicht werden, muss eine tiefere Analyse erfolgen. Hier kann beispielsweise untersucht werden, welche der einer bestimmten Kompetenz zugeordneten Items wie oft von den Assessment-Teilnehmern richtig beantwortet wurden, die über ein bestimmtes Kompetenzniveau verfügen. Dabei sind sowohl das Kompetenzniveau als auch die Items immer auf eine bestimmte Kompetenz bezogen.

Kompetenzniveau	Schwelle	Item	Lösungen	Personen	Anteil
Funktionaler Zusammenhang kaum beherrscht	0,2	EX10.1	1	2	50,00%
Funktionaler Zusammenhang teilweise beherrscht	0,4	EX10.1	8	11	72,73%
Funktionaler Zusammenhang weitgehend beherrscht	0,6	EX10.1	38	49	77,55%
Funktionaler Zusammenhang klar beherrscht	0,8	EX10.1	12	14	85,71%
Funktionaler Zusammenhang kaum beherrscht	0,2	EX10.3	1	2	50,00%
Funktionaler Zusammenhang teilweise beherrscht	0,4	EX10.3	11	11	100,00%
Funktionaler Zusammenhang weitgehend beherrscht	0,6	EX10.3	47	49	95,92%
Funktionaler Zusammenhang klar beherrscht	0,8	EX10.3	14	14	100,00%
Funktionaler Zusammenhang kaum beherrscht	0,2	EX10.4	2	2	100,00%
Funktionaler Zusammenhang teilweise beherrscht	0,4	EX10.4	6	11	54,55%
Funktionaler Zusammenhang weitgehend beherrscht	0,6	EX10.4	38	49	77,55%
Funktionaler Zusammenhang klar beherrscht	0,8	EX10.4	12	14	85,71%
Funktionaler Zusammenhang kaum beherrscht	0,2	EX12.2	1	2	50,00%
Funktionaler Zusammenhang teilweise beherrscht	0,4	EX12.2	10	11	90,91%
Funktionaler Zusammenhang weitgehend beherrscht	0,6	EX12.2	48	49	97,96%
Funktionaler Zusammenhang klar beherrscht	0,8	EX12.2	14	14	100,00%
Funktionaler Zusammenhang kaum beherrscht	0,2	EX12.3	1	2	50,00%
Funktionaler Zusammenhang teilweise beherrscht	0,4	EX12.3	9	11	81,82%
Funktionaler Zusammenhang weitgehend beherrscht	0,6	EX12.3	48	49	97,96%
Funktionaler Zusammenhang klar beherrscht	0,8	EX12.3	14	14	100,00%
Funktionaler Zusammenhang teilweise beherrscht	0,4	EX12.4	7	11	63,64%
Funktionaler Zusammenhang weitgehend beherrscht	0,6	EX12.4	41	49	83,67%
Funktionaler Zusammenhang klar beherrscht	0,8	EX12.4	14	14	100,00%
Funktionaler Zusammenhang teilweise beherrscht	0,4	EX16.1	1	11	9,09%
Funktionaler Zusammenhang weitgehend beherrscht	0,6	EX16.1	8	49	16,33%
Funktionaler Zusammenhang klar beherrscht	0,8	EX16.1	9	14	64,29%

Abbildung 75 Lösungen pro Item für Kompetenzniveaus einer Kompetenz

Eine solche Analyse ist in Abbildung 75 dargestellt, wobei für die Kompetenz 101 jeweils das Niveau und der dazu definierte Schwellenwert dargestellt wird. Für eine beispielhafte Auswahl der dazugehörigen Items wird angegeben, wie viele der Personen, die dem entsprechenden

Niveau zugeordnet werden, das entsprechende Item richtig beantwortet haben. Diese Zahl ist auf die Gesamtzahl der Personen bezogen, die insgesamt dem Kompetenzniveau zugeordnet werden. Man erkennt, dass der Anteil der richtigen Lösungen in den meisten Fällen mit zunehmendem Niveau ansteigt, wie es zu erwarten war. Bei kleineren Anzahlen kann es auch zufällige Schwankungen geben. Aus den Anteilen erkennt man zusätzlich, welche Items auf welchem Niveau besondere Schwierigkeiten verursachen. Die inhaltliche Analyse dieser Items kann dann zur Festlegung der für die jeweiligen Kompetenzstufen sinnvollen Feedback- und Unterstützungsaktionenklassen dienen.

Kompetenzniveau	Feedback	Vors
Funktionaler Zusammenhang teilweise beherrscht	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) = a \cdot b^x$ a den Anfangswert und b den Wachstumsfaktor bei exponentiellem Wachstum angibt.	Aufgaben, die den Anfangswert bei exponentiellem Wachstum ermitteln.
Funktionaler Zusammenhang teilweise beherrscht	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) = a \cdot b^x$ a den Anfangswert und b den Wachstumsfaktor bei exponentiellem Wachstum angibt.	Aufgaben, die den Wachstumsfaktor bei exponentiellem Wachstum ermitteln.
Funktionaler Zusammenhang kaum beherrscht	Die Bedeutung des Exponenten ist unklar. Vermutlich wird Periode falsch eingegeben.	Aufgaben zum Aufbau einer Exponentialfunktion.
Funktionaler Zusammenhang nicht beherrscht	Der Grundaufbau einer Exponentialfunktion ist nicht bekannt.	Aufgaben zum Aufbau einer Exponentialfunktion.
Funktionaler Zusammenhang teilweise beherrscht	Die Bedeutung der Parameter a und b als Anfangswert und Wachstumsfaktor ist unklar.	Aufgaben mit Angabe von Anfangswert und Wachstumsfaktor.
Funktionaler Zusammenhang weitgehend beherrscht	Logischer Fehler bei der Bestimmung der Basis.	Aufgaben zur Ermittlung der Parameter für eine Exponentialfunktion.
Funktionaler Zusammenhang kaum beherrscht	Es ist nicht klar, dass zur Beschreibung linearer Wachstums eine lineare Funktion verwendet werden muss.	Aufgaben mit verschiedenen Funktionstypen.
Funktionaler Zusammenhang kaum beherrscht	Es ist nicht klar, dass zur Beschreibung exponentiellen Wachstums eine Exponentialfunktion verwendet werden muss.	Aufgaben mit verschiedenen Funktionstypen.
Funktionaler Zusammenhang nicht beherrscht	Es ist nicht klar, dass zur Beschreibung linearer Wachstums eine lineare Funktion verwendet werden muss.	Aufgaben mit verschiedenen Funktionstypen.
Funktionaler Zusammenhang nicht beherrscht	Es ist nicht klar, dass zur Beschreibung exponentiellen Wachstums eine Exponentialfunktion verwendet werden muss.	Aufgaben mit verschiedenen Funktionstypen.
Funktionaler Zusammenhang kaum beherrscht	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) = a \cdot b^x$ a den Anfangswert und b den Wachstumsfaktor bei exponentiellem Wachstum angibt.	Aufgaben, die den Anfangswert bei exponentiellem Wachstum ermitteln.
Funktionaler Zusammenhang nicht beherrscht	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) = a \cdot b^x$ a den Anfangswert und b den Wachstumsfaktor bei exponentiellem Wachstum angibt.	Aufgaben, die den Anfangswert bei exponentiellem Wachstum ermitteln.
Funktionaler Zusammenhang teilweise beherrscht	Es ist nicht klar, dass bei einer linearen Funktion $f(x) = ax + b$ a die Steigung und b den Anfangswert bei linearer Zunahme angibt.	Aufgaben, die den Anfangswert bei linearer Zunahme ermitteln.
Funktionaler Zusammenhang kaum beherrscht	Es ist nicht klar, dass bei einer linearen Funktion $f(x) = ax + b$ a die Steigung und b den Anfangswert bei linearer Zunahme angibt.	Aufgaben, die den Anfangswert bei linearer Zunahme ermitteln.
Funktionaler Zusammenhang nicht beherrscht	Es ist nicht klar, dass bei einer linearen Funktion $f(x) = ax + b$ a die Steigung und b den Anfangswert bei linearer Zunahme angibt.	Aufgaben, die den Anfangswert bei linearer Zunahme ermitteln.

Abbildung 76 Ausschnitt Zuordnung Niveau und Feedback

Ein weiteres sinnvolles Hilfsmittel für diese Festlegung kann auch eine Fehleranalyse sein, wie sie im Anhang B beispielhaft zu sehen ist. Dabei geht es nicht um eine vollständige Analyse aller möglichen Fehlerarten und deren vollständige Zuordnung, sondern um die punktuelle Auswertung auf Basis erkennbarer Fehlermuster.

Die Ergebnisse dieser Analysen können für den detaillierten Aufbau der Feedbackstruktur genutzt werden.

So kann man beispielsweise am Item EX12.2 ablesen, dass für dieses Item ab einer Schwelle von 0,4, also „Funktionaler Zusammenhang teilweise beherrscht“ die Lösungshäufigkeit bereits über 90% liegt. Die durch das Item geforderte Fähigkeit, den richtigen Funktionstyp zuzuordnen, sollte also insbesondere Personen mit niedrigerem Kompetenzlevel in einem Feedback als problematisch aufgezeigt werden. Bestätigt wird dies durch das analoge Item EX12.3. Die Items EX10.1 und EX10.4 zeigen, dass die Fähigkeit, den richtigen Startwert, beziehungsweise den Funktionswert an der Stelle 0, zuzuordnen, selbst bei höheren Kompetenzniveaus nicht besonders ausgeprägt ist.

Das Item EX16.1 erfordert die komplette Eingabe einer Funktionsgleichung. Entsprechend wird dies relativ selten komplett gelöst. Hier ist beispielsweise eine Fehleranalyse wie im Anhang B angegeben sinnvoll, um detaillierte Informationen zu erhalten.

Aus den so entwickelten Vorschlägen kann dann im CSP ein direkter Zusammenhang zwischen dem Kompetenzniveau und dem Feedback erstellt werden, der ausschnittsweise in Abbildung 76 dargestellt ist.

8.7 Zeitbedarf

Abschließend sollen noch einige Angaben zum Zeitbedarf für die Entwicklung und Durchführung des Assessments gemacht werden. Der Zeitbedarf für die Domänenanalyse und die Beschreibung des Domäneninhaltes ist erheblich. Er lässt sich schlecht abschätzen, da er zumindest vom Umfang der Domäne, den vorliegenden Materialien, dem gewünschten Detaillierungsgrad und der Erfahrung der Modellierer abhängt. Bestimmte Ausschnitte wie die Kompetenzskalen sind einmalig zu erstellen. Dieser Aufwand kann nicht berücksichtigt werden. Eine einzelne Person würde nach den Erfahrungen mit dem dokumentierten Ausschnitt für die gesamte Mathematik der 11. Klassenstufe circa einen Monat benötigen.

Der Aufwand für den Entwurf eines e-Assessments ist größer als der Aufwand für den Entwurf eines klassischen paper-and-pencil-Assessments. Hier ist nach der bisherigen Erkenntnis je nach Erfahrung und verwendetem Werkzeug sicher ein Faktor im Bereich 2-10 anzusetzen. Dabei ist allerdings zu berücksichtigen, dass ein e-Assessment, zumal wenn es statistisch analysiert werden soll, immer einige Sorgfalt erfordert. Allerdings sind durch die Vorarbeit bei der Domänenanalyse, eine bessere Ausgewogenheit und ein klarerer Bezug zu den zu prüfenden Fähigkeiten leichter zu realisieren.

Ein weiterer Vorteil liegt in der Wiederverwendbarkeit und dem sukzessiven Lernen. So können einmal designte Items nicht nur wiederverwendet werden, es stehen vielmehr bereits Erfahrungen über mögliche Antworten und deren mögliche Bewertung zur Verfügung. Zudem können bessere Vergleiche mit früheren Assessments gezogen werden. Somit dürfte nach mehrmaliger Anwendung der Aufwand nicht mehr wesentlich über dem eines klassischen Assessments liegen.

Erhebliche Zeiteinsparungen ergeben sich bei der Ermittlung der Ergebnisse. Bei dem hier verwendeten Assessment ergaben sich nach der Durchführung für 76 Assessment-Teilnehmer:

- Download der Daten ca. 10 Minuten,
- zwei Datensätze korrigieren (1 undefinierbarer Bedienungsfehler, 1 versehentliches Ausschalten des Rechners durch einen anderen Assessment-Teilnehmer.) ca. 20 Minuten,
- Behebung der Rundungsfehler, ca. 15 Minuten,
- Behebung sonstiger RAW ERROR TYPEs ca. 30 Minuten,
- Erstellung der Bewertung (Gesamt-Assessment) ca. 15 Minuten,
- statistische Analyse (Gesamt-Assessment) ca. 60 Minuten.

Somit waren ca. 2,5 Stunden für die Erstellung von Ergebnissen einschließlich einer globalen Kompetenzschätzung notwendig. Die Korrektur einer klassischen paper-and-pencil-Klausur in der gymnasialen Oberstufe ist sicher von vielen Faktoren abhängig. Allerdings dürfte eine Schätzung von 30 bis 60 Minuten pro Klausur aus eigener Erfahrung nicht unrealistisch hoch sein. Somit stehen 2,5 Stunden etwa 38 bis 76 Stunden bei einer konventionellen Korrektur gegenüber.

Generell kann man feststellen, je offener ein Item ist, desto höher ist der Aufwand für dessen Aufbereitung und Analyse, daher sind hier eher wenige Items empfehlenswert. Diese können bei der Analyse genutzt werden, um Fehlerhypothesen zu verfeinern, wie dies mit dem Item EX16.1 hier exemplarisch geschehen ist. Damit kann ein höherer Aufwand in deren Analyse und das gezielte Feedback für die Assessment-Teilnehmer investiert werden.

Somit ergibt sich als Fazit, dass der höhere Aufwand für die erstmalige Erstellung der Domänenanalyse bereits nach 4 - 5 derartigen Assessments wieder eingespart werden kann.

9 Zusammenfassung

Hinsichtlich der ursprünglichen Frage nach einem Modell für die Kompetenzbeschreibung wird ein dreistufiges Informationsmodell als Meta-Modell, das in der Unified Modeling Language (UML) beschrieben ist, entwickelt. Das Modell wird als Core-Situation-Personal-Modell (CSP) bezeichnet und beinhaltet eine Reihe von Teilmodellen, die Ziele, Kompetenzen, Kompetenzniveaus, Inhalte, Assessments und tatsächliche Ergebnisse mit manifesten Variablen für die Ergebnisse und latenten Variablen für die stochastische Kompetenzschätzung mit psychometrischen Modellen beschreiben. Zusätzlich sind ein Vorgehensmodell und zwei graphische Notationen für die Modellierung des Inhalts und des Zusammenhangs zwischen Ergebnissen und Kompetenzen Bestandteil des Modells.

Eine Voraussetzung für die Entwicklung des Modells ist die Klärung der Begrifflichkeiten. Dafür werden die Begriffe Kompetenz, Qualifikation und Performanz in Bezug auf den Erfolg in einem Assessment gegeneinander abgegrenzt und insbesondere der Kompetenzbegriff in einer operationalisierten Form als Kriterium für die Erfolgskontrolle ausgewählt. Viele Begriffe werden auch im Zusammenhang mit large-scale Assessments wie PISA oder TIMMS verwendet, obwohl dies nicht der Fokus dieser Arbeit ist. Es werden aber andere Ansätze wie Wilsons „Four Building Blocks“ oder das „Evidence-Centered Design“ ebenso wie Ergebnisse des DFG-Programms „Kompetenzmodelle“ berücksichtigt.

Dabei wird klar, dass nur eine domänenspezifische Sichtweise eine hinreichend detaillierte Beschreibung und Messung erlaubt. Allgemeinere Kompetenzkonzepte wie beispielsweise die Schlüsselqualifikationen können nicht der erste Ansatzpunkt sein. Das Modell ist allerdings nicht auf eine bestimmte Domäne beschränkt, sondern erlaubt sowohl die Beschreibung verschiedener Domänen als auch den Austausch zwischen diesen Beschreibungen.

Dies gehört ebenso zu den Anforderungen, die bei der Entwicklung des Modells im Blickpunkt standen wie die Analysierbarkeit, die Erklärungsfähigkeit der Ergebnisse, eine verständliche, sprachlich orientierte Beschreibung der Inhalte, die Wiederverwendbarkeit der dokumentierten Modellbestandteile, die Berücksichtigung psychometrischer Modelle, um eine größere Allgemeingültigkeit der Ergebnisse zu erreichen, und letztlich auch die Erzielung einer Arbeitersparnis gegenüber klassischen Klausuren.

Das Ergebnis ist das CSP-Informationsmodell mit seinen drei Ebenen CORE (Ziele, Kompetenzen, Domäneninhalte, Item-Prototypen), SITUATION (Assessments, Items, Response Types, manifeste Variablen, latente (stochastische) Variablen, Gruppierungen) und PERSONAL (individuelle Ergebnisse, ursprüngliche und typisierte Antworten der Assessment-Teilnehmer). Kompetenzen können sowohl in einem Kompetenzstrukturmodell mit verschiedenen Granularitäten und Hierarchien modelliert als auch miteinander und mit den Kompetenzniveaus von Taxonomien in einem Kompetenzniveaumodell matrixartig zu einem gemeinsamen Modell verbunden werden. Dabei werden sowohl für die Strukturen als auch die Niveaus jeweils mehrdimensionale Modelle zugelassen.

Für die konkrete Anwendung wird ergänzend zum CSP-Informationsmodell ein zielorientiertes Vorgehensmodell zur Erarbeitung der Modellinhalte, vom Domäneninhalt über die Erstellung des Assessments bis zu dessen Auswertung mit insgesamt 15 Schritten dargestellt.

Für die Beschreibung des Domäneninhaltes wird eine graphische, semiformale Darstellung vorgeschlagen, deren Elemente unmittelbar auf das Informationsmodell zurückführbar sind. Außerdem wird analog eine graphische Darstellung für die Beschreibung der Kompetenzermittlung und –messung aus den Ergebnissen des Assessments auf Basis der Domäneninhaltsbeschreibung entwickelt.

Beispielhaft wird das Modell in eine relationale Datenbank übertragen und am Beispiel von Inhalten der Mathematikdomäne „Lineares und exponentielles Wachstum“ gefüllt. Ein e-Assessment mit 76 Teilnehmern einer 11. Klasse eines Gymnasiums wird dokumentiert und ausgewertet. Dabei wird auf die Mittel der relationalen Datenbank, auf Excel und die Statistiksoftware R zurückgegriffen. Die Ergebnisse werden ebenso mit denen einer klassischen paper-and-pencil-Klausur wie mit einer statistisch mittels eines psychometrischen Modells (RASCH-Modell) ermittelten Kompetenz verglichen.

Die Gesamtanalyse der Assessment-Ergebnisse zeigt

- eine hohe Korrelation zwischen dem Score (Bepunktung) des e-Assessments und der parallelen klassischen Klausur,
- eine sehr hohe Korrelation zwischen dem Score des e-Assessments und den Traits.

Daraus lässt sich trotz der kleinen Menge von Assessment-Teilnehmern zumindest ein Hinweis auf eine ähnliche Bewertung in einer klassischen Klausur und dem e-Assessment mit dem CSP-Modell ableiten. Der Zusammenhang zwischen manifestem Score und Trait lässt es außerdem zu, die Ergebnisse des CSP-Modells auch als allgemeingültiger als die einer klassischen Klausur anzusehen.

Basierend auf der Struktur des Modells können Gruppen von Items als Variablen zusammengefasst werden, die sich auf bestimmte Kompetenzen und Inhalte beziehen, beispielsweise auf „Exponentialfunktionen“ oder „lineare Funktionen“ oder „die Bedeutung von deren Parametern“.

Damit werden individuelle Aussagen hinsichtlich des erreichten Kompetenzniveaus bezogen auf bestimmte Kompetenzen der Domäne möglich.

- Das ermöglicht in Form verschiedener Variablen eine (mehrdimensionale) Analyse bezogen auf bestimmte Inhalte und Kompetenzen sowohl mit einem manifesten Score als auch mit stochastisch geschätzten Traits.

Dabei können die stochastischen Ergebnisse auch zur Unterstützung der manifesten Bewertung genutzt werden.

- Schließlich ist auch eine Individualdiagnostik für die Lernenden und Gruppendiagnostik für die Lehrenden als Feedback möglich. Stärken und Schwächen können auf Basis der mehrdimensionalen Analyse als mögliche „Ursachen“ identifiziert werden.

Die Ergebnisse können auf Basis der verbundenen Inhalte analysiert und (hier) exemplarisch mit Ursachen und Vorschlägen zu deren Behebung verbunden werden.

- Damit besteht eine Erklärungsfähigkeit und eine Lenkungsfunktion mit der Möglichkeit sowohl dem Lernenden ein individuelles Feedback und Korrekturmöglichkeiten aufzuzeigen als auch der Lehrpersonen ein Feedback über Stärken und Schwächen der Gruppe zu geben und so Justierungen des Bildungsprozesses vorzunehmen.

Diese Analyse ist auch als reine Beobachtung, also ohne Bewertung möglich.

Bezieht man sich auf die anfängliche Frage, ob sich ein Modell für die Kompetenzbeschreibung mit einer Erklärungskomponente für das Feedback ermitteln lässt, das eine Leistungskontrolle hinsichtlich der Verfügbarkeit dieser Kompetenzen auf „classroom-level“ ermöglicht, so lässt sich dies hier in Form des CSP bestätigen.

Da das Modell sowohl formative als auch summative Aussagen basierend auf demselben Assessment und demselben Domänenmodell ermöglicht, also Beobachtung und Bewertung, ist auch diese zusätzliche Anforderung erfüllt.

Betrachtet man weiter den zeitlichen Aufwand müssen zwei Aspekte unterschieden werden. Zunächst erfordert die Nutzung des CSP-Modells Vorarbeiten bei der Dokumentation des Domäneninhaltes wie auch Mehrarbeit bei der Gestaltung des Assessments, der einige Wochen betragen kann. Diese Vorarbeit zahlt sich aber durch die erhöhte Transparenz und Wiederverwendbarkeit von Assessments als Ganzes oder von einzelnen Teilen wieder aus. Der zweite Aspekt sind erhebliche Zeiteinsparungen bei der Auswertung der Ergebnisse. Hier haben sich bereits bei einer Gruppe von 76 Personen geschätzte Einsparungen von 90% ergeben. Diese Einsparungen erhöhen oder verringern sich entsprechend der Anzahl der Assessment-Teilnehmer.

Der relativ kleine Testumfang kann keinen vollständigen Nachweis der Validität der Ergebnisse sowohl hinsichtlich der manifesten Bewertung und der geschätzten Traits als auch hinsichtlich des Vergleichs mit einer paper-and-pencil-Klausur liefern. Trotzdem geben die Korrelationen einen deutlichen Hinweis auf die Nutzbarkeit des CSP zur Messung. Letztlich ist bei den bestehenden Differenzen auch die Frage nach der Validität klassischer Klausuren hinsichtlich der Ziele des Bildungsprozesses relevant.

In jedem Fall ergibt sich bereits durch die Externalisierung der angestrebten Ergebnisse des Bildungsprozesses ein Nutzen. Die strukturierte Modellierung erlaubt außerdem eine nachvollziehbare Modellierung von Assessments mit ihren Items mit klarem Bezug zu Kompetenzen und Inhalten, die für eine Bewertung ebenso wie für eine psychometrische Analyse genutzt werden können.

Die Nutzung des CSP-Modells kann somit Lernenden wie Lehrenden bei der Justierung des Lernprozesses helfen. Darüber hinaus bietet die Externalisierung auch anderen Stakeholdern des Bildungsprozesses die Möglichkeit der Beurteilung des Erfolgs.

Die möglichen Ersparnisse beim Aufwand für Assessments werden umso größer je mehr sie für mehrere Lerngruppen eingesetzt werden, die Wiederverwendbarkeit konsequent genutzt wird oder eine Zentralisierung der Assessments, zumindest aber der Inhalte, erfolgt. Generell ist dabei außerdem eine stärkere software-technische Unterstützung und insbesondere Integration der Werkzeuge sinnvoll, die aber die Möglichkeiten dieser Arbeit übersteigt.

Das Modell sollte auch in anderen Domänen getestet werden, die für eine strukturierte Beschreibung weniger geeignet erschienen als die Mathematik, was aber nicht im Bereich dieser Arbeit liegt.

10 Ausblick

10.1 Weitere Aufgaben und Domänen

Die Anwendung des Modells wird mit einer Domäne aus dem Bereich Mathematik und einem e-Assessment untersucht. Dabei werden viele geschlossene Aufgabenstellungen verwendet, also solche Aufgaben bei denen der Startpunkt, der Lösungsweg und die Antwortalternativen weitgehend vorgegeben sind. Der Weg, um zu einem Ziel in Form einer Lösung zu gelangen, ist also weitgehend geplant, kann unter Umständen aber mehrere Schritte beinhalten. Es gibt verschiedene Items, die schrittweise hinsichtlich der Vorgabe der Antwortalternativen geöffnet sind.

Neben dem schrittweisen Öffnen der Antwortalternativen können offenere Aufgabenstellungen entwickelt werden, die aus dem Bereich des Starts der Problemlösung, des Weges und dem Ziel dorthin, verschiedene Bereiche vorgeben. So können ein Startpunkt und eine Lösung vorgegeben sein und eine Begründung des Weges kann erfragt werden. Oder eine Anfangssituation wird aus einem Weg und einem Ergebnis abgeleitet. Letztlich kann auch aus einer offenen Situation eine Problemstellung erarbeitet werden. Verschiedene Klassifikationen findet man beispielsweise bei Büchter und Leuders. (Büchter und Leuders 2011, 92ff.).

Einen besonderen Aspekt stellen polyvalente Aufgaben dar, bei denen zu einer gegebenen Problematik zumindest eine grundlegende Lösung und weitere höherwertige Antworten existieren. (Hellmig und andere 2015).

Das CSP bietet verschiedene Möglichkeiten, diesen Aufgabentypen Rechnung zu tragen. So können die einzelnen Antwortalternativen in Form von Response Types unterschiedlich bewertet werden. Die Bewertung wie sie in Abschnitt 6.3.2. beschrieben wird, lässt ausdrücklich unterschiedliche Scores für verschiedene Alternativen, insbesondere auch verschiedene als richtig eingestufte, aber unter Umständen unterschiedlich bewertete Lösungen zu. Auch nicht antizipierte Lösungen können ergänzt und bewertet werden. Testweise werden auch in dem hier durchgeführten Assessment zwei Items zur Beschreibung von Reihenfolgen vorgegebener Prozessschritte verwendet, die allerdings, auch wegen ihrer inhaltlichen Ausrichtung, nicht gewertet werden.

Sollte die unterschiedliche Bewertung wegen der Verwendung bestimmter stochastischer Modelle zu Problemen führen, kann ein Item entsprechend den Antworten in mehrere Items aufgespalten werden und die einzelnen Items entsprechend den enthaltenen Kompetenzen in getrennten stochastischen Modellen analysiert werden. Wegen der inneren Abhängigkeit können aber nicht mehrere dieser Items in einem stochastischen Modell verwendet werden.

Das hier verwendete Assessment enthält nur einen kleinen Anteil etwas offenerer Fragen. Bei anderen, offeneren Formen des Assessments ist unter Umständen ein Vorverarbeitungsschritt notwendig. So musste bereits bei versuchsweise verwendeten prozessorientierten Items die Struktur des Prozesses und der damit verbundene Zielerreichungsgrad festgelegt werden. Generell öffnet eine Vorverarbeitung den möglichen Weg einer Integration derartiger Items und Assessments.

So ergibt sich eine derartige Thematik bei der Berücksichtigung anderer Fächer insbesondere im gesellschaftswissenschaftlichen und sprachlich-künstlerischen Bereich. Hier wird man bei der Bewertung zwischen dem formalisierbaren und dem nicht formalisierbaren, eher ästhetischen

Bereich unterscheiden müssen. Der ästhetische Bereich liegt aus jetziger Sicht außerhalb der Reichweite eines Systems wie dem CSP.

Betrachtet man den formalisierbaren Bereich, so kann man im Fach Deutsch auf die Kriterien für die Bewertung von Prüfungsleitungen zurückgreifen. Für die sachliche Richtigkeit sind beispielsweise vorgesehen (Kultusministerkonferenz 2002, 27):

- Folgerichtigkeit und Begründetheit der Aussagen,
- Vielfalt der Gesichtspunkte und ihre jeweilige Bedeutsamkeit,
- Differenziertheit des Verstehens und Darstellens,
- Herstellung geeigneter Zusammenhänge,
- Grad der Selbstständigkeit,
- Klarheit in Aufbau und Sprache,
- Sicherheit im Umgang mit der Fachsprache und -methode und
- Erfüllung standardsprachlicher Normen.

Die automatische Prüfung der Erfüllung standardsprachlicher Normen scheint am leichtesten zu automatisieren zu sein. Hier könnten Aspekte wie Grammatik, Sprachumfang, Wortwahl, und Satzlänge in Metriken ermittelt werden. Für diese, aber insbesondere auch andere Aspekte wäre wie geschildert eine Vorverarbeitung denkbar, die die relevanten Informationen aus einem Text in strukturierter Form extrahiert. Auch wenn dies im Moment noch schwer erkennbar ist, gibt es im Bereich der Qualitative Content Analysis viele Ansätze, beispielsweise in einem Überblick bei Hsieh und Shannon (Hsieh und Shannon 2015).

Dabei muss ein umfangreicher Textanalyseprozess erfolgen, der die relevanten RESPONSE TYPEs ermittelt. Derartige Systeme sind nicht vollkommene Utopie, sondern es gibt durchaus ernst zu nehmende Ansätze¹³. Die weitere Verfolgung der Möglichkeiten derartiger vorgelagerter Messprozesse ist nicht Gegenstand dieser Arbeit.

Schließlich kann man auch ohne derartige Automatisierungstechniken zumindest Teile des CSP nutzen, indem manuell ermittelte und vorverdichtete Ergebnisse beispielsweise mittels der in Abschnitt 7.8.2.1 beschriebenen CNMs für die Kompetenzdokumentation genutzt werden. In diesem Zusammenhang wird dort auf das IQB hingewiesen, dass auch für das Fach Deutsch Beschreibungen für Kompetenzstufen angegeben hat.

10.2 Weitere Teilmodelle

10.2.1 Qualifikationen

Das CSP-Modell kann neben der Kompetenz auch die Qualifikation beschreiben, die sich laut Definition beide aus den Fähigkeiten herleiten lassen. Daher gibt es für das CSP einen Entwurf für ein Teilmodell für diesen Bereich. Bereits in Kapitel 2 wurden Qualifikationen als Kompetenz- und Fähigkeitsbündel, die als Anforderung für die Bewältigung von Prozessen in Handlungsräumen (parametrisierte Situationen) definiert werden, beschrieben. Qualifikationen werden daher hier als Zusammenfassung beobachtbarer Fähigkeiten und messbarer Kompetenzen angesehen.

¹³ Es existieren auch softwaregestützte Ansätze, beispielsweise (Online 2010).

Ein Beispiel für die Definition von Qualifikationszielen über Fähigkeiten kann der Erläuterung der Universität Bayreuth mit den Qualifikationszielen eines Bachelor-Studiengangs entnommen werden.

Beispiel Bachelor-Qualifikationsziele

Qualifikationsziele (deutsch):

Nach dem erfolgreichen Absolvieren dieses Kurses sollten die Studierenden in der Lage sein

- Die Bedeutung der Planung in Unternehmen zu erläutern und die wesentlichen Schritte eines Planungsprozesses zu benennen
- Die wesentlichen Bestandteile einer Entscheidungssituation zu beschreiben und zu erläutern
- Methoden für Entscheidungsprobleme unter mehrfacher Zielsetzung, unter Ungewissheit sowie unter Risiko zu erklären und zur Lösung von entsprechenden Problemen anzuwenden
- Ausgewählte Methoden der Projektplanung zu erläutern und anzuwenden
- Grundlegende Methoden der Finanzmathematik auf Investitions- und Finanzierungsprobleme anzuwenden
- Die Grundlagen der Buchhaltung, der Bilanzierung und der Kostenrechnung zu erläutern und Methoden aus diesen Bereichen auf einfache Problemstellungen anzuwenden
- Die Bedeutung des Controllings im Unternehmen zu erläutern
- Ausgewählte Instrumente des Controllings auf Planungssituationen anzuwenden

(Bayreuth 2014)

Die Formulierung „Planung ist im Unternehmen von Bedeutung“ ist im Sinne des CSP ein **FACT**. Der **OPERATOR** „erläutern“ steht in Verbindung zu dieser Aussage, da diese Aussage erläutert werden soll. Hieraus ergibt sich dann eine konkrete Fähigkeit. Im obigen Beispiel bezieht man sich in der Erläuterung auf die Bloomsche Taxonomie. In der Bloomschen Taxonomie steht diese Aussage entsprechend dem Operator auf dem Niveau 2 „Verständnis“ und wäre im Modell entsprechend mit dem Qualifikationsniveau zu verbinden. Die Fähigkeit wird außerdem dem Ziel des hier als Beispiel angegebenen Bachelorstudiengangs zugeordnet. Dies entspricht der ausdrücklich als Qualifikationsziel benannten Logik. Analog kann mit den anderen angegebenen Zielen verfahren werden.

Dies soll nur beispielhaft einen Ansatz für die Modellierung im CSP darstellen, um den Zusammenhang zum Modell aufzuzeigen und soll nicht weiter vertieft werden.

Entsprechend der Definition werden die Kompetenzen und die Skills gebündelt und letztlich zu verschiedenen Arten von **QUALIFICATION BUNDLES** zusammengefasst. Die **COMPETENCE BUNDLES** werden aus den spezifischen Kompetenzen auf einem bestimmten Niveau gebildet. Das könnten beispielsweise die Erkenntnisse über den Unterschied zwischen linearem Wachstum und exponentiellem Wachstum aus dem Anforderungsbereich II sein. Dies kann mit **SKILL BUNDLES** kombiniert werden, beispielsweise der Fähigkeit einen Funktionswert einer Exponentialfunktion zu bestimmen. Zusammen mit anderen Kompetenzen und Fähigkeiten kommt man so zu **QUALIFICATION BUNDLES**, die letztlich eine Person befähigen, einen bestimmten

Prozessschritt auszuführen, also beispielsweise in einer Situation zu entscheiden, dass es sich um exponentielles Wachstum handelt und einen Prognosewert aus der entsprechenden Funktion zu bestimmen.

Daher werden die **QUALIFICATION BUNDLES** mit einem Prozessschritt(**STEP**) verbunden und so die Beziehung zur Prozesssicht hergestellt. Die entsprechenden Kompetenzen unterstützen die Durchführung des Prozessschrittes.

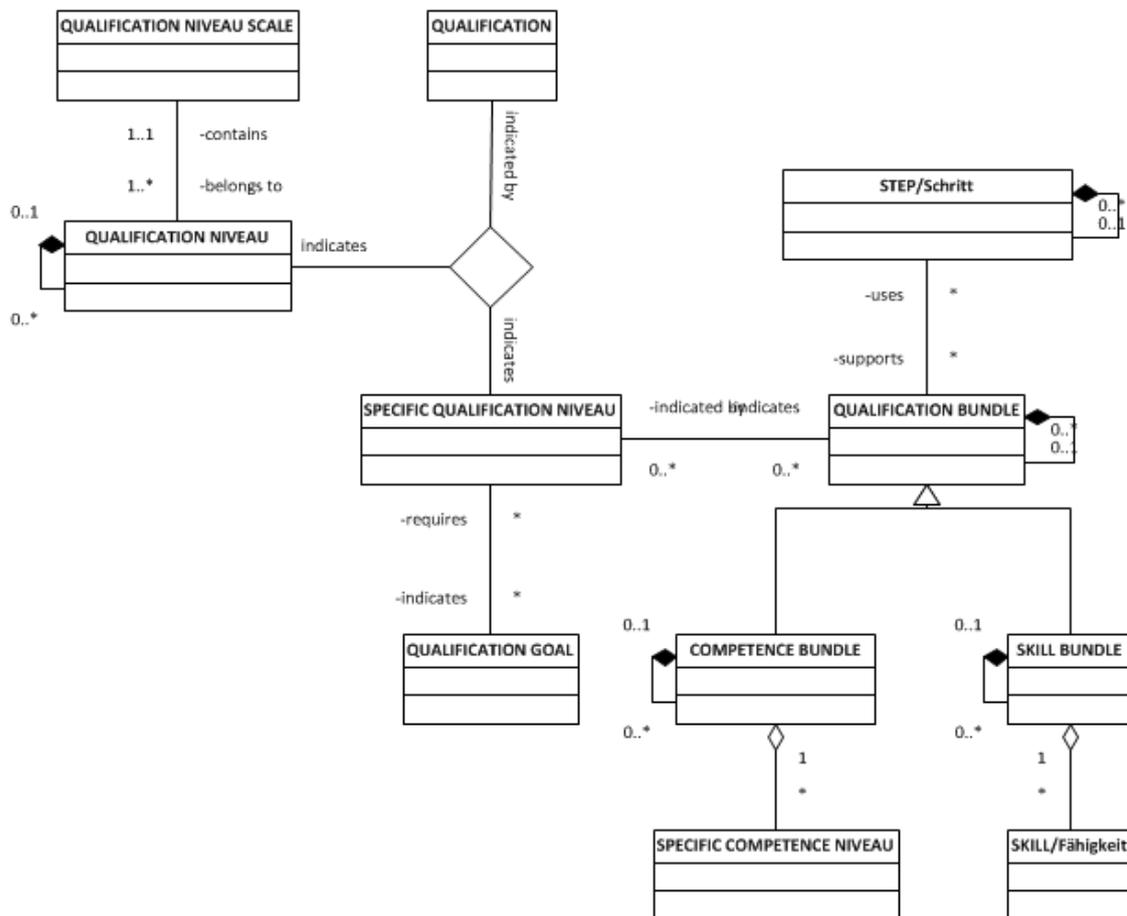


Abbildung 77 Bereich Qualification

Die andere Sichtweise ergibt sich aus der formalen Definition von Qualifikationen. So definiert der Europäische Qualifikationsrahmen für lebenslanges Lernen (EQR) (European Qualifications Framework (EQF)) (European Commission 2015) acht Qualifikationsniveaus, die durch ihre Kenntnisse, Fertigkeiten und Kompetenzen beschrieben sind. Dabei wird nicht direkt auf die formale Art des Erwerbs eines bestimmten Niveaus eingegangen, um nicht bestimmte Bildungssysteme zu bevorzugen oder zu benachteiligen. Aus diesem Europäischen Qualifikationsrahmen sind dann nationale Qualifikationsrahmen hervorgegangen, so auch der Deutsche Qualifikationsrahmen (DQR) (BMBF, DQR - Deutscher Qualifikationsrahmen 2011), der konkrete auch formale Abschlüsse auf die Qualifikationsniveaus des EQR bezieht.

Diese Qualifikationsrahmen können im Rahmen des CSP als **QUALIFICATION NIVEAU SCALES** abgebildet werden und die entsprechenden Niveaus als **QUALIFICATION NIVEAUS**. So können konkrete formale oder informelle **QUALIFICATION**s einem solchen Niveau zugeordnet und

durch eine Summe von **SPECIFIC QUALIFICATION NIVEAUs**, die letztlich die Einzelqualifikationen mit ihrem zugehörigen Niveau beschreiben, abgebildet werden.

Diese **SPECIFIC QUALIFICATION NIVEAUs** ihrerseits können zu den in formalen Anforderungen, beispielsweise der Industrie- und Handelskammern für einen bestimmten Berufsabschluss oder einer Universität wie der oben erwähnten Universität Bayreuth für einen bestimmten Bachelor-Abschluss, definierten **QUALIFICATION GOALS** in Bezug gesetzt werden.

Andererseits können sie durch konkrete **QUALIFICATION BUNDLES** weiter spezifiziert werden. Hier kommt die Möglichkeit hinzu, dass es sich bei den **QUALIFICATION BUNDLES** nicht nur um beobachtbare Kompetenzen und messbare Fähigkeiten handelt, sondern diese nach erfolgten Assessments auch als beobachtete Kompetenzen und gemessene Fähigkeiten beschrieben werden können und daher für Aussagen über die tatsächlich für die Durchführung eines Prozesses hilfreichen Kompetenzen genutzt werden könnten.

10.2.2 Competence Intensive Prozess (CIP)

Der Bereich des prozeduralen Wissens kann ebenso wie die Competence Intensive Processes, die noch definiert werden, oder die Situation-Oriented-Processes des SITUATION-Modells als Prozesse verfeinert und modelliert werden. Der Bereich Procedure, wie in Abbildung 31 beispielhaft modelliert, stellt die Ausgestaltung des Modells für einen Prozess dar. Diese Möglichkeit einer klassischen Prozessbeschreibung wird für Competence Intensive Processes (CIP) erweitert.

Der Begriff Competence Intensive Process (CIP) beschreibt einen Prozess, der erst durch den Einsatz von Kompetenzen und Qualifikationen zu vorgegebenen Zielen geführt werden kann, wobei diese dessen Nutzen beschreiben und bei der Beschreibung des Prozesses mitmodelliert werden.

Im Folgenden sollen also Prozesse betrachtet werden, von denen angenommen wird, dass es sich um kompetenzintensive Prozesse handelt, die im Rahmen eines komplexen Systems einen messbaren Nutzen erfüllen, also um in der realen Welt von einem System, oder in einem beruflichen Umfeld, auszuführende Prozesse.

Der Grundgedanke bei der Ausführung des CIP ist die Erreichung eines oder mehrerer Ziele in einem messbaren Umfang. Ein Ziel kann dabei wiederum aus einer externen Spezifikation abgeleitet werden. Daher wurde der Typ des **UBIQUITY GOAL** in das Modell aufgenommen. Ein solches **UBIQUITY GOAL** kann ganz oder teilweise als **TARGET VARIABLE** in die Zielvorstellung des Prozesses übernommen werden. Diese **TARGET VARIABLES** können zu **TARGET BUNDLES** zusammengefasst werden, was nicht mehr bedeutet, als die einzelnen Ziele zu einem Vektor aus Einzelkomponenten zusammenzufassen. Dieses **TARGET BUNDLE** beschreibt die Erfüllung der Ziele eines **COMPETENCE INTENSIVE PROCESS (CIP)**.

Ein **CIP** wird wie jeder Prozess durch eine Reihe von aufeinanderfolgenden Schritten realisiert. Geht man davon aus, dass ein kompetenzintensiver Prozess wegen seiner Komplexität und der innovativen Elemente aus Sicht der Prozesssubjekte nicht vollständig in seiner Struktur vorherbestimmt ist, so liegt es nahe, ihn primär durch das mit dem Prozess verfolgte Ziel, die Aufgabe des Prozesses, zu beschreiben. So können Teile oder die gesamte Prozessstruktur zunächst als unbekannt angesehen werden.

Die strukturelle Gestaltung der **CIPs** kann daher in unterschiedlichen Freiheitsgraden sinnvoll sein. Dies kann von einer bekannten nur in kleinen Bereichen mit Lücken versehenen Struktur über einen Zustand mit bekannten **STEPs** aber unbekannter Struktur, die Kombination bekannter **STEPs** bis zur kompletten Neukonstruktion von Prozessen reichen.

Die Bausteine der Prozessstruktur sollen weiterhin als **STEPs**, also als Schritte, bezeichnet werden. Ein **CIP** besteht aus einem oder mehreren **STEPs**, deren Struktur nicht vor der Ausführung bekannt sein muss, die nicht einmal alle komplett bekannt sein müssen. Jeder definierte **STEP** wird auf einen **SIMPLE PROCESS** zurückgeführt, wie er bereits im Rahmen des prozeduralen Wissens mit dem Typ **PROCEDURE** verbunden worden ist. Das entsprechende, mit diesem **SIMPLE PROCESS** verknüpfte Modell bleibt unverändert.

Ein kompetenzorientierter Prozess soll also nicht vollständig festgelegt sein, sondern wird durch zielorientierte Entscheidungen während seiner Ausführung zumindest teilweise erst entstehen oder konstruiert werden.

Dies bedeutet für den Ausführenden nicht nur die Wahlmöglichkeit bei antizipierten Verzweigungen, sondern die Möglichkeit der strukturellen Gestaltung des Prozesses während seiner Durchführung.¹⁴ Dies ist nicht nur bei ganz oder teilweise unbekanntem Prozessen notwendig, sondern kann auch generell bei der Beurteilung des Erfolgs eines Prozesses hilfreich sein. Denn eine solche strukturelle Gestaltung setzt die Verfügbarkeit von Kriterien für Entscheidungen bei der Gestaltung voraus. Solche Kriterien können gerade in Form von Zielen des Prozesses vorgegeben sein. Dies bietet gleichzeitig die Möglichkeit die Ziele in die Prozessgestaltung zu integrieren, was ein oft vernachlässigter Aspekt ist (Soffer und Wand 2005, 663). Das CSP erlaubt mit einer solchen Erweiterung dann die Beschreibung der Ziele bevor der Prozess komplett bekannt sein muss. Für das CSP ist der Prozessgedanke also einerseits eine Folge der externen Zielvorgabe, andererseits eine Möglichkeit, Prozesse aufgrund der Ergebnisse von Assessments zu strukturieren und Aussagen über die verwendeten **QUALIFICATION BUNDLES** zu machen.

Bezieht man dies umgekehrt wieder auf Assessments, so können vordefinierte Prozesse bei der Gestaltung von Assessments als Leitfaden dienen, andererseits können in Assessments beobachtete Prozesse zur verallgemeinerten Formulierung eines **CIP** dienen. Außerdem können in Assessments beobachtete und gemessene Kompetenzen und Fähigkeiten zu den durchgeführten **STEPs** und deren Struktur in Bezug gesetzt werden.

Die Dokumentation eines **CIP** besteht somit aus den bekannten Elementen. Dies sind zunächst die **TARGET VARIABLES** gebündelt zum **TARGET BUNDLE**. Hinzu kommen die **STEPs** soweit diese bereits bekannt sind. Jeder **STEP** beginnt mit einem zuvor erreichten **STATE** und endet in einem danach gültigen **STATE**. Schließlich sind die **QUALIFICATION BUNDLES** anzugeben.

Diese Beschreibungen können noch etwas formalisiert werden. Generell soll gelten, dass ein **CIP** definiert ist als

$CIP = (O, S, P, K)$

¹⁴ Eine solche Möglichkeit bietet beispielsweise die BPMN in Form der ad-hoc-Prozesse ebenfalls (Allweyer 2009, 97).

O = Zielerreichungsvektor, der durch die Sollwerte der Ziele (O_i) im **TARGET BUNDLE** beschrieben ist. Die Ziele, die durch die Ausführung des Prozesses erreicht werden sollen, werden im Zielerreichungsvektor gemessen. Der Zielerreichungsvektor gibt also Auskunft darüber in welchem Umfang die einzelnen Ziele des Prozesses erreicht werden und setzt sich aus den einzelnen Zielvariablen zusammen, die den einzelnen Prozesszielen zugeordnet sind.

S = Menge von Statusvariablen, die zusammen den Zustandsraum bilden und als Statusvektor (S_i) in jedem **STATE** des **CIP** bestimmte Istwerte haben.

P = Menge gegebener oder entwickelter **STEPS** (auf atomarer Ebene, die den CIP repräsentierenden Elementarprozesse) und die als Teil der Problemlösung zur Verfügung stehen.

K = Benötigte **QUALIFICATION BUNDLES** als Menge der benötigten Kompetenzen und Fähigkeiten.

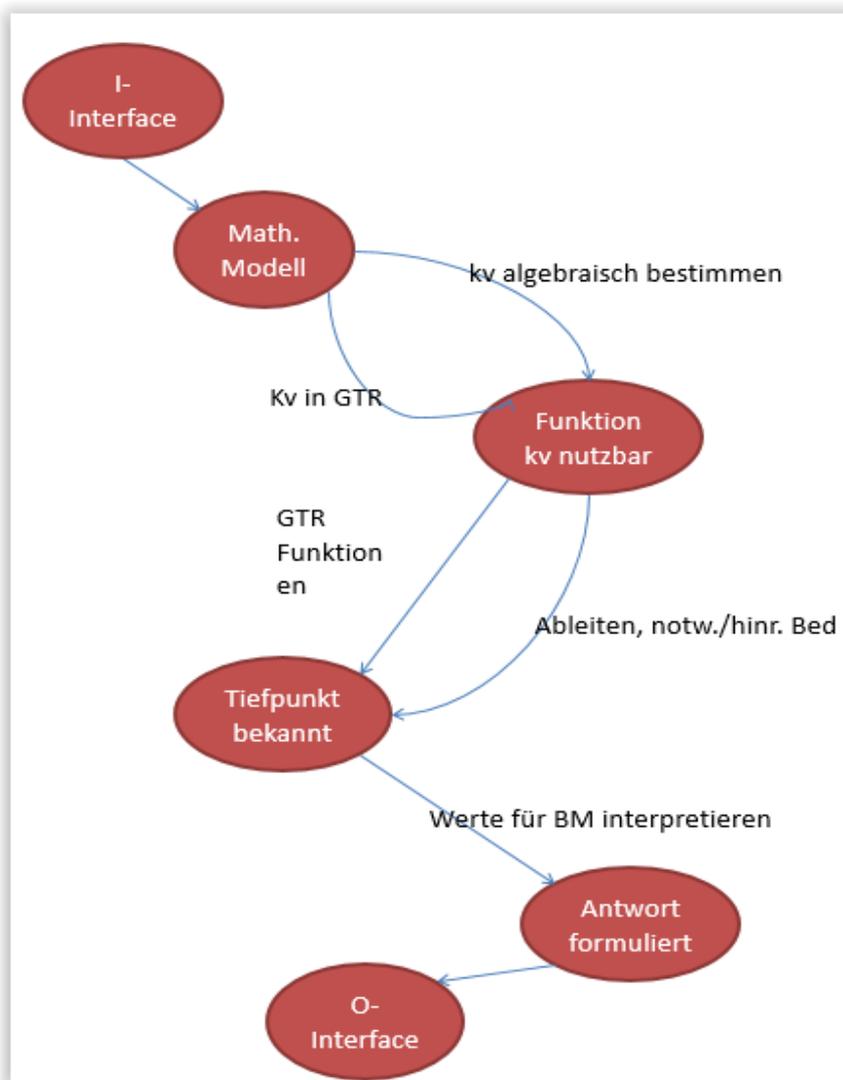


Abbildung 78 Struktur eines CIP mit STEPs und STATES

Die Auswahl und Reihenfolge der **STEPS** und **STATES** formen den tatsächlich ausgeführten **CIP**. In Abbildung 78 ist beispielhaft diese Struktur angegeben. Es geht inhaltlich dabei darum, das Betriebsminimum (BM) aus einer Kostenfunktion zu bestimmen. Dazu kann die Funktion der

Eine zweite Erweiterung der Prozesse stellt die hier untergeordnete Ebene der bereits angesprochenen **QUALIFICATION BUNDLES** dar. Für einen kompetenzorientierten Prozess stellt die Modellierung der Qualifikationen einen wesentlichen Aspekt bei der Formulierung des Prozesses dar. Hier kann man auch wissensintensive Prozesse betrachten. In diesem Bereich der wissensintensiven Prozesse findet man Ansätze, um das benötigte Wissen einschließlich der Konversionen zwischen stillschweigendem (tacit) Wissen und explizitem Wissen darzustellen. Explizites Wissen kann unmittelbar im Prozess genutzt werden. Stillschweigendes Wissen ist nur indirekt beobachtbar. Gerade die Berücksichtigung von stillschweigendem Wissen und dessen Nutzung bei der Prozessdurchführung stellt aber einen wesentlichen Aspekt wissensintensiver Prozesse dar. Daher ist die Erweiterung um eine Ebene von Wissensobjekten notwendig, wie sie beispielsweise die bereits erwähnte KMDL bietet. Sprachen wie die KMDL bieten die Möglichkeit der Modellierung von Informationsobjekten (explizit), beziehungsweise Wissensobjekten (tacit) und deren Übergang. Damit wird eine wichtige Brücke im Übergang zwischen Prozessbeschreibung und Wissensbeschreibung geschaffen. Problematisch bleibt dann, wie das Wissen in den Wissensobjekten modelliert werden kann und wie es zu den Kompetenzen im Zusammenhang steht.

Das „Wissen“ ist im CSP in den **FACTS**, **SKILLS** und untergeordneten Typen lokalisiert. Es könnte über die **QUALIFICATION BUNDLES** zusammengefasst und so den **STEPS** zugeordnet werden.

Der Status entsteht durch eine reine (in der Praxis oftmals subjektive) Messung. Der Status stellt aber nicht eine einzige Punktzahl dar, sondern bewertet jede der Statusvariablen, so dass im **TARGET BUNDLE** die Zielerreichung differenziert beschrieben werden kann.

So ist es neben einer Bewertung des erreichten Endzustandes über die Rollen und Fakten möglich, Rückschlüsse auf die Beherrschung dieser Fakten zu ziehen. Von dort kann die Folgerung wiederum auf die Lernziele¹⁵ und Kompetenzen übertragen werden. Zu beachten ist, dass die Logik hier stets über einen Soll-Ist-Vergleich der Statusvariablen erfolgt.

10.2.3 Idee eines Modellbereichs Situation Process Model

Der Prozessgedanke lässt sich auch auf das **SITUATION**-Modell und damit die Gestaltung von Assessments übertragen. Die Idee ist, Aufgabenstellungen zu modellieren, die mehr als einen Schritt umfassen und diese Schritte bewusst von den Assessment-Teilnehmern im Sinne eines CIP dynamisch strukturieren zu lassen. Das Lösen der Aufgabe kann auch als das Durchführen eines Prozesses beschrieben werden und der Assessment-Teilnehmer trifft insbesondere eigenständige Entscheidungen über die auszuführenden Schritte und deren Reihenfolge, die ausdrücklich in die Bewertung aufgenommen werden.

So kann man eine **EXERCISE** und deren Bewertung so formulieren, dass deutlich wird „[...] welche Teilkompetenzen für die erfolgreiche Bearbeitung erforderlich sind und wie für entsprechende Teillösungen eine Teilhonorierung erfolgt.“ (Büchter und Leuders 2011, 182). Dies beschreibt die Situation, dass in Aufgaben bestimmte Bündel von Anforderungen konstruiert werden. Dabei werden die Anforderungsniveaus zumeist differenziert (Büchter und Leuders 2011, 104). Im Modell bedeutet dies, dass den **ASSESSMENT ITEMS** und deren Gruppierungen entsprechende Typen der Prozessmodellierung gegenübergestellt werden, die die einzelnen

¹⁵ Wie auch von Schott und Ghanbari gefordert, (Schott und Ghanbari 2008, 75).

Schritte und Entscheidungen umfassen. Dann können letztlich den Teillösungen aus der Prozessbearbeitung **MANIFEST VARIABLES** und **ASSESSMENT ITEMS** zugeordnet werden.

Dabei erscheint es sinnvoll, die Aufgabe, die den gesamten Prozess repräsentiert, diesem zuzuordnen. Dies geschieht auch hier als **CIP**. Die Gliederung eines **CIP** entspricht exakt der, die bereits als Idee beschrieben wurde. Die einzelne Aktivität als kleinste Einheit wird dem Item zugeordnet. Die entsprechenden Gruppierungen dazwischen repräsentieren jeweils die Struktur im anderen Teil. Die Ziele des Prozesses werden im **TARGET BUNDLE** beschrieben. Sie werden herausgegriffen, um die Modellierung zu verdeutlichen. Die Ziele werden in einzelne **TARGET VARIABLES** zerlegt, die die einzelnen Ziele des Prozesses beschreiben und auch den jeweils erreichten Erfolg messen. Diese Erfolgsmessung geschieht jeweils bezogen auf einen ausgeführten **STEP** des Prozesses mittels Auswertung des erreichten **STATES**. Die mit den **TARGET VARIABLES** gemessenen Ziele können dann über eine Verbindung zur **MANIFEST VARIABLE** in die Gesamtmessung der **EXERCISE** integriert werden.

Ein **STEP** im Sinne des Modells entspricht dem Übergang von einem definierten messbaren Zustand hinsichtlich der Zielerreichung des Gesamtprozesses in einen anderen messbaren Zustand. Die **TARGET VARIABLES** können somit auch die Strategie messen, die im Rahmen des Gesamtprozesses für die Bewältigung und Sequentialisierung der **STEPS** verwendet wird. Die Wahl der Strategie zur Erreichung und damit letztlich die Ausgestaltung dieses Schrittes und die Auswahl des Elementarprozesses kann dem Lernenden überlassen bleiben. Diese Strategie kann im einen Extremfall entweder durch die Aufgabe komplett festgelegt sein oder im anderen Extremfall komplett neu entstehen.

10.3 Sozio-technische Systeme

Die Nutzung für derartige Konzepte erscheint möglich, da strukturelle Parallelen gegeben sind. So ist es gerade Ziel von Bildungsprozessen, Kompetenzen zur Handlung in realen Situationen zu vermitteln. Dementsprechend handelt es sich bei der Gestaltung von Assessments um eine Art „Simulation“ realer Handlungssituationen, so dass deren Messung analog zur Messung dieser realen Situationen gesehen werden kann, auch wenn in der Praxis sicherlich viele Hindernisse zu überwinden und Kompromisse einzugehen sind.

Die Bewertung und Beurteilung von Kompetenzen setzt die Beobachtbarkeit voraus. Beobachtung bedeutet, die Handlung einer Person in irgendeiner Form wahrnehmen zu können. Das wiederum erfordert eine Abfolge von Aktivitäten der Person, oder anders ausgedrückt, die Ausführung eines Prozesses, bei der die Fähigkeiten (Skills) der Person sichtbar werden.

Mit der geschilderten Erweiterung des Modells um CIPs kann jede Problemlösung als Ausführung eines Prozesses verstanden werden. Die für den Erfolg eines solchen Prozesses relevanten Fragen können sein:

- Welche Ziele verfolgt ein Anwender in einer bestimmten Situation?
- Wodurch ist der Erfolg bei der Erreichung dieser Ziele im Einzelnen messbar?
- Welche Schritte führt er in welcher Reihenfolge aus?
- Wie tragen die einzelnen Schritte zum Gesamterfolg bei?

Beispiele solcher Prozesse sind:

- 1) Ein Anwender möchte mit seinem neuen Handy zu einem bisher unbekanntem Ziel navigieren, das ihm ein Freund ebenfalls auf dieses Handy noch mitteilen möchte.
- 2) Ein Sanitärinstallateur erhält den Auftrag, ein Bad in einem vorgegebenen Stil zu gestalten, das eine Dusche, eine große Badewanne, Waschbecken und Toilette beinhaltet.
- 3) Ein Schüler erhält eine zweistündige Mathematiklausur, die er mittels Papier, Bleistift, Geodreieck und Taschenrechner bearbeiten soll.

Betrachtet man die aufgezählten Prozesse, so erfordern sie von den Ausführenden jeweils einige „Vorkenntnisse“, also Wissen oder Kompetenzen, die für diesen Prozess relevant sind. Die Frage ist, welche Vorkenntnisse zum erfolgreichen Bearbeiten des Prozesses hilfreich sind. Fragestellungen hierzu können etwa sein:

- Welche Informationen und welches Wissen sind geeignet, zum Erfolg beizutragen?
- Welche Fähigkeiten, Kompetenzen und Qualifikationen sind geeignet zum Erfolg beizutragen?

Mit der Erweiterung um CIPs entstehen genau diese Zusammenhänge im CSP-Modell. Wird die Aufgabe dann als Assessment dokumentiert, könnte es das CSP-Modell ermöglichen, auch diese Zusammenhänge darzustellen und zu messen.

A Anhang

Hier werden Detailergebnisse der statistischen Analysen zur Ermittlung der Traits erläutert.

A.1 Vorgehensweise bei der statistischen Analyse

Zur Vorbereitung der Analyse werden die Daten des Assessments in der Excel-Datei M11V1R2 genutzt. Diese stammen aus dem Download in Excel-Format der Erfassung mit SoSciSurvey.de. Sie wurden in ein Datenblatt gleichen Namens im CSV-Format übertragen.

Für die Analyse mit dem RASCH-Modell wird die Software R in der Version 3.01 mit dem Erweiterungspaket eRM verwendet. Die entsprechenden Befehle in R sind im Folgenden angegeben.

```
(1) M11V2 <-  
    read.table("C:/Rostock/Umfrage/Ergebnisse/komplett/M11V1R2.csv",  
    sep = ";", header = TRUE, row.names = 1)  
(2) ncol(M11V2)  
(3) nrow(M11V2)  
(4) rm1 <- RM(M11V2)  
(5) rm1$betapar  
(6) pp2 <- person.parameter(rm1)  
(7) pp2  
(8) summary(pp2)  
(9) lr2 <- LRtest(rm1, splitcr="mean", se=TRUE)  
(10) lr2  
(11) plotGOF(lr2, beta.subset="all", conf=list(gamma=0.95, col=1))  
(12) plotjointICC(rm1)  
(13) Waldtest(rm1,splitcr="mean")  
(14) beta <- data.frame(colnames(M11V2),rm1$betapar)  
(15) write.table(beta,  
    "C:/Rostock/Umfrage/Ergebnisse/komplett/M11V2betafertig.csv",  
    sep=";", dec = ",", col.names=TRUE, row.names=TRUE, quote=TRUE,  
    na="NA")  
(16) theta <- data.frame(rownames(M11V2),pp2$thetapar)  
(17) write.table(theta,  
    "C:/Rostock/Umfrage/Ergebnisse/komplett/M11V2thetafertig.csv",  
    sep=";", dec = ",", col.names=TRUE, row.names=TRUE, quote=TRUE,  
    na="NA")  
(18) library(relimp, pos=4)  
(19) library(eRm, pos=4)
```

Listing 1 R-Script für die statistische Analyse

Die einzelnen Befehle des Listing 1 sollen kurz erläutert werden. Die Befehle je Zeile sind nicht zwingend in der Reihenfolge auszuführen, sondern können einzeln gestartet werden und es wird jeweils nur die Zeile ausgeführt. Sie bauen jedoch teilweise aufeinander auf. Die Pfadangaben sind rechnerspezifisch.

Die Daten werden zunächst aus den Basisdaten des Assessments geladen. Die Datei M11V1R2.csv enthält die Ergebnisse als dichotome Werte (0 = falsch, 1 = richtig) mit Semikolon getrennt. Jede Zeile entspricht einer Person, die Reihenfolge der Items ist festgelegt und die Namen stehen in einer Kopfzeile. Die komplette Datei steht als Matrix in M11V2 (1). Die Anzahl der Zeilen und

Spalten kann zur Kontrolle ermittelt werden (2,3). Mit dem Befehl RM werden die Daten entsprechend dem RASCH-Model analysiert und die kompletten Analyseergebnisse in rm1 gespeichert (4). Die geschätzten Schwierigkeiten pro Item können getrennt ausgegeben werden (5). Die Kompetenzwertschätzung muss getrennt vorbereitet werden und wird in pp2 abgelegt (6). Die Kompetenzwerte (θ_i) können jetzt angezeigt werden (7), wie auch weitere Parameter (8). Nach einer Modellerstellung kann jeweils ein LR-Test (10) und ein Wald-Test (13) durchgeführt werden. Ebenso kann die Anzeige für den grafischen Test erstellt (11) und die ICC-Kurven dargestellt werden (12). Die übrigen Befehle dienen der Erstellung von Ausgaben.

A.2 Erläuterungen zur ersten statistischen Analyse

Die erste Analyse basiert auf den Gesamtdaten des Assessments soweit sie die betrachtete Domäne betreffen. Im Vorfeld werden also nur die Items, die das Wachstum und Exponentialfunktionen betreffen verwendet (W, EX und EL). Der Ablauf besteht nach dem Laden in der Erzeugung des RASCH-Schätzmodells, dem Check mit dem LR-Test und einem Wald-Test.

Wald test on item level (z-values):								
Item value	z-statistic	p-value	Item value	z-statistic	p-value	Item value	z-statistic	p-value
beta EX02_02	1.006	0.315	beta EX02_03	-1.026	0.305	beta EX02_04	0.084	0.933
beta EX03_05	0.364	0.716	beta EX03_11	1.117	0.264	beta EX07_07	0.588	0.557
beta EX07_08	0.290	0.772	beta EX07_02	1.284	0.199	beta EX07_10	-0.523	0.601
beta EX04_01	1.086	0.278	beta EX04_02	1.066	0.286	beta EX05_01	0.598	0.550
beta EX05_02	1.066	0.286	beta EX06_03	0.593	0.553	beta EX08_01	1.139	0.255
beta EX09_01	0.290	0.772	beta EX10_01	0.281	0.779	beta EX10_02	-0.434	0.664
beta EX10_04	1.363	0.173	beta EX11_02	1.020	0.308	beta EX11_04	-1.454	0.146
beta EX11_05	-0.258	0.796	beta EX11_06	0.759	0.448	beta EX11_07	-1.122	0.262
beta EX12_01	-0.259	0.796	beta EX12_04	0.140	0.889	beta EX12_05	-0.211	0.833
beta EX12_06	0.475	0.635	beta EX12_07	-0.211	0.833	beta EX13_02	-0.149	0.882
beta EX13_04	1.609	0.108	beta EX13_05	0.589	0.556	beta EX14_01	1.037	0.300
beta EX14_02	0.292	0.770	beta EX15_01	0.521	0.602	beta W101_01	-0.593	0.553
beta W101_02	0.520	0.603	beta W101_03	-0.739	0.460	beta W101_04	0.434	0.664
beta W102_01	0.165	0.869	beta W102_03	0.101	0.920	beta W102_04	1.548	0.122
beta W103_01	-0.378	0.705	beta W103_03	0.945	0.345	beta W103_04	0.165	0.869
beta W104_01	0.303	0.762	beta W104_02	0.625	0.532	beta W104_03	1.576	0.115
beta W106_02	-0.669	0.503	beta W107_01	-0.259	0.796	beta W107_02	-0.583	0.560
beta W109_02	-0.259	0.796	beta W109_03	-1.020	0.308	beta W109_05	-0.258	0.796
beta W110_01	0.764	0.445	beta W110_02	1.561	0.119	beta W110_03	-1.552	0.121
beta W110_04	-1.304	0.192	beta BO01_01	-0.494	0.621	beta BO01_02	-0.786	0.432
beta BO01_03	0.625	0.532	beta BO01_04	-0.817	0.414	beta BO02_06	0.326	0.744
beta BO02_07	0.326	0.744	beta BO02_01	-0.039	0.969	beta BO02_02	0.281	0.779
beta BO02_03	0.944	0.345	beta BO02_09	0.586	0.558	beta BO02_10	1.384	0.166
beta BO02_12	-0.247	0.805	beta BO03_01	-0.690	0.490	beta BO03_02	0.408	0.683
beta BO03_03	-0.960	0.337	beta BO03_04	-1.455	0.146	beta BO03_05	-1.300	0.193
beta BO03_06	-0.291	0.771	beta BO03_07	-1.132	0.258	beta BO03_09	0.116	0.908
beta BO03_10	0.871	0.384	beta BO03_11	0.116	0.908	beta BO05	1.348	0.178
beta EL01_01	-1.454	0.146	beta EL01_02	-1.454	0.146	beta EL01_03	-1.558	0.119
beta EL01_04	0.362	0.717						

Tabelle 30 Ergebnisse des Wald-Tests nach dem letzten Durchgang für 85 Items

Es werden alle Items ausgeschlossen, die im Wald-Test nicht dem geforderten Niveau $p=0,05$ entsprechen. Dann wird der Ablauf wiederholt. Dieser Vorgang wird insgesamt siebenmal wiederholt, bis sowohl der LR-Test mit 0,943 einen guten Testparameter liefert als auch weder der LR-Test noch der Wald-Test noch der abschließende graphische Test signifikante Verletzungen bei einem Item zeigen.

In Tabelle 30 sind die Ergebnisse des Wald-Tests für die verbliebenen Items angegeben. Man erkennt, dass kein Item unter dem gewählten Niveau ($p=0,05$) liegt.

Ergänzend werden dann die verbliebenen Items einem graphischen Check unterzogen, der bei diesem Prozess bisher nur rein informativ und nicht zum Ausschluss von Items genutzt wurde. Dabei erkennt man in Abbildung 80, dass es auch hier keine signifikanten Verletzungen gibt, da alle Konfidenzumbegungen der Items die Hauptdiagonale schneiden und somit keine signifikante Abweichung vorliegt. Das Splitkriterium ist „Mean“.

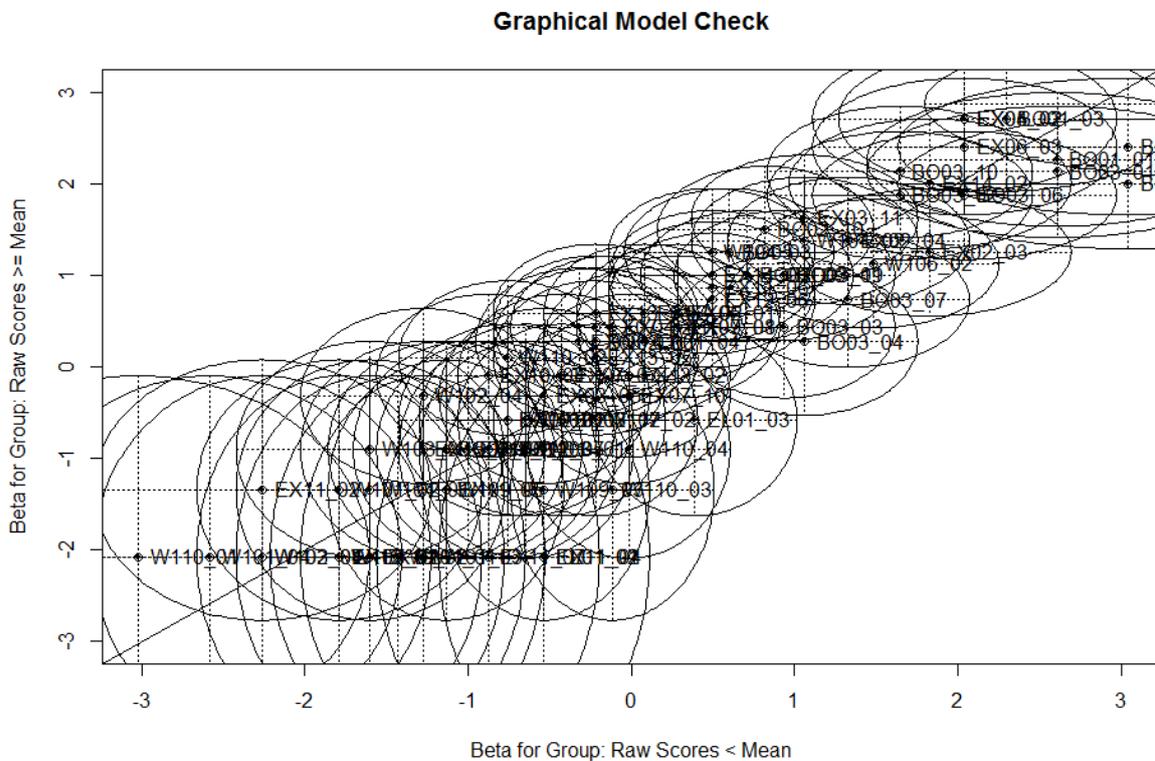


Abbildung 80 Graphischer Test

In Tabelle 31 sind die geschätzten Schwierigkeiten β_j für die einzelnen Items wiedergegeben. Zu beachten ist dabei, dass R mit dem Modul eRm nicht die Schwierigkeit, sondern die Einfachheit eines Items berechnet. Je höher also ein β -Wert ist, als desto einfacher wird das Item geschätzt.

Nach dem Test können die so gewonnenen Modellparameter untersucht werden. Zunächst sollen die geschätzten Schwierigkeiten β_j für die einzelnen Items j betrachtet werden. Auch wenn in Abbildung 81 die einzelnen Items nicht beschriftet sind, sieht man doch gut die gleichmäßige Schwierigkeitsverteilung. In der Abbildung ist auf der Abszisse die latente Variable dargestellt, die allen ausgewählten Items des e-Assessments entspricht. Von links nach rechts nimmt die Schwierigkeit eines Items ab.

Auf der Ordinate ist die Wahrscheinlichkeit einer richtigen Lösung in Abhängigkeit der vorhandenen Kompetenz aufgetragen. Je weiter rechts die ICC-Kurve eines Items verläuft, desto geringer muss die Kompetenz sein, um zu einer richtigen Antwort zu gelangen. Man kann auch sagen, je weiter rechts die Kurve verläuft, desto einfacher ist das Item.

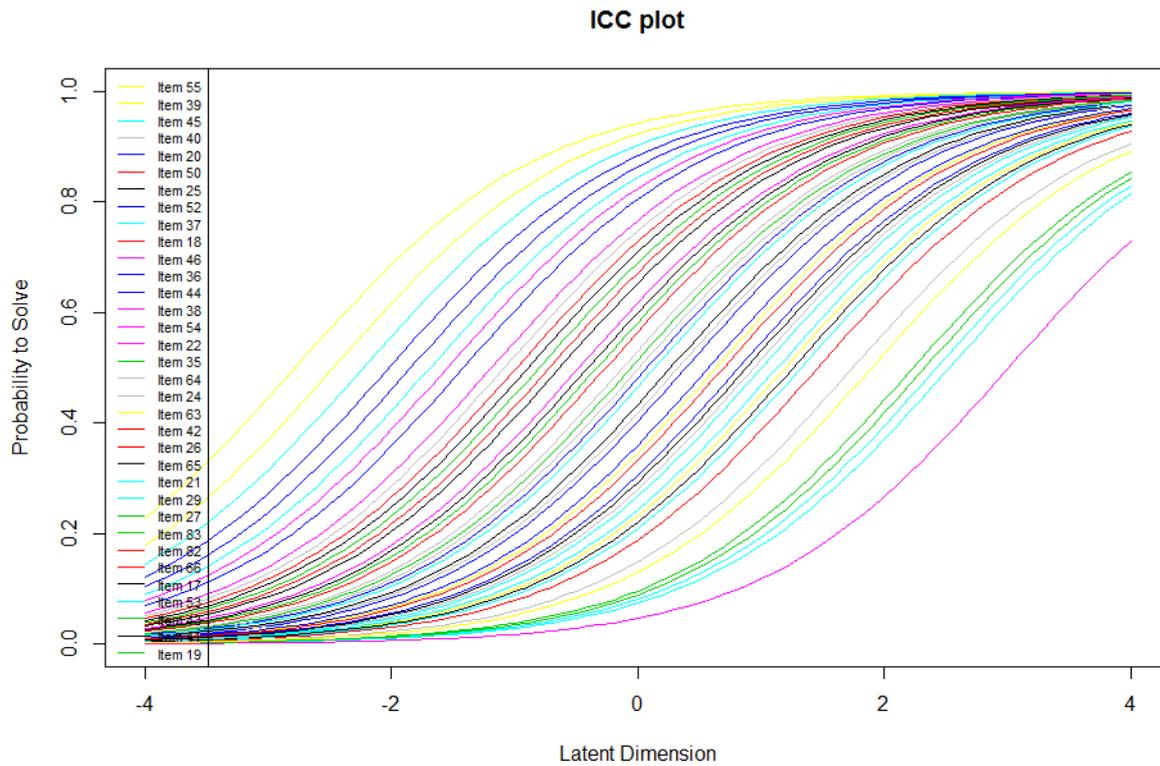


Abbildung 81 ICC-Plot für die verbliebenen Items

Item	B-Wert
EX02_02	-0,26704759
EX02_03	-1,4619357
EX02_04	-1,32746245
EX03_05	0,47725146
EX03_11	-1,32746245
EX07_07	0,32983111
EX07_08	-0,32940785
EX07_02	-0,01163419
EX07_10	0,12118852
EX04_01	-0,0765913
EX04_02	-2,42799564
EX05_01	-0,39135856
EX05_02	-2,42799557
EX06_03	-2,23766461
EX08_01	0,12118846

EX09_01	-0,3294079
EX10_01	0,71565792
EX10_02	1,68504124
EX10_04	0,63348346
EX11_02	2,01915897
EX11_04	0,80104152
EX11_05	1,18551083
EX11_06	-0,63678771
EX11_07	1,08157909
EX12_01	1,84185194
EX12_04	0,98338675
EX12_05	0,80104145
EX12_06	-0,57562342
EX12_07	0,80104151
EX13_02	0,05424953
EX13_04	-0,0765913
EX13_05	0,12118852
EX14_01	-0,69795466
EX14_02	-1,90008811
EX15_01	1,18551083
W101_01	1,54390154
W101_02	1,6850413
W101_03	1,41510113
W101_04	2,47082592
W102_01	2,2245068
W102_03	0,63348348
W102_04	0,98338684
W103_01	0,71565789
W103_03	1,41510118
W103_04	2,2245068
W104_01	1,54390157
W104_02	-1,19691123
W104_03	-0,82060862
W106_02	-1,2617535
W107_01	1,84185197
W107_02	0,32983099
W109_02	1,84185192
W109_03	0,7156579
W109_05	1,18551083
W110_01	2,78139357
W110_02	0,47725145
W110_03	0,40256991
W110_04	0,25881431
BO01_01	-2,33062739
BO01_02	-2,53044154
BO01_03	-2,53044154

BO01_04	-3,0117284
BO02_06	1,08157906
BO02_07	1,08157909
BO02_01	0,89009902
BO02_02	0,71565797
BO02_03	0,05424952
BO02_09	-0,82060869
BO02_10	-1,13283106
BO02_12	0,4772514
BO03_01	-2,2376646
BO03_02	-1,74689176
BO03_03	-0,69795483
BO03_04	-0,69795474
BO03_05	-2,23766461
BO03_06	-1,90008805
BO03_07	-1,00657226
BO03_09	-0,94421265
BO03_10	-1,90008808
BO03_11	-0,94421255
BO05	-0,88225219
EL01_01	0,80104144
EL01_02	0,80104144
EL01_03	-0,07659127
EL01_04	-0,14074078

Tabelle 31 Schwierigkeitswerte der einzelnen Items

ML estimated ability parameters (without spline interpolated values)				
Person	Geschätztes θ	Std-Abweichung	2,5%	97,5%
58	2.659357995	0.3717723	1.93069764	3.38801834
59	0.387078402	0.2588994	-0.12035503	0.89451184
60	-0.007655246	0.2547836	-0.50702186	0.49171137
61	0.730404577	0.2657290	0.20958540	1.25122375
62	-0.331288588	0.2545309	-0.83015990	0.16758272
63	0.057355632	0.2551812	-0.44279040	0.55750167
64	0.590893263	0.2625924	0.07622155	1.10556497
65	0.387078402	0.2588994	-0.12035503	0.89451184
66	-1.210494556	0.2700955	-1.73987201	-0.68111710
67	0.946738262	0.2715807	0.41444980	1.47902672
68	0.946738262	0.2715807	0.41444980	1.47902672
69	0.254014119	0.2570591	-0.24981253	0.75784076
70	0.057355632	0.2551812	-0.44279040	0.55750167
71	1.332962819	0.2851395	0.77409974	1.89182590
72	-0.266554828	0.2543462	-0.76506424	0.23195459
73	1.767263109	0.3056894	1.16812282	2.36640339
74	-0.526493113	0.2558305	-1.02791175	-0.02507448
75	1.415232930	0.2885738	0.84963871	1.98082715

76	0.057355632	0.2551812	-0.44279040	0.55750167
77	1.960893322	0.3169617	1.33965989	2.58212675
78	2.659357995	0.3717723	1.93069764	3.38801834
79	1.862325638	0.3110447	1.25268918	2.47196210
80	1.675325023	0.3008208	1.08572719	2.26492286
81	2.282613522	0.3391438	1.61790396	2.94732309
83	2.659357995	0.3717723	1.93069764	3.38801834
84	-0.266554828	0.2543462	-0.76506424	0.23195459
85	0.946738262	0.2715807	0.41444980	1.47902672
86	1.021108852	0.2738743	0.48432505	1.55789265
87	2.170438046	0.3308778	1.52192949	2.81894660
88	3.130806461	0.4256385	2.29657033	3.96504259
89	0.873564555	0.2694673	0.34541827	1.40171084
90	0.454386855	0.2600021	-0.05520793	0.96398164
92	1.675325023	0.3008208	1.08572719	2.26492286
93	1.173897959	0.2790548	0.62696056	1.72083536
94	0.730404577	0.2657290	0.20958540	1.25122375
95	0.320309123	0.2579197	-0.18520419	0.82582244
96	0.873564555	0.2694673	0.34541827	1.40171084
97	-0.201884178	0.2542800	-0.70026386	0.29649550
100	0.522301830	0.2612325	0.01029554	1.03430812
101	0.387078402	0.2588994	-0.12035503	0.89451184
102	1.415232930	0.2885738	0.84963871	1.98082715
103	1.415232930	0.2885738	0.84963871	1.98082715
104	-0.461193408	0.2552686	-0.96151059	0.03912377
105	1.586187153	0.2963763	1.00530020	2.16707411
106	1.415232930	0.2885738	0.84963871	1.98082715
107	0.254014119	0.2570591	-0.24981253	0.75784076
108	0.522301830	0.2612325	0.01029554	1.03430812
109	-0.266554828	0.2543462	-0.76506424	0.23195459
110	0.254014119	0.2570591	-0.24981253	0.75784076
111	0.320309123	0.2579197	-0.18520419	0.82582244
112	0.254014119	0.2570591	-0.24981253	0.75784076
113	-0.396146971	0.2548359	-0.89561607	0.10332213
114	-0.137216708	0.2543327	-0.63569973	0.36126631
115	0.454386855	0.2600021	-0.05520793	0.96398164
116	-0.137216708	0.2543327	-0.63569973	0.36126631
117	-0.724610439	0.2583436	-1.23095464	-0.21826624
118	0.188130778	0.2563169	-0.31424121	0.69050277
120	1.415232930	0.2885738	0.84963871	1.98082715
121	1.096788001	0.2763615	0.55512947	1.63844653
122	0.946738262	0.2715807	0.41444980	1.47902672
123	-0.266554828	0.2543462	-0.76506424	0.23195459
124	1.173897959	0.2790548	0.62696056	1.72083536
125	0.522301830	0.2612325	0.01029554	1.03430812
126	0.188130778	0.2563169	-0.31424121	0.69050277
127	0.254014119	0.2570591	-0.24981253	0.75784076
128	0.122597996	0.2556917	-0.37854851	0.62374450
129	1.021108852	0.2738743	0.48432505	1.55789265
130	0.122597996	0.2556917	-0.37854851	0.62374450
131	0.387078402	0.2588994	-0.12035503	0.89451184

132	2.958954059	0.4040777	2.16697625	3.75093187
133	1.415232930	0.2885738	0.84963871	1.98082715
134	0.946738262	0.2715807	0.41444980	1.47902672
135	0.387078402	0.2588994	-0.12035503	0.89451184
136	1.252573221	0.2819747	0.69991290	1.80523355
137	-1.435921893	0.2785453	-1.98186065	-0.88998314
138	0.590893263	0.2625924	0.07622155	1.10556497

Tabelle 32 Geschätzte Kompetenzen der Assessment-Teilnehmer

In Tabelle 32 sind die geschätzten Kompetenzen θ_i aller Assessment-Teilnehmer zusammen mit einer Standardabweichung und den Grenzen eines 95%-Konfidenzintervalls angegeben.

A.3 Erläuterungen zur zweiten statistischen Analyse

Um den Einfluss der Vorgehensweise auf mögliche Ergebnisse zu testen, wird die statistische Analyse wiederholt. Diesmal wird statt des Wald-Tests der graphische Test nach jedem Durchlauf als Kriterium für den Ausschluss von Items verwendet. Die Ergebnisse sind in der Arbeit in Abschnitt 8.3 dokumentiert.

Hier sind ergänzend die Ergebnisse des graphischen Modelltests in Abbildung 82 wiedergegeben. Man sieht, dass wiederum alle verbliebenen Items mit ihren Konfidenzumgebungen die Hauptdiagonale überdecken. Außerdem sind weniger sehr leichte oder sehr schwere Items als bei der ersten Vorgehensweise enthalten.

Auch die ICC-Kurven bilden ein homogenes Spektrum, wie Abbildung 83 zeigt. Die übrigen Modellparameter sind ebenfalls in Abschnitt 8.3 der Arbeit dokumentiert. Damit hat auch die zweite Analyse ein befriedigendes Modell ergeben.

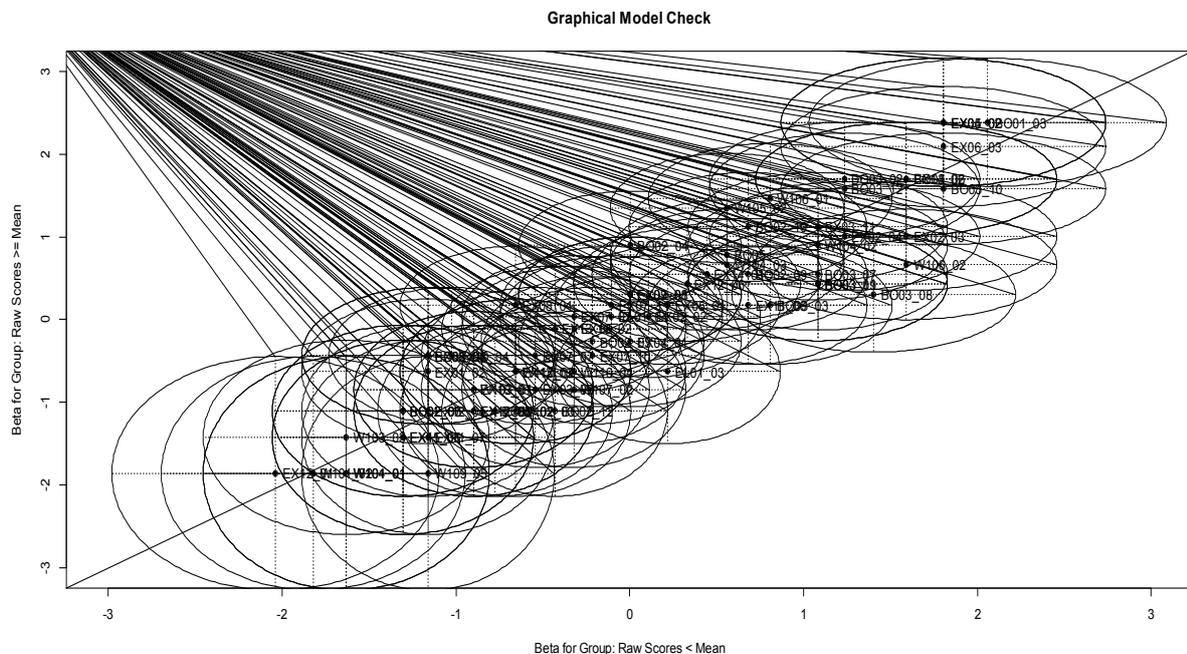


Abbildung 82 Graphischer Modelltest

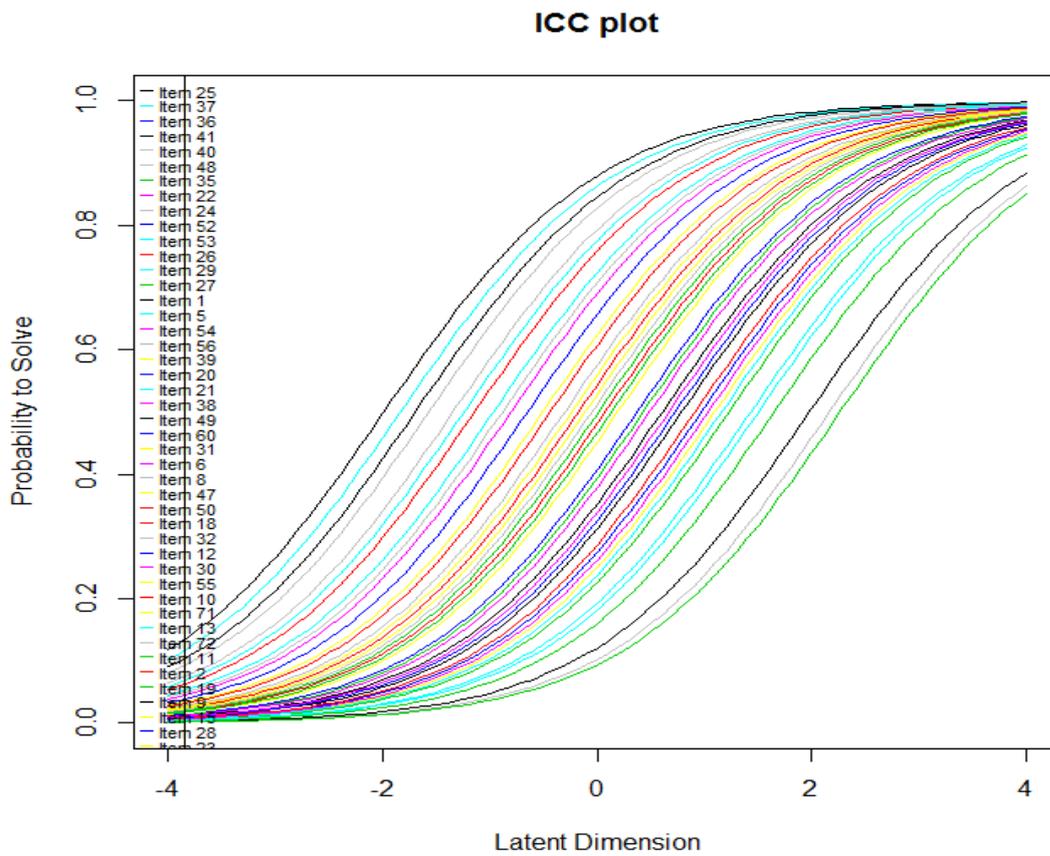


Abbildung 83 ICC-Kurven der zweiten Analyse

A.4 Erläuterungen zu den Typen des CSP-Informationsmodells

Hier werden die Tabellen beschrieben, die bei der Umsetzung des CSP-Informationsmodells in Access entwickelt worden sind.

A.4.1 Modellbereich CONTENT

FCT_FACT

Ein Fakt stellt die Grundeinheit des beschriebenen Wissens dar. Er hat eine eigene Kurzbezeichnung und besteht aus mehreren Valenzen. Die erste Valenz sollte ein Prädikat sein.

FCT_FACT		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung des Fakts	X
definition	Definition des Fakts	X
description	Weitere Erläuterungen	
is_of	-> FCT_FACT_TYPE	X

FCT_FACT_TYPE

Klassifikation der FACTs. Hier kann beispielsweise eine Einordnung nach deklarativem, prozeduralem oder anderem Wissen erfolgen.

FCT_FACT_TYPE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Typisierung für die Fakten	X
definition	Langtext und Definition des Typs	X
description	Weitere Erläuterungen	

FCT_VALENCE

Ein Platzhalter innerhalb eines FACTs. Die Valenzen werden entsprechend ihrer Reihenfolge innerhalb des FACTs nummeriert. Eine Valenz kann entweder von einer ENTITY oder wiederum von einem FACT ausgefüllt werden, aber nicht von beiden gleichzeitig.

FCT_VALENCE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung der Valenz, Prädikat oder andere Angabe, die der Beschriftung der Valenz dient.	X
description	Weitere Erläuterungen	
position	Position, die die Valenz im Fakt einnimmt, von 1 an nummeriert.	X
implemented_by_entity	-> FCT_ENTITY, die Entität, die die Valenz ausfüllt. Entweder muss dieses Feld oder implemented_by_fact belegt sein.	(X)
implemented_by_fact	-> FCT_FACT, Fakt, der die Valenz ausfüllt. Entweder muss dieses Feld oder implemented_by_entity belegt sein.	(X)
used_by	-> FCT_FACT, Fakt, in dem die Valenz verwendet wird.	X
is_of	-> VALENCE_TYPE, Typisierung der Valenz, grammatische Rolle	X

FCT_VALENCE_TYPE

Hier kann eine Typisierung der Valenz erfolgen. In kasus-basierten Sprachen kann dies beispielsweise der Kasus sein, es kann sich aber auch um temporale, kausale oder andere Ergänzungen handeln.

FCT_VALENCE_TYPE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Typisierung für die Valenzen	X
definition	Langtext und Definition des Typs	X
description	Weitere Erläuterungen	

FCT_CONSTRAINT

Hier werden die verwendeten CONSTRAINTs dokumentiert. Diese Regeln lassen auch die Dokumentation einer automatischen Prüfung zu, sofern eine geeignete Sprache zur Verfügung steht.

FCT_CONSTRAINT		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung des Constraints	X
definition	Langtext und Definition des Constraints	X
description	Weitere Erläuterungen	
rule	Regel, die als Constraint angewendet werden soll	
is_of	-> FCT_CONSTRAINT_TYPE, Typisierung des Constraints	X

FCT_CONSTRAINT_TYPE

Hier erfolgt die Klassifizierung der CONSTRAINTs, beispielsweise als UND, ODER oder NICHT. Es können letztlich beliebige weitere CONSTRAINT TYPEs definiert werden.

FCT_CONSTRAINT_TYPE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Typisierung für die Constraints	X
definition	Langtext und Definition des Typs	X
description	Weitere Erläuterungen	

ENTITY

Eine Klasse oder Instanz eines konkreten oder abstrakten Dings, das für die Besetzung der VALENCE eines FACTs genutzt werden kann.

FCT_ENTITY		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung der Entity	X
definition	Definition der Entity	
description	Weitere Erläuterungen	
part_of	-> FCT_ENTITY, übergeordnete ENTITY	
qualified_by	-> FCT_QUALIFIER, Bezeichnungsart	
area	Markierung T/F, ob eine eigene Modellierung für die ENTITY existiert.	

QUALIFIER

Bezeichner für ENTITYs, beispielsweise der Nachname einer Person.

FCT_QUALIFIER		
Feld	Befüllung	Muss
uid	Freie Vergabe	X
sname	Benennungsart	X
definition	Langtext und Definition der Bezeichnung	
description	Weitere Erläuterungen	

CAUSES_FACT

Aus den Kompetenzen werden Fakten hergeleitet. Dabei können grundsätzlich aus einer Kompetenz mehrere Fakten entstehen und umgekehrt kann ein FACT mehreren Kompetenzen zugeordnet werden.

FCT_CAUSES_FACT		
Feld	Befüllung	Muss
caused_by	-> CSD_COMPETENCE	X
causes	-> FCT_FACT	X
support_indication	Wert zwischen 0 und 1 als Hinweis auf den Zusammenhang.	

A.4.2 Modellbereich ITEM

ITM_SNIPPET

Ein Aspekt eines FACTs, der für die Beurteilung der Bedeutung des FACTs bekannt sein muss.

ITM_SNIPPET		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung des Snippets	X
description	Weitere Erläuterungen	
belongs_to	-> FCT-FACT, Fakt aus dem das Snippet abgeleitet wurde	X

ITM_ABSTRACT_ITEM

Ein Einzelaspekt, der in einem ASSESSMENT durch konkrete Ausgestaltung in Items thematisiert werden soll.

ITM_ABSTRACT_ITEM		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbeschreibung des Items	X
definition	Definition des Items als Aussage	X
description	Weitere Erläuterungen	
answer	Angabe, ob das Item verifiziert oder falsifiziert werden soll.	X
belongs_tp	-> ITM_SNIPPET, Snippet aus dem das Item abgeleitet wurde.	X

ITM_CONTEXT_ELEMENT

Repräsentiert **FACTs**, **SNIPPETs** und **ABSTRACT ITEMs**. Der Typ dient so als Anker für andere Teile des Modells, um verallgemeinerte Bezüge zu den repräsentierten Typen herstellen zu können.

ITM_CONTEXT_ELEMENT		
Feld	Befüllung	muss
is_a	-> FCT_FACT oder -> ITM_SNIPPET oder -> ITM_ITEM	X
type	Art des repräsentierten Inhaltes, also FACT, SNIPPET oder ITEM	X

A.4.3 Modellbereich SKILL

SKI_OPERATOR

Hier werden die zur Spezifikation der Fähigkeiten verwendeten Verben dokumentiert. Der **OPERATOR** wird durch eine externe Spezifikation definiert, beispielsweise die Bildungsstandards oder ein Curriculum.

SKI_OPERATOR		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung des Operators	X
description	Weitere Erläuterungen	
used_by	-> CSP_CURRICULUM_STANDARD	X

SKI_SKILL

Hier werden die Fähigkeitsbeschreibungen zusammengeführt, also wie die Fähigkeit hinsichtlich eines **CONTEXT ELEMENTs** zu bewerten ist.

SKI_SKILL		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung des SKILL	X
description	Weitere Erläuterungen	
uses	-> SKI_OPERATOR, verwendeter Operator	X
references	-> ITM_CONTEXT_ELEMENT	X
references_type	-> ITM_CONTEXT_ELEMENT	X
weight	Gewichtung des Skills bei der Aggregation zum CONTEXT_ITEM	X
aggregate	Aggregatsfunktion für die Skills	

SKI_SKILL_INDICATING

Hier wird die Zuordnung von **SKILLs** entsprechend den **COMPETENCE METRICS** zu messbaren **SPECIFIC COMPETENCE NIVEAU**s abgelegt.

SKI_SKILL_INDICATING		
Feld	Befüllung	muss
indicates	-> SKI_SKILL oder -> CONTEXT_ELEMENT	X
is_indicated	-> CSD_SPECIFIC_COMPETENCE_NIVEAU	X
by_using	-> SKI_COMPETENCE_METRICS	X
weight	Gewichtung des Skills für die Metrik, Zahl zwischen 0 und 100%	X
type	Typ der Zuordnung SKILL oder CONTEXT_ELEMENT	X

SKI_CNM_AGGREGATE

Hier wird die hierarchische Anordnung von Anordnung der SKILL INDICATING hinterlegt. Dadurch wird es möglich verschiedene Zuordnungsaspekte miteinander zu verbinden.

SKI_CNM_AGGREGATE		
Feld	Befüllung	muss
is_master	-> SKI_SKILL_INDICATING	X
Operator	AND, OR oder spezifisch	X

SKI_LEVELINDICATOR

Verbindung zwischen SKILL INICATING und SPECIFIC COMPETENCE NIVEAU.

SKI_LEVELINICATOR		
Feld	Befüllung	muss
is_indicated	-> CSD_SPECIFIC_COMPETENCE_NIVEAU	X
defines	-> SKI_SKILL_INICATING	X
threshold	Level, ab dem ein Niveau erreicht ist, Zahl zwischen 0 und 1	X

SKI_COMPETENCE_NIVEAU_METRICS

Die COMPETENCE_NIVEAU_METRICS beschreiben den Zusammenhang in Formel- oder Algorithmusform, wie aus den gemessenen SKILLS die entsprechenden SPECIFIC COMPETENCE NIVEAUs bestimmt werden können.

SKI_COMPETENCE_NIVEAU_METRICS		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung der Metrik	X
definition	Genaue Definition, hier beispielsweise durch eine Competence Map mit Schwellwerten.	X
description	Algorithmus oder Formel zur Berechnung des Vergleichswertes für das Kompetenzniveau	X

SKI_COMPETENCE_NIVEAU_OPERATOR

Hier wird die Zuordnung des Operators zu einem Kompetenzniveau dokumentiert.

SKI_COMPETENCE_NIVEAU_OPERATOR		
Feld	Befüllung	muss
uses	-> SKI_SKILL	X
indicates	-> CSD_COMPETENCE_NIVEAU	X

SKI_QUALIFICATION_NIVEAU_OPERATOR

Hier wird die Zuordnung des Operators zu einem Qualifikationsniveau dokumentiert (Erweiterung).

SKI_QUALIFICATION_NIVEAU_OPERATOR		
Feld	Befüllung	muss
uses	-> SKI_SKILL	X
indicates	-> CSD_QUALIFICATION_NIVEAU	X

A.4.4 Modellbereich COMPETENCE SPECIFICATION

CSP_CURRICULUM_STANDARDS

Dieser Typ ist eine Spezialisierung der EDUCATIONAL_STANDARDS, die wiederum eine Spezialisierung der EXTERNAL_SPECIFICATIONs sind. Andere Spezialisierungen und zusätzliche Modellierungen externer Anforderungen sind sinnvoll.

CSP_CURRICULUM_STANDARDS		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung des Standards	X
definition	Genaue Angabe des Standards	
description	Weitere Erläuterungen zum Standard	
subject	Inhaltlicher Bezug	X
release_date	Datum der Herausgabe	X
valid_from	Gültigkeitsbeginn	X
valid_to	Gültigkeitsende	
version	Version des Standards	X
valid_for	Gültigkeitsbereich	X
responsible	Herausgeber	X

CSP_EXTERNAL_GOAL

Hier werden die externen Ziele beschrieben, die sich aus den externen Anforderungen ergeben.

CSP_EXTERNAL_GOAL		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung des Ziels	X
definition	Definition	
description	Weitere Erläuterungen	
type	Typ des Ziels COMPETENCE, QUALIFICATION, ...	X

CSP_DEFINE_GOAL

Verbindung zwischen den Zielen und den sie definierenden externen Anforderungen.

CSP_DEFINE_GOAL		
Feld	Befüllung	muss
defined_by	-> CSP_CURRICULUM_STANDARD	X
defines	-> CSP_EXTERNAL_GOAL	X
source	Genauere Angabe der Stelle im Standard	

A.4.5 Modellbereich COMPETENCE DIMENSION

CSD_ABSTRACT_DIMENSION

Dies ist eine allgemeine aus der Theorie abgeleitete oberste Gliederungsebene von Kompetenzen ohne Bezug zu einer externen Spezifikation. Sie dient der Einordnung in Standarddimensionen.

CSD_ABSTRACT_DIMENSION		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung	X
definition	Definition	
description	Weitere Erläuterungen	
part_of	-> CSD_ABSTRACT_DIMENSION	

CSD_COMPETENCE_DIMENSION

Bezeichnung einer Kompetenzdimension. Sie steht in Bezug zu einer externen Spezifikation und/oder kann einer oder mehreren abstrakten Dimensionen zugeordnet werden.

CSD_COMPETENCE_DIMENSION		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung der Dimension	X
description	Genauere Angabe der Dimension	
supports	->CSP_CURRICULUM_STANDARD Zugeordneter Standard	X

CSD_SUPPORTS_ABSTRACT_DIMENSION

Dies ist die Darstellung der Beziehung zwischen CSD_ABSTRACT_DIMENSION und CSD_COMPETENCE_DIMENSION. Hier wird eine Zuordnung der Kompetenzdimension zu abstrakten Kompetenzdimensionen getroffen.

CSD_SUPPORTS_ABSTRACT_DIMENSION		
Feld	Befüllung	muss
supports	-> CSD_ABSTRACT_DIMENSION	X
includes	-> CSD_COMPETENCE_DIMENSION	X

CSD_COMPETENCE

Dieser Typ beschreibt eine einzelne Kompetenz. Damit ist dies auf unterster Ebene die feinste messbare Kompetenz. Dies entspricht einer latenten Variablen, die in einem Assessment gemessen werden soll.

CSD_COMPETENCE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung der Kompetenz	X
definition	Kurzbeschreibung	
description	Genauere Angabe der Kompetenz	
part_of	-> CSD_COMPETENCE, vorhanden soweit nicht oberste Kompetenz	(X)
belongs_to_dimension	-> CSD_COMPETENCE_DIMENSION	X
belongs_to_level	-> COMPETENCE_LEVEL zugehöriges Level	

CSD_COMPETENCE_LEVEL

Ebene für eine Strukturierung von Kompetenzen in Ebenen. Kann als zusätzliches Strukturierungsmittel für die Kompetenzdimensionen verwendet werden.

CSD_COMPETENCE_LEVEL		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung der Kompetenzebene	X
definition	Kurzbeschreibung	
description	Genauere Angabe der Kompetenz	
part_of	-> CSD_COMPETENCE_LEVEL	(X)
belongs_to	-> CSD_COMPETENCE_DIMENSION	X

CSD_COMPETENCE_NIVEAU_SCALE

Skalen für die Niveaus der Kompetenzmessung. Eine Skala umfasst eine Reihe von Kompetenzniveaus. Dies können die Bloomsche Taxonomie, die SOLO-Taxonomie, die Anforderungsbereiche oder andere zu verwendende Niveaustrukturen sein.

CSD_COMPETENCE_NIVEAU_SCALE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung der Skala	X
definition	Definition	
description	Genauere Angabe der Skala	

CSD_COMPETENCE_NIVEAU

Die einzelnen Niveaus für die Kompetenzskalen. Jedes Niveau unterhalb des höchsten Niveaus verweist auf das übergeordnete Niveau. Mehrere parallel untergeordnete Niveaus sind erlaubt.

CSD_COMPETENCE_NIVEAU		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung des Niveaus	X
definition	Definition	
description	Genauere Angabe des Niveaus	
measured_by	-> COMPETENCE_NIVEAU_SCALE, zugehörige Skala	X
under	-> COMPETENCE_NIVEAU, übergeordnetes Niveau	

CSD_SPECIFIC_COMPETENCE_NIVEAU

Hier wird der Zusammenhang zwischen einer Kompetenz und der Ausprägung spezifischer Niveaus gemessen. Dies stellt die Verbindung zwischen Kompetenzstruktur- und Kompetenzniveaumodell dar. In den meisten Fällen wird sich das auf das Anforderungsniveau einer Skala beziehen. Wie bereits dargestellt, kann es sich aber auch um ein mehrdimensionales System von Skalen handeln, so dass hier mehrere Skalen zugeordnet werden können. Daher wird die Zuordnung zu den Niveaus in CSD_APPLIES_NIVEAU vorgenommen, was sich auf diesen Typ bezieht. Das Niveau wird außerdem von SKI_SKILL_INDICATING referenziert, was sich bei der Messung der Fähigkeiten auf das Erreichen eines spezifischen Niveaus bezieht.

CSD_SPECIFIC_COMPETENCE_NIVEAU		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bezeichnung des Niveaus	X
definition	Definition	
description	Genauere Angabe des Niveaus	
related_to	-> CSD_COMPETENCE	X
supports	-> CSP_EXTERNAL_GOAL	

CSD_APPLIES_NIVEAU

Die Zuordnung von **COMPETENCE_NIVEAU**s zu **SPECIFIC_COMPETENCE_NIVEAU**s, also die Anwendung einer allgemeinen Skala auf ein spezifisches Kompetenzniveau.

CSD_APPLIES_NIVEAU		
Feld	Befüllung	muss
applies	-> CSD_COMPETENCE_NIVEAU	X
to	-> CSD_SPECIFIC_COMPETENCE_NIVEAU	X

A.4.6 Modellbereich FEEDBACK

FBCK_EDUCATION_COURSE

Beschreibung eines Bildungsprozesses, der zu den im Assessment zu prüfenden Kompetenzen führt.

FBCK_EDUCATION_COURSE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung des Kurses	X
definition	Definition des Kurses	
description	Zusatzbeschreibung	

FBCK_EDUCATION_MODULE

Eine Unterrichtseinheit, beispielsweise eine Lernsituation.

FBCK_EDUCATION_MODULE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung des Moduls	X
definition	Definition des Moduls	
description	Zusatzbeschreibung	
is_part_of	-> FBCK_EDUCATION_COURSE	X

FBCK_EDUCATION_MATERIAL

Material, das eine Unterrichtseinheit unterstützt.

FBCK_EDUCATION_MATERIAL		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung des Materials	X
definition	Definition des Materials	
description	Zusatzbeschreibung	
supplies	-> FBCK_EDUCATION_MODULE	X

FBCK_MODULE_GOAL

Beziehung zwischen einer Unterrichtseinheit und einem damit verbundenen Ziel.

FBCK_MODULE_GOAL		
Feld	Befüllung	muss
targets	-> CSP_EXTERNAL_GOAL	X
targeted_by	-> FBCK_EDUCATION_MODULE	X

FBCK_MODULE_COMPETENCE_NIVEAU

Verbindung zwischen einer Unterrichtseinheit und einem angestrebten Kompetenzniveau für eine bestimmte Kompetenz.

FBCK_MODULE_COMPETENCE_NIVEAU		
Feld	Befüllung	muss
targets	-> CSD_SPECIFIC_COMPETENCE_NIVEAU	X
targeted_by	-> FBCK_EDUCATION_MODULE	X

FBCK_MODULE_CONTEXT

Verbindung zwischen einer Unterrichtseinheit und einem dort zu vermittelnden FACT, SNIPPET oder ABSTRACT ITEM.

FBCK_MODULE_CONTEXT		
Feld	Befüllung	muss
targets	-> ITM_CONTEXT_ELEMENT	X
targeted_by	-> FBCK_EDUCATION_MODULE	X

FBCK_ERRORTYPE

Ein typisierter Fehler als Ergebnis einer inhaltlichen Analyse oder einer Fehleranalyse.

FBCK_ERROR_TYPE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung des Fehlers	X
definition	Definition des Fehlers	
description	Zusatzbeschreibung	
causes	-> FBCK_FEEDBACK	

FBCK_FEEDBACK

Eine Ursachenbeschreibung für typisierte Fehler.

FBCK_FEEDBACK		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung des Feedbacks	X
definition	Ursachenbeschreibung	
description	Weitere Analysen	
related_to	-> FBCK_EDUCATION_MODULE	X

FBCK_ACTION

Eine Maßnahme, die aufgrund ermittelter Fehlerursachen ergriffen werden kann.

FBCK_ACTION		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung der Aktion	X
definition	Beschreibung der Maßnahme	
description	Weitere Hinweise	
related_to	-> FBCK_FEEDBACK	X

FBCK_FEEDBACK_MATERIAL

Material, das bei der Durchführung der Feedback-Maßnahmen unterstützend verwendet werden kann.

FBCK_FEEDBACK_MATERIAL		
Feld	Befüllung	Muss
uid	Freie Vergabe	X
sname	Benennung des Materials	X
definition	Definition des Materials	
description	Zusatzbeschreibung	
supplies	-> FBCK_ACTION	X

FBCK_FEEDBACK_CONTEXT

Beziehung zwischen FEEDBACK und einen CONTEXT ELEMENT.

FBCK_FEEDBACK_CONTEXT		
Feld	Befüllung	Muss
indicates	-> ITM_CONTEXT_ELEMENT	X
type	FACT, SNIPPET oder ITEM	X
related_to	-> FBCK_FEEDBACK	X

FBCK_FEEDBACK_SKILL

Beziehung zwischen FEEDBACK und SKILL.

FBCK_FEEDBACK_SKILL		
Feld	Befüllung	muss
indicates	-> SKI_SKILL	X
related_to	-> FBCK_FEEDBACK	X

FBCK_FEEDBACK_ERROR_TYPE

Beziehung zwischen dem Fehlertyp und dem Antworttyp im Assessment.

FBCK_FEEDBACK_ERROR		
Feld	Befüllung	muss
indicates	-> FBCK_ERROR_TYPE	X
related_to	-> SCO_RESPONSE_TYPE	X

FBCK_FEEDBACK_SCN

Beziehung zwischen FEEDBACK und SPECIFIC COMPTENCE NIVEAU.

FBCK_FEEDBACK_SCN		
Feld	Befüllung	muss
indicates	-> CSD_SPECIFIC_COMPETENCE_NIVEAU	X
related_to	-> FBCK_FEEDBACK	X

A.4.7 Modellbereich ASSESSMENT

ASS_SETTING

Das **SETTING** beschreibt ein komplettes Assessment. Es kann mehrmals konkret genutzt werden. Auch konventionelle Settings außerhalb des CSP-Modells, beispielsweise parallel geschriebene Klausuren können hier hinterlegt werden, um einen Vergleich mit einem e-Assessment vornehmen zu können.

ASS_SETTING		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung des Assessments	X
description	Weitere Erläuterungen	
type	Art des Settings CSP/CONVENTIONAL oder weitere.	X

ASS_SETTING_EXECUTION

Die **SETTING_EXECUTION** beschreibt die tatsächliche Durchführung eines Assessments. Sie bezieht sich immer auf ein **QUESTIONNAIRE** und das zugehörige **SETTING**. Sie ist auch ein zentraler Anker im PERSONAL-Modell, um die beobachteten und bewerteten Ergebnisse dieser Assessment-Ausführung zuzuordnen.

ASS_SETTING_EXECUTION		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung	
description	Weitere Erläuterungen	
participants	Anzahl tatsächlich beobachtbarer Teilnehmer	
day	Tag der Ausführung	
uses	-> ASS_QUESTIONNAIRE	X
described_by	-> ASS_EXECUTION_PARAMETERS	X

ASS_EXECUTION_DESCRIPTION

Hier werden die planbaren Parameter der Assessment-Ausführung gespeichert. Diese sind unter Umständen auch für mehrere Ausführungen desselben oder verschiedener **SETTINGS** nutzbar. Die Attribute können bei Bedarf erweitert werden.

ASS_EXECUTION_PARAMETERS		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung	
duration	Dauer des Assessments	
utilities	Erlaubte Hilfsmittel	
objectivity	Angaben zur Objektivität	
reliability	Angaben zur Reliabilität	
validity	Angaben zur Validität	

ASS_EXECUTION_PROBLEM

Hier wird eine Problemklasse dokumentiert, die während der Assessment-Ausführung auftreten kann, beispielsweise „Absturz des PC während des Assessments“ oder „Nichterscheinen eines Teilnehmers“. Dies stellt eine Sammlung denkbarer Probleme und empfohlener Maßnahmen zu deren Behebung dar. Durch tatsächlich aufgetretene Probleme und empfohlene Lösungen kann diese Sammlung ständig erweitert werden.

ASS_EXECUTION_PROBLEM		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung des Problems	X
description	Weitere Erläuterungen	
recommended _action	Empfohlene Maßnahmen	

ASS_EXECUTION_PROBLEM_ACTION

Hier werden die tatsächlich während einer Assessment-Ausführung aufgetretenen Probleme dokumentiert. Dabei werden die jeweiligen Problemklassen der Assessment-Ausführung zugeordnet und die tatsächlich getroffenen Maßnahmen dokumentiert.

ASS_EXECUTION_PROBLEM_ACTION		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung des Problems	X
description	Weitere Erläuterungen	
action_taken	Tatsächliche Maßnahmen	X
in	-> ASS_SETTING_EXECUTION	X
referred_to	-> ASS_EXECUTION_PROBLEM	X

ASS_RAW_ERROR_TYPE

Im Rahmen der Assessment-Ausführung können abhängig vom verwendeten Werkzeug Antworten entstehen, die offensichtlich Bedienungsfehlern oder Eingabefehlern zuzuordnen sind. Diese sollen korrigiert werden, wenn sie nicht Teil der zu messenden Fähigkeiten sind. In RAW ERROR TYPE werden grundsätzlich mögliche Fehlerklassen dieser Art gesammelt.

ASS_RAW_ERROR_TYPE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung des Problems	X
description	Weitere Erläuterungen	
recommended _action	Empfohlene Maßnahmen	
objectivity	Angaben zur Objektivität der Maßnahme	
reliability	Angaben zur Reliabilität der Maßnahme	
validity	Angaben zur Validität der Maßnahme	

ASS_RAW_ERROR_TYPE_ACTION

Hier werden die tatsächlich während einer Assessment-Ausführung korrigierten Fehler dokumentiert. Dabei werden die jeweiligen Fehlerklassen der Assessment-Ausführung zugeordnet und die tatsächlich getroffenen Maßnahmen dokumentiert.

ASS_RAW_ERROR_TYPE_ACTION		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung des Problems	X
description	Weitere Erläuterungen	
action_taken	Tatsächliche Maßnahmen	X
in	-> ASS_SETTING_EXECUTION	X
referred_to	-> ASS_RAW_ERROR_TYPE	X

ASS_QUESTIONNAIRE

Ein kompletter Fragebogen oder kompletter Aufgabensatz. Hier werden alle Aufgaben gebündelt, die ein Assessment-Teilnehmer bei einer Assessment-Durchführung zu bearbeiten hat. Es besteht die Möglichkeit mehrerer QUESTIONNAIREs pro SETTING anzulegen, wenn den Teilnehmern beispielsweise unterschiedliche Aufgabensätze zur Verfügung gestellt werden sollen.

ASS_QUESTIONNAIRE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Beschreibung des Fragebogens	X
description	Weitere Erläuterungen	
is_part_of	-> ASS_SETTING, Zuordnung des Settings	X

ASS_ASSESSMENT_ITEM

Hier wird ein einzelnes Item dokumentiert. So kann es sich beispielsweise um die einfache Auswahl zwischen den Antwortmöglichkeiten „Ja“ und „Nein“ handeln. Es kann sich aber auch um die Auswahl aus einer Reihe von Werten oder die freie Eingabe eines Wortes oder einer Zahl handeln.

ASS_ASSESSMENT_ITEM		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbeschreibung	
definition	Angezeigter Text des Items	X
description	Weitere Erläuterungen	
answer	Vorgeschlagene Lösung	
control_degree	Geschätzter Aufwand für die manuelle Kontrolle und Nachbearbeitung der Lösung 0 = kein Aufwand, typisch bei festen Antwortalternativen 1 = geringer Aufwand, Rundungen, Schreibweise 2 = Größerer Aufwand, wegen Rechnungen 3 = Größerer Aufwand wegen mehrdeutiger Lösungen, Reihenfolgen, usw.	
technical_dependency	Markierung, ob das Item von eingesetzter Technologie abhängt.	X
estimated_difficulty	Geschätzter Schwierigkeitsgrad, 1 am einfachsten, dann aufsteigend	
openness	Offenheit der Frage, 0 am geringsten (geschlossene Frage), dann aufsteigend	

ASS_CAUSES_ITEM

Beziehung zwischen dem **ABSTRACT ITEM** und dem daraus abgeleiteten **ASSESSMENT ITEM**.
Zusätzlich kann in Vorbereitung der Bewertung ein Hinweis auf einen damit gemessenen **SKILL** hinterlegt werden.

ASS_CAUSES_ITEM		
Feld	Befüllung	muss
based_on	-> ITM_ABSTRACT_ITEM	X
applied_to	-> ASS_ASSESSMENT_ITEM	X
indicates	-> SKI_SKILL, Hinweis auf möglichen Skill	
support_indicator	Prozentzahl zwischen 0 und 100 als Hinweis auf die Stärke des Zusammenhangs	X

ASS_ITEMSET_ITEM

Hier wird die Beziehung hinterlegt, welche Items zu welchem Item Set gehören. Dadurch können Items in verschiedenen Item Sets wiederverwendet werden.

ASS_ITEMSET_ITEM		
Feld	Befüllung	muss
part_of	-> ASS_ASSESSMENT_ITEMSET	X
contains	-> ASS_ASSESSMENT_ITEM	X

ASS_ASSESSMENT_ITEMSET

Gruppierung von Items zu einem Set, beispielsweise für eine Aufgabe oder Teilaufgabe.
ITEMSETs können beliebig oft ineinander verschachtelt werden.

ASS_ASSESSMENT_ITEMSET		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbeschreibung	
definition	Angezeigter Text des Item Sets	X
description	Weitere Erläuterungen	

ASS_ITEMSET_ITEMSET

Verschachtelung von **ITEMSETs** ineinander. Dadurch können beliebig tiefe Schachtelungen von Aufgaben erzeugt werden.

ASS_ITEMSET_ITEMSET		
Feld	Befüllung	muss
part_of	-> ASS_ASSESSMENT_ITEMSET	X
contains	-> ASS_ASSESSMENT_ITEMSET	X

ASS_ITEMSET_EXERCISE

Hier wird die Beziehung hinterlegt, welche **ITEMSETs** zu welcher **EXERCISE** gehören. Dadurch können **ITEMSETs** in verschiedenen **EXERCISEs** wiederverwendet werden.

ASS_ITEMSET_EXERCISE		
Feld	Befüllung	muss
part_of	-> ASS_EXERCISE	X
contains	-> ASS_ASSESSMENT_ITEMSET	X

ASS_EXERCISE

Komplette Zusammenfassung einer Aufgabe. Dies stellt die oberste Grundstruktur dar, die in einen QUESTIONNAIRE eingefügt werden kann.

ASS_EXERCISE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbeschreibung	
definition	Erläuterung	X
description	Weitere Erläuterungen	

ASS_EXERCISE_QUESTIONNAIRE

Hier wird die Verbindung hinterlegt, welche EXERCISEs zu welchem QUESTIONNAIRE gehören. Dadurch können EXERCISEs in verschiedenen QUESTIONNAIREs wiederverwendet werden.

ASS_EXERCISE_QUESTIONNAIRE		
Feld	Befüllung	muss
part_of	-> ASS_QUESTIONNAIRE	X
contains	-> ASS_EXERCISE	X

ASS_ASSESSMENT_ELEMENT

Abstraktion der Typen, die die Assessment-Struktur beschreiben. Dient zum einen der Verknüpfung aller Typen zur Wiederverwendung und kann zum anderen für andere Bereiche des Modells als Anker genutzt werden.

ASS_ASSESSMENT_ELEMENT		
Feld	Befüllung	muss
is_a	Verbindung zu Exercise, Itemset oder Item	X
type	Typ: EXERCISE, ITEMSET oder ITEM	X

ASS_MATERIAL

Zusätzliches Material wie beispielsweise Beschreibungen, Quellen, Diagramme, Bilder und ähnliches, die die Bearbeitung des ASSESSMENT ELEMENTs unterstützen.

ASS_MATERIAL		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbeschreibung	
definition	Erläuterung	X
description	Weitere Elemente	
supplies	-> ASS_ASSESSMENT_ELEMENT	

ASS_ELEMENT_POOL

Hier wird die Verbindung hinterlegt, welche ASSESSMENT ELEMENTs zu welchem ASSESSMENT POOL gehören. Diese Zuordnung erlaubt die mehrfache Wiederverwendung auch in Sammlungen die unabhängig von einem konkreten Assessment sind.

ASS_ELEMENT_POOL		
Feld	Befüllung	muss
part_of	-> ASS_ASSESSMENT_POOL	X
contains	-> ASS_ASSESSMENT_ELEMENT	X

ASS_ASSESSMENT_POOL

Dies erlaubt die Sammlung von Aufgaben und Teilaufgaben unabhängig von einem konkreten Assessment. Diese Aufgaben können sich auf Klassenstufen aber auch auf übergreifende Aufgaben wie Abituraufgaben beziehen. Hier können weitere Attribute hinzugefügt werden, um den Zweck des Pools genauer zu charakterisieren.

ASS_ASSESSMENT_POOL		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbeschreibung	
definition	Erläuterung	X
description	Weitere Erläuterungen	

A.4.8 Modellbereich SCORE

SCO_MANIFEST_VARIABLE

Die **SCO_MANIFEST_VARIABLE** repräsentiert die Messung, also die Bewertung eines Items auf einer bestimmten Messskala unter dem Aspekt eines Skills. Die **MANIFEST VARIABLE** können somit zunächst aus den **ASSESSMENT ITEMS** erzeugt werden, da grundsätzlich jedes **ASSESSMENT ITEM**, das Bestandteil eines **ASSESSMENT** ist, auch gemessen werden soll.

SCO_MANIFEST_VARIABLE		
Feld	Befüllung	muss
uid	Freie Vergabe, kann uid in ASS_ASSESSMENT_ITEM entsprechen	X
measures	FK zu ASS_ASSESSMENT_ITEM	X
sname	Zusatzbeschreibung	
maximum	-> SCO_SCORE Maximaler Score über Fremdschlüssel	X
outoming_ space_type	Predefined oder ex-post im Sinn der „Four building blocks“	X
is_part_of	FK zu SCO_VARIABLE_GROUP	X

SCO_SCORE_GROUP

Gruppe von manifesten Variablen, die zu einer gemeinsamen Bewertung einer **SETTING EXECUTION** gehören, beispielsweise nach einer bestimmten Messskala. Auch inhaltliche Gruppierungen nach Kompetenzen, Skills oder anderen Kriterien können hier beschrieben werden.

SCO_SCORE_GROUP		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung	
definition	Angabe beispielsweise der Bewertung oder inhaltlichen Gruppierung	X
description	Zusatzbeschreibung	
based_on	-> ASS_SETTING_EXECUTION	X

SCO_MEASUREMENT_SCALE

Der Inhalt wird innerhalb des SCORE-Modells neu erstellt. Eine Skala erhält einen Namen, so dass bei jeder Bewertung eine eindeutige Zuordnung der verwendeten Skala erfolgen kann. Es sind keine Voraussetzungen über die Messbarkeit, also den Skalentyp getroffen, dieser soll hier aber angegeben werden, um eine adäquate Nutzung der damit ermittelten Bewertungen sicherzustellen. Es kann sich beispielsweise um die Noten oder Notenpunkte handeln.

SCO_MEASUREMENT_SCALE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Skalenbezeichnung	X
description	Zusatzbeschreibung	
skalentyp	Beispielsweise: Nominal, Ordinal, Differenz, Verhältnis, Absolut	X
maximum	Größte Ausprägung	X

SCO_SCORE

Dokumentation der möglichen Einzelbewertungen, beispielsweise eine mögliche Punktzahl für einen **SCO_RESPONSE_TYPE**. Die Werte müssen der zugeordneten Skala entsprechen. Auf einer klassischen Notenskala würden hier sechs Einträge für die Noten 1 bis 6 stehen.

SCO_SCORE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Bewertung, beispielsweise Punktzahl	X
definition	Bewertung in Langform als Text	X
description	Zusatzbeschreibung	
uses	->SCO_MEASUREMENT_SCALE	X

SCO_RESPONSE_TYPE

Hier werden die typisierten Antworten der **ASSESSMENT ITEMS** festgelegt. Diese Werte können als predefined-Werte bei der Konstruktion des Assessments festgelegt oder als ex-post-Werte erst nach der Durchführung als tatsächliche Antworten gekennzeichnet werden. Die predefined-Werte werden aus der Aufgabenstellung des Items abgeleitet. Zumindest eine richtige Antwort muss definiert sein. Die Anzahl falscher Antworten ist nicht festgelegt. Die ex-post Antworten können aus den tatsächlichen Antworten erzeugt werden, indem bisher fehlende Antwortmöglichkeiten ergänzt werden. Eine Typisierung im Sinne typischer Fehler ist nicht nur möglich, sondern wünschenswert, um ein effizientes Feedback erzeugen zu können.

SCO_RESPONSE_TYPE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
correct	T/F, markiert, ob es sich grundsätzlich um eine korrekte Antwort handelt	X
definition	Eigentlicher Antwortwert	X
description	Zusätzliche Beschreibung der Antwort	
outcoming_space_type	Predefined oder ex-post im Sinn der „Four building blocks“	X
belongs-to	->ASS_ASSESSMENT_ITEM, wird bei der Erstellung gefüllt	X
link_to_competence	->CSD_COMPETENCE_NIVEAU	
link_to_qualification	->QUA_QUALIFICATION_NIVEAU	

SCO_SCORING

Der Typ **SCORING** ist das Herzstück der manifesten Bewertung. Hier werden den einzelnen Antworten auf die einzelnen Items Scores, also Bewertungen zugeordnet. Im einfachsten Fall bedeutet dies eine typische Bepunktung der Antwort mit vollen oder Teilpunkten. Dabei wird jeweils einer bestimmten typisierten Antwort (**RESPONSE_TYPE**) eines Items (repräsentiert durch die **MANIFEST_VARIABLE**) ein **SCO_SCORE** bezogen auf einen **SKI_SKILL** zugeordnet.

SCO_SCORING		
Feld	Befüllung	muss
score_for	-> SCO_RESPONSE_TYPE, die betreffende Antwort	X
score_by	-> SCO_SCORE, der zugeordnete Score	X
measured_for	-> SCO_MANIFEST_VARIABLE, die Variable für die der Score gilt	X
related_to	-> SKI_SKILL Skill, auf den sich die Bewertung bezieht.	X

Die Befüllung der Tabelle kann manuell oder in einfachen Fällen durch eine Prozedur erfolgen, die basierend auf den vorhandenen Informationen, insbesondere Richtigkeit und Schwierigkeit aber auch anderen Informationen, die insbesondere im **RESPONSE TYPE** und in der **MANIFEST VARIABLE** vorhanden sind, einen Wert zuordnet.

SCO_VARIABLE_AGGREGATION_RULE

Regel für die Aggregation der Scores von **MANIFEST VARIABLE**s zum Score einer aggregierten **MANIFEST VARIABLE**. Im Regelfall wird es sich um eine einfache Summation handeln.

SCO_VARIABLE_AGGREGATION_RULE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Aggregationsoperation/-regel	X
definition	Definition der Aggregationsoperation	X
description	Zusatzbeschreibung	

SCO_AGGREGATE

Zuordnung von **MANIFEST VARIABLE** zu einer Gruppe, die selbst wieder eine **MANIFEST VARIABLE** ist.

SCO_AGGREGATE		
Feld	Befüllung	muss
belongs_to	-> SCO_MANIFEST_VARIABLE, Zuordnung zu einer untergeordneten Gruppe	X
part_of	-> SCO_MANIFEST_VARIABLE, Zuordnung zu einer übergeordneten Gruppe	X
controlled_by	-> SCO_VARIABLE_AGGREGATION_RULE Aggregationsvorschrift	X
weight	Gewichtung der Aggregation	X

A.4.9 Modellbereich STATISTIK

STA_STATISTICAL_MODEL_TYPE

Hier werden die Grundtypen der verwendeten statistischen Modelle hinterlegt, beispielsweise das hier verwendete RASCH-Modell aus dem Bereich der IRT. Es können grundlegende Eigenschaften und Einschränkungen dokumentiert werden.

STA_STATISTICAL_MODEL_TYPE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung des Modells	X
definition	Definition des Modells	
description	Zusatzbeschreibung	
statistic_type	Modelltyp CTT/IRT	X

STA_STATISTICAL_MODEL

Das Modell ist eine Instanz eines Modelltyps und verweist auf diesen als zugeordnetem Modelltyp. Es bezieht sich immer auf eine konkrete Durchführung eines Assessments, also eine **SETTING_EXECUTION**, für die umgekehrt mehrere Modelle existieren können. Hier werden alle Tests zusammengeführt und die Kompetenzschätzung auf eine latente Variable bezogen.

STA_STATISTICAL_MODEL		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung des Modells	X
definition	Definition des Modells	
description	Zusatzbeschreibung	
uses	-> STA_STATISTICAL_MODEL_TYPE, zugeordneter Modelltyp	X
is_part_of	-> ASS_SETTING_EXECUTION, zugeordnete Assessment-Ausführung	X
is_successor_of	-> STA_STATISTICAL_MODEL, eventuell vorher verwendetes Modell, auf dem dieses Modell basiert	
indicates	-> STA_LATENT_VARIABLE	X

STA_QUALITY_TEST

Hier werden statistische Testverfahren hinterlegt und gegebenenfalls beschrieben, die auf die verwendeten Modelltypen angewendet werden können.

STA_QUALITY_TEST		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung des Tests	X
definition	Definition der Funktion und Vorgehensweise	
description	Zusatzbeschreibung	
applicable	-> STA_STATISTICAL_MODEL_TYPE, Modelle, auf die der Test anwendbar ist	X

STA_APPLICABLE_TEST

Auflösung der Beziehung zwischen Modelltypen und Testverfahren, um die Anwendbarkeit verschiedener Testverfahren auf bestimmte Modelltypen zu beschreiben.

STA_APPLICABLE_TEST		
Feld	Befüllung	muss
applicable_to	-> STA_STATISTICAL_MODEL_TYPE	X
testable_by	-> STA_QUALITY_TEST	X

STA_STATISTICAL_QUALITY_PARAMETER_LR

Dokumentation der Parameter für die tatsächliche Durchführung des LR-Tests nach Anderson in einem statistischen Modell.

STA_STATISTICAL_QUALITY_PARAMETER_LR		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung	
description	Weitere Erläuterungen	
software	Hinweise zur verwendeten Software	
described_by	-> STA_STATISTICAL_QUALITY_TEST, zugrundeliegender Test	X
applies_to	-> STA_STATISTICAL_MODEL, Model für das der Test durchgeführt wurde	X
LR-Value	χ^2 -Wert, entsprechende den Freiheitsgraden und dem p-Wert	X
CHI-Square	Freiheitsgrade	X
p-Value	Testgröße, Modell ist valide, wenn sie über einem Fehlerniveau liegt, beispielsweise 0,05	X

STA_STATISTICAL_QUALITY_PARAMETERS_WALDTEST

Dokumentation der Parameter für die tatsächliche Durchführung des Waldtests in einem statistischen Modell.

STA_STATISTICAL_QUALITY_PARAMETER_WALDTEST		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung	
description	Weitere Erläuterungen	
software	Hinweise zur verwendeten Software	
described_by	-> STA_STATISTICAL_QUALITY_TEST, zugrunde liegender Test	X
applies_to	-> STA_STATISTICAL_MODEL, Model für das der Test durchgeführt wurde	X
applies_to_variable	-> STA_STATISTICAL_VARIABLE, betroffene Variable	X
z-statistics	z-Wert	X
p-value	Testgröße, Variable ist valide, wenn sie über einem Fehlerniveau liegt, beispielsweise 0,05	X

STA_STATISTICAL_QUALITY_PARAMETERS_GRAPHICS

Dokumentation der Parameter für die tatsächliche Durchführung des graphischen Abweichungstests in einem statistischen Modell.

STA_STATISTICAL_QUALITY_PARAMETER_GRAPHICS		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung	
description	Weitere Erläuterungen	
software	Hinweise zur verwendeten Software	
described_by	-> STA_STATISTICAL_QUALITY_TEST, zugrundeliegender Test	X
applies_to	-> STA_STATISTICAL_MODEL, Modell für das der Test durchgeführt wurde	X
graphic	Graphische Darstellung zur Auswertung	

STA_LATENT_VARIABLE

Beschreibung einer latenten Variablen. Die Variable kann verschiedenen statistischen Modellen zugeordnet werden, in denen ihre Werte bestimmt werden. Die Variable kann unmittelbar zur Indikation eines Skills verwendet werden.

STA_LATENT_VARIABLE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung der Variablen	
definition	Definition der zu messenden Eigenschaft	X
description	Zusätzliche Beschreibungen	
indicates_skill	-> SKI_SKILL	

STA_STATISTICAL_VARIABLE

Beschreibung einer statistischen Variablen. Diese Variable ist immer einer manifesten Variable zugeordnet und enthält zusätzlich statistische Schätzparameter wie beispielsweise die Schwierigkeit/Einfachheit, wie sie im zugehörigen statistischen Modelle geschätzt wurde. Die statistical_easyness ist die Umkehrung des Schwierigkeitsgrades, der beim RASCH-Modell auch durch das Modell bestimmt werden kann.

STA_STATISTICAL_VARIABLE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung der Variablen	
measured_by	-> STA_STATISTICAL_MODEL, Modell mit dem der Messwert ermittelt wurde	X
statistical_easyness	Wert, der der Variablen durch das Modell zugeordnet wurde, hier aus eRM Modul aus R direkt übernommen	X
indicated_by	-> SCO_MANIFEST_VARIABLE, manifeste Variable oder Gruppe, die dieser Variablen entspricht.	X

A.4.10 Modellbereich PERSONAL

PERS_GROUP

Eine Gruppe von Personen, die an einem Assessment teilnimmt, beispielsweise ein Kurs oder eine Klasse.

PERS_GROUP		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Benennung der Lerngruppe	X
definition	Definition der Lerngruppe	
description	Weitere Erläuterungen	

PERS_PERSON

Hier werden die persönlichen Daten der Assessment-Teilnehmer gespeichert, die für die Analyse relevant sind. Da es sich um potenziell unter Datenschutzaspekten sensible Informationen handelt, die aber andererseits für Analysen relevant sein können, ist eine Trennung von den persönlichen Daten vorgenommen worden, die für die Analyse nicht unmittelbar relevant sind. Hier dürfen also nur anonyme Daten abgelegt werden. Sollen weitere Aspekte in die Analyse einbezogen werden, kann der Typ entsprechend erweitert werden.

Der „sname“ kann als Kurzbezeichnung oder Pseudonym einer Person verwendet werden, um die Analyse zu personalisieren. Es kann ein Link zu den tatsächlichen persönlichen Daten gespeichert werden, die nicht zwingend im selben System sein müssen. Schließlich kann die **PERSON** einer **GROUP** zugeordnet werden.

PERS_PERSON		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung der Person, Pseudonym	
definition	Definition	
description	Weitere Erläuterungen	
has_personal_data	-> PERS_PERSONAL_DATA	
part_of	-> PERS_GROUP	

PERS_PERSONAL_DATA

Dieser Typ muss nicht im CSP-Modell gespeichert sein. Er kann in einem externen Verwaltungssystem liegen. Er dient der Speicherung persönlicher nicht unmittelbar für das Assessment relevanter Informationen zu einer tatsächlichen Person.

PERS_PERSONAL_DATA		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung der Person, Pseudonym	
definition	Definition	
description	Weitere Erläuterungen	
lastname	Nachname	
firstname	Vorname	

PERS_PERSONAL_ITEM_RESPONSE

Dieser Typ speichert die ursprünglichen Originalantworten und die als tatsächlich beabsichtigt rekonstruierten Antworten. Die Antwort wird dem Assessment-Teilnehmer, der Durchführung und der für die Messung des Items relevanten manifesten Variable zugeordnet.

PERS_PERSONAL_ITEM_RESPONSE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung	
definition	Definition	
description	Weitere Erläuterungen	
original_answer	Originalantwort	X
answer	Als beabsichtigt eingestufte Antwort	X
belongs_to	-> PERS_PERSON	X
abstract_link	-> SCO_MANIFEST_VARIABLE	X
setting_link	-> ASS_SETTING_EXECUTION	X

PERS_PERSONAL_ITEM_RESPONSE_TYPE

Hier werden die im Rahmen des Outcome Modeling (Abschnitt 7.11) typisierten Antworten gespeichert. Die Zuordnung erfolgt prinzipiell wie beim PERSONAL ITEM RESPONSE. Dies ist die abschließende Dokumentation der beobachteten Antworten im Assessment.

PERS_PERSONAL_ITEM_RESPONSE_TYPE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung	
definition	Definition	
description	Weitere Erläuterungen	
answer	Typisierte Antwort	X
belongs_to	-> PERS_PERSON	X
abstract_link	-> SCO_MANIFEST_VARIABLE	X
setting_link	-> ASS_SETTING_EXECUTION	X

PERS_PERSONAL_ITEM_SCORE

Hier werden die Bewertungen der beobachteten Antworten gespeichert, die mittels des SCORING im Modellbereich SCORE vorgenommen werden. Hier, wie auch in allen anderen Bewertungstabellen wird ein Maximalwert mitgeführt. Dieser wird pro Person hinterlegt, da er sich unter Umständen in einem Assessment beispielweise wegen eines Nachteilsausgleichs für gehandicapte Personen unterscheiden kann.

PERS_PERSONAL_ITEM_SCORE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung	
indicator	Indikator, der das Scoring beinhaltet	X
maximum	Maximal erreichbares Scoring	X
group	Zugehörige Variablengruppe	X
belongs_to	-> PERS_PERSON	X
abstract_link	-> SCO_MANIFEST_VARIABLE	X
setting_link	-> ASS_SETTING_EXECUTION	X

PERS_PERSONAL_AGGREGATED_SCORE

Hier können für aggregierte manifeste Variablen Scores und das jeweils erreichbare Maximum als materialisierte Views gespeichert werden. Dies dient als Hilfe für weitere Analysen.

PERS_PERSONAL_AGGREGATED_SCORE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung, beispielsweise Skill oder Fact-Beschreibung	
group_id	Zuordnung der Gruppe, kann aus MANIFEST VARIABLE kommen	X
score	Bewertung	X
maximum	Maximal erreichbare Bewertung	X
belongs_to	-> PERS_PERSON	X
abstract_link	-> SCO_MANIFEST_VARIABLE	X
setting_link	-> ASS_SETTING_EXECUTION	X

PERS_PERSONAL_MANIFEST_SCORE

Tatsächliche manifeste Gesamtbewertung für eine Person in einer Assessment-Ausführung. Es kann bei unterschiedlichen Bewertungen über die group_id zwischen den unterschiedlichen Bewertungen unterschieden werden.

Außerdem kann dieser Typ für die direkte Dokumentation extern durchgeführter Assessments, beispielsweise konventioneller Klausuren genutzt werden, deren Ergebnisse zu Vergleichszwecken dokumentiert werden sollen. Die Unterscheidung kann über die **SETTING EXECUTION** erfolgen.

PERS_PERSONAL_MANIFEST_SCORE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung	
group_id	Zuordnung der Gruppen für die der Score gilt, 0 ist Basis	X
score	Score	X
maximum	Maximal erreichbares Scoring	X
belongs_to	-> PERS_PERSON	X
setting_link	-> ASS_SETTING_EXECUTION	X

PERS_PERSONAL_TRAIT_SCORE

Tatsächliche kompetenzorientierte Gesamtbewertung für eine Person in einer Assessment-Ausführung. Es kann unterschiedliche Bewertungen geben, wenn unterschiedliche statistische Modelle eingesetzt wurden oder das Modell bei mehrmaliger Anwendung wegen nicht-deterministischer Komponenten in der Durchführung zu unterschiedlichen Schätzungen geführt hat. Jedes statistische Modell wird durch die zugehörige latente Variable bestimmt, so dass hier eine Verbindung zu dieser Variable hergestellt wird.

PERS_PERSONAL_TRAIT_SCORE		
Feld	Befüllung	muss
uid	Freie Vergabe	X
sname	Kurzbezeichnung	
score	Scoring gemäß statistischem Modell, also der Trait	X
belongs_to	-> PERS_PERSON, wird automatisch übernommen	X
setting_link	-> ASS_SETTING_EXECUTION	X
abstract_link	-> STA_LATENT_VARIABLE	X

B. Fehleranalyse

Das Modell bietet durch die feine Dokumentation grundsätzlich auch Ansatzmöglichkeiten für eine detaillierte Fehleranalyse. Dies sollte aber nur der punktuellen Ergänzung von Ergebnissen dienen, um nicht in die Problematik behavioristischer Ansätze zu kommen.

B.1 Analyse der Kombination zweiter Items

Die Möglichkeit einer individuellen Fehleranalyse soll am Beispiel von zwei Items gezeigt werden, die beide zum Fakt 1107 beitragen und den Aspekt der Bedeutung des Parameters a in der Exponentialfunktion $f(x) = a \cdot b^x$ als Anfangswert zum Zeitpunkt $x=0$ für das exponentielle Wachstum haben. Dazu sind die beiden Items EX15.1 und W103.1 ausgewählt worden.

7. Die Gewapa GmbH beschreibt ihren Umsatz mit der Gleichung $U(x) = 300000 \cdot 1,08^x$. [W103]

	ja	nein
Der Umsatz im Jahr 0 liegt bei 324000,-	<input type="radio"/>	<input type="radio"/>
Das jährliche Wachstum liegt bei 8%.	<input type="radio"/>	<input type="radio"/>
Es kommt jährlich absolut gesehen immer derselbe Umsatz hinzu.	<input type="radio"/>	<input type="radio"/>
Es handelt sich um lineares Wachstum.	<input type="radio"/>	<input type="radio"/>

Abbildung 85 Item W103

16. Von einer Insektenart sind am Anfang 5 Exemplare vorhanden. Nach 3 Zeiteinheiten sind es 320 Exemplare. Das Wachstum soll durch eine Exponentialfunktion $f(x) = a \cdot b^x$ beschrieben werden. Bestimmen Sie die Funktionsgleichung. [EX15]

$f(x) =$ \cdot x

Abbildung 84 Item EX15

Das Item W103 bietet die fertige Exponentialfunktion $U(x) = 300000 \cdot 1,08^x$ an und erfordert als W103.1 die Entscheidung, ob die Aussage „Der Umsatz im Jahr 0 liegt bei 324,000 richtig oder falsch ist .

Das Item EX15 beschreibt in einem Text eine Situation zum exponentiellen Wachstum und erfordert die Eingabe der Parameter a und b der Exponentialfunktion. Die Funktion ist strukturell vorgegeben, es muss auch keine Entscheidung über die Art des Wachstums getroffen werden.

Das Item EX15.1 beschreibt als erstes Eingabefeld speziell den Parameter a . Beide Items erfordern die Bedeutung des Parameters a zu kennen, das zweite Item erfordert zusätzlich aus dem Text heraus den Wert zu ermitteln und einzusetzen.

W103.1/EX15.1	Richtig	Falsch	
Richtig	47	10	59
Falsch	15	2	17
	64	12	76

Tabelle 33 Vergleich W103.1 und EX15.1

In Tabelle 33 sind die richtigen und falschen Antworten für beide Items als Zeilen respektive als Spalten dargestellt. Insgesamt wurde also W103.1 zu 78% richtig beantwortet und EX15.1 zu 84%. Man sieht außerdem, dass die Wahrscheinlichkeit EX15.1 richtig zu beantworten, wenn man

W103.1 richtig beantwortet hat bei $P_{W103.1}(EX15.1) = \frac{47}{59} = 0,80$ recht hoch ist, wie auch umgekehrt die Wahrscheinlichkeit W103.1 richtig zu beantworten, wenn EX15.1 richtig beantwortet wurde ($P_{EX15.1}(W103.1) = \frac{47}{64} = 0,73$). Umgekehrt sollte man, wenn es nur um die Bedeutung von a geht, erwarten, dass Assessment-Teilnehmer, die bereits W103.1 falsch beantworten auch bei EX15.1 einen Fehler machen, da hier noch zusätzlich der Wert aus dem Text extrahiert werden muss.

14. Ein Taubenzüchter hat einen Taubenschlag, der ganzjährig bewohnbar ist. Dir Fütterung der Tauben erfolgt mit Mais und Weizenkörnern im Verhältnis 50:50. Seine Taubenzucht bestand vor drei Jahren aus 40 Tauben. Dieses Jahr hat er 135 Tauben gezählt, die jährlich etwa 200kg Futter fressen. [EX16]

Geben Sie eine Funktionsgleichung an, die die Zunahme an Tauben richtig beschreibt.

f(x) =

Abbildung 86 Item EX16.1

answer	belongs_to	abstract_link
-9	106	EX15.1
40*^t	106	EX16.1
1	106	W103.1
320	107	EX15.1
200/135*40	107	EX16.1
1	107	W103.1
320	110	EX15.1
40*1,3	110	EX16.1
1	110	W103.1
320	112	EX15.1
-9	112	EX16.1
1	112	W103.1
3	114	EX15.1
40*1,3	114	EX16.1
2	114	W103.1
315	117	EX15.1
200*54^t	117	EX16.1
2	117	W103.1
4	127	EX15.1
40*1,5^3	127	EX16.1
1	127	W103.1
0	137	EX15.1
135x*200^x	137	EX16.1
1	137	W103.1
3	67	EX15.1
40*2^135*x	67	EX16.1
1	67	W103.1
-9	70	EX15.1
-9	70	EX16.1
1	70	W103.1
-9	71	EX15.1
-9	71	EX16.1
1	71	W103.1
320	84	EX15.1
40*0,97^3	84	EX16.1
1	84	W103.1

Abbildung 87 Ergebnisse für Item EX16.1 mit richtiger Antwort auf Item W103.1

Tatsächlich zeigt Tabelle 33, dass selten beide Items falsch beantwortet wurden, sondern die falschen Antworten auf jeweils ein Item überwiegen. Auffällig ist auch, dass das als schwieriger eingeschätzte Item EX15.1 öfter richtig beantwortet wurde. Somit scheint in W103.1 ein weiterer bisher unbekannter Schwierigkeitsgrad zu liegen.

Führt man die Analyse weiter, zeigt sich, dass von den 15 falschen Antworten für W103.1 bei denen EX15.1 richtig beantwortet wurde, auch in EX16.1 wo eine noch komplexere Beschreibung zur Bestimmung des Anfangswertes gegeben ist, 12 richtige Antworten vorhanden sind. Dies legt den Schluss nahe, dass das Problem hier weder das Verständnis des Parameters a als solches noch das Verständnis des Textes ist. Außerdem haben bei der Angabe des Wachstums in Höhe von 8% in W103.2 von diesen 15 wiederum 14 richtig geantwortet. Also besteht auch grundsätzlich die Fähigkeit, die Funktion zu verstehen. Die Hypothese wäre, dass das Einsetzen von $x=0$ fehlerhaft war, man also entweder die Umsetzung des Startzeitpunktes auf den Wert 0 oder das Einsetzen als solches nicht verstanden hat. Diese ist zu prüfen.

Ein Indiz dafür, dass die These, dass das Einsetzen eines Wertes kein generelles Problem darstellt, liefert das Item W104.1.

8. Die Gewapa GmbH beschreibt ihren Umsatz mit der Gleichung $U(x) = 300000 * 1,08^x$. x ist in Jahren mit $x=0$ für das Jahr 2010 angegeben. [W104]

Der Umsatz nach 4 Jahren beträgt (gerundet in ganzen Euro). Euro

Bestimmen Sie die Anzahl Jahre (ganzzahlig) nach denen sich der Umsatz um mindestens 50% erhöht hat. Jahre

Bestimmen Sie den Umsatz für 2009, wenn er sich von 2009 auf 2010 auch um 8% erhöht hätte (gerundet in ganzen Euro). Euro

Abbildung 88 Item W104

Hier soll der Wert $x=4$ eingesetzt werden. Sieht man von Rundungsfehlern ab, gelingt dies 12 der betrachteten Assessment Teilnehmer, die $x=0$ in W103.1 nicht verwendet haben. Dies deutet darauf hin, dass eher das Erkennen von $x=0$ als Anfangswert problematisch war. Ein weiteres Indiz in diese Richtung kann EX14.1 liefern, wo ebenfalls $x=0$ in die Funktion einzusetzen ist. Die Aufgabe wurde von 50% der Assessment-Teilnehmer gelöst, von den 15 Teilnehmern hier aber nur von 6 Teilnehmern, was zumindest deutlich unter den 50% liegt. Somit kann man erkennen, dass eine Fehleranalyse unter Einbeziehung mehrerer Items zumindest im Bereich der Plausibilisierung möglich ist. Beide W103.1 und EX14.1 beruhen im Übrigen im Modell auf demselben SNIPPET SWE25 „Anfangswert a ist Funktionswert an der Stelle 0“.

Jetzt sollen die 10 Antworten in Tabelle 33 analysiert werden, bei denen W103.1 richtig, EX15.1 aber falsch beantwortet wurde. Hier ist die These, dass die Schwierigkeit in der Extraktion des Wertes aus dem Text gelegen hat. Diese wird gestützt von der Tatsache, dass nur noch bei 5 Teilnehmern EX16.1 hinsichtlich des Anfangswertes richtig. Hier ist es auch interessant, das Item W107.1 zu betrachten, das die Möglichkeit des Ablesens aus einer Formel mit der Entnahme der relevanten Information aus einem Text kombiniert.

12. Die Messung der Vermehrungsrate von Forellen hat folgenden funktionalen Zusammenhang ergeben [W107]

$$f(t) = 110 \cdot 1,10^t$$

Bei der Übertragung der Formel auf einen Teich mit 15% Wachstumsrate und 200 Fischen Anfangsbestand ergäbe sich dann (ganzzahlig und bei Bedarf mit zwei Nachkommastellen):

f(t) = * ^t

Abbildung 89 Item W107

Stellt man W107.1 und EX15.1 gegenüber, ergibt sich die Situation, die in Tabelle 34 dargestellt ist.

W107.1/EX15.1	Richtig	Falsch	
Richtig	57	12	69
Falsch	7	0	7
	64	12	76

Tabelle 34 Kombination Item W107.1 und Item EX15.1

Das Bild bei der Gegenüberstellung von EX15.1 mit W101.3 in Tabelle 33 und mit W107.1 Tabelle 34 ist sehr ähnlich. In beiden Fällen gibt es einen hohen Anteil gemeinsamer Lösungen, was auch zu erwarten war. Interessant ist, dass keine der 10 Personen mit einer falschen Lösung bei EX15.1 und einer richtigen Lösung bei W103.1 eine falsche Lösung in W107.1 hat. Alle sind richtig. Das legt die These nahe, dass die zusätzliche Hilfe in Form einer analogen Formel das Verständnis erleichtert hat, dass also die Präsentation eines Beispiels die Schwierigkeit des Ablesens aus dem Text überwunden hat.

B.2 Analyse eines einzelnen offenen Items

Am Beispiel des Items EX16.1 lässt sich erkennen, dass gerade offene Fragestellungen viele Möglichkeiten der Interpretation von Lösungen bieten. EX16.1 wurde schon in obigen Vergleichen einige Male herangezogen und ist in Abbildung 86 dargestellt. Gleiche Antworten sind dabei gezählt und die Anzahl als Häufigkeit angegeben worden.

Antworten	Häufigkeit
$40 \cdot 1,5^x$	18
$40 \cdot 1,3^x$	1
$40 \cdot 1,5^3$	8
$40 \times 1,50^3$	1
$40 \cdot 3^x$	2
$40 \cdot 2,9^x$	1
$40 \cdot 1,54^x$	1
$40 \cdot 337,5^t$	2
$40 \cdot t$	1
$40 \cdot 1,55^3$	1
$40 \cdot 0,97^3$	1
$40 \cdot 2^{135 \cdot x}$	1
$40 \cdot 1,3$	2
$40 \cdot 1,$	1
$40 \cdot$	3
$135 \cdot 1,5^x$	1
$135x \cdot 200^x$	1
$135 \cdot 1,88^{(x/3)}$	1
$26,7 \cdot 1,05^x$	1
$135 \cdot 7,0^t$	1
$200 \cdot 54^t$	1
$40 + 1,5^x$	2
$40 + 95^x$	2
$135 + 200^x$	1
$40 + 1,12^x$	1
$40 + 32x$	3
$40 + 31,67^x$	1
$40 \cdot 3,35t$	1
$40 \cdot (9/11)x + 40$	1
$40 \cdot 200x + 95$	1
$40 \cdot (135 + 200) + 40$	1
$200/135 \cdot 40$	1
0	1
-9	10
Summe	76

Alle Antworten sind in Tabelle 35 mit Häufigkeiten dargestellt. Das sind 34 verschiedene **RESPONSE TYPES**. Der erste **RESPONSE TYPE** repräsentiert die richtige Antwort. Die übrigen 33 weisen auf verschiedene Fehler hin, wobei in einer Antwort auch mehrere Fehler stecken können.

Die Antworten sind sortiert worden, wobei eine Sortierung nach der Art des Fehlers erfolgt ist. So sind zunächst die Antworten mit dem im Prinzip richtigen Funktionstyp aufgeführt. Innerhalb dieser Antworten sind diejenigen mit dem richtigen Anfangswert nach oben sortiert. Dann geht es noch um den Wachstumsfaktor, den Namen der Variablen und insbesondere die Verwendung einer Variablen statt eines festen Wertes.

Gerade der letztgenannte Aspekt hätte auch als Sortierkriterium verwendet werden können, geht es doch hier um eine grundlegende Eigenschaft einer Funktion.

Als nächstes sind die Antworten mit richtigem Funktionstyp und falschem Anfangswert aufgeführt. Danach sind die Antworten mit falschem Funktionstyp, im Allgemeinen mit einer linearen Funktion aufgeführt. Hier zeigen sich auch wieder einige Antworten mit scheinbar unterschiedlichem Verständnisgrad.

Schließlich sind reine Zahlenangaben aufgeführt, bei denen eine Analyse schon schwerfällt.

Die Angabe -9 steht dafür, dass keine Antwort gegeben wurde.

Tabelle 35 Antworten für EX16 mit absoluten Häufigkeiten

Klassifiziert man die Fehler nach relevanten Fehlertypen, so kann man als Kriterium beispielsweise die in Tabelle 36 gewählten Kriterien verwenden.

Kriterium	Erfüllt	Nicht erfüllt	unspezifisch
Richtiger Funktionstyp	44	21	11
Variable verwendet	46	19	11
Richtiger Anfangswert	56	8	12
Richtiger Wachstumsfaktor	30	30	16

Tabelle 36 Vier verschiedene Kategorien von Fehlerursachen

Man erkennt, dass etwa 58% der Assessment-Teilnehmer eine richtige Vorstellung von dem benötigten Funktionstyp haben, 28% haben eine falsche Vorstellung, sehen diese aber als richtig an. Etwa dieselben Werte ergeben sich für die Erkenntnis, dass bei Angabe einer Funktion die Variable verwendet werden muss und nicht ein (wahrscheinlich zur Bestimmung des Wachstumsfaktors) verwendeter Zahlenwert. Der Anfangswert kann zu 74% richtig bestimmt werden, der Wachstumsfaktor dagegen nur zu 39%. So können gezielt Ursachen aus einem Item herausgefiltert werden, die sich auf der Ebene von **SNIPPETs** oder **FACTs** bewegen.

Über eine solche Einteilung hinaus können aber auch mögliche Fehlerursachen analysiert und den **RESPONSE TYPEs** zugeordnet werden. So zeigt Tabelle 37 eine Auswahl möglicher Fehlerursachen und die entsprechenden **RESPONSE TYPEs**. Die entsprechende Zuordnung erfolgt im Modell mittels der Typen **FBCK_ERROR_TYPE** und dem verbindenden **FBCK_RESPONSE_ERROR**. Eine ausführliche Übersicht ist in Tabelle 38 enthalten.

Vermute Ursache	Betroffene Response Types
3 Jahre werden als Wachstumsfaktor genutzt	$40 \cdot 1,3^x$; $40 \cdot 3^x$; $40 \cdot 1,3$
Linearer Ansatz $135 = 40 + b \cdot 3$	$40 + 31,67x$; $40 + 32x^{16}$
Vertauschung der Werte, Ansatz $40 = 135 \cdot b^3$	$40 \cdot 1,3^x$, $40 \cdot 1,3$; $40 \cdot 0,97^{3^{17}}$
Aufgabenstellung falsch verstanden, Addition von Tauben und Futter	$40 \cdot 3,35t$; $40 \cdot (135 + 200) + 40^{18}$
Aufgabenstellung falsch verstanden, 200kg Futter direkt verwendet	$135x \cdot 200^x$; $200 \cdot 54^t$; $135 + 200^x$; $40 \cdot 200x + 95^{19}$; $40 \cdot (135 + 200) + 40$; $200/135 \cdot 40$

Tabelle 37 Fehler und mögliche Ursachen bei EX16

Bei der Analyse möglicher Fehlerursachen können sich auch Situationen zeigen, bei denen man bei einzelnen Ergebnissen ganze Fehlerketten vermuten kann. So kann der **RESPONSE TYPE**

$$135 \cdot 1,88^{(x/3)}$$

zunächst mit der Annahme des falschen Anfangswertes 135 erklärt werden. Bei der Wahl des Exponenten wird aber scheinbar eine Art „Multiplikations/Divisions-Logik für die drei Jahre“ verwendet. Dafür spricht die Angabe „ $x/3$ “. Dann könnte es dem Assessment-Teilnehmer logisch erschienen sein, 40 zu verdreifachen, da die Jahre durch 3 geteilt wurden. Wenn 135 als Anfangswert genutzt wird, könnte er $120 = 135 \cdot b^{(x/3)}$ als Ansatz zur Bestimmung von b gewählt haben. $120/135$ ist 0,88. Wurzeln werden nicht gezogen, da mit Multiplikation/Division gearbeitet wird. Störend wäre dann, dass 0,88 kleiner als 1 und damit offensichtlich falsch ist. Dieses Problem kann man „beheben“, wenn man berücksichtigt, dass die Kaninchen schon „da“ sind, also wie bei Kapital und Verzinsung 1 addiert werden muss. Somit kommt man zum Wachstumsfaktor 1,88. Man sieht hier eine ganze Kette von Fehlentscheidungen und somit Ursachen, die die zunächst eher kurios erscheinende Antwort erklären können.

¹⁶ Mit entsprechender Rundung

¹⁷ Mit fehlender Variablen, insbesondere aber Fehler bei Interpretation als abnehmender Prozentsatz

¹⁸ Mit Fehler des linearen Ansatzes, fehlender Variablen und doppelter Verwendung des Anfangswertes

¹⁹ 95 als Differenz der Tauben („ $135 - 40 = 95$ “)

B.3 Analyse einzelner Assessment-Teilnehmer

Eine andere Perspektive der Analyse kann in der Konzentration auf einzelne Assessment-Teilnehmer und die Wiederkehr ähnlicher Fehlertypen und somit auf einen verfestigten Kompetenzmangel liegen. So kann beispielsweise die fehlerhafte Bestimmung des Wachstumsfaktors in EX16 in der Praxis oft auf fehlende Fähigkeiten in der algebraischen Umformung zurückgeführt werden, also auf aus Sicht der Oberstufe elementare Probleme. Das Assessment sollte nicht primär derartige elementare Umformungen testen, allerdings sind diese elementaren Fähigkeiten oft die Prämisse für eine erfolgreiche Bearbeitung. Daher ist eine andere elementare Fähigkeit aufgenommen worden, die auch bei der Bestimmung des Wachstumsfaktors eine Rolle spielt, die Umwandlung einer Prozentangabe in eine Dezimalzahl. Diese soll hier ebenfalls analysiert werden, wobei diesmal der Aspekt noch detaillierter auf die Zusammenhänge von Lösungen verschiedener Items bei einzelnen Assessment-Teilnehmern gelegt wird.

So ist interessant, dass der Assessment-Teilnehmer 84 seine Antwort „ $40 \cdot 0,97^3$ “ scheinbar auch mit falscher Anwendung der Prozentlogik ermittelt hat, indem er sowohl die Zehnerpotenz als auch die Vergrößerung nicht richtig interpretiert hat. Zu prüfen wären daher seine Kenntnisse der Prozentrechnung. Mit dem Item EL01 steht ein entsprechendes Item zur elementaren Prüfung der Kenntnisse bei der Umwandlung von Prozentwerten in Dezimalzahlen zur Verfügung, siehe Abbildung 91. Hier zeigt sich, dass dieser Assessment-Teilnehmer nur einer von zwei Teilnehmern ist, die dieses Item überhaupt nicht bearbeitet haben. Der andere ist Nummer 123 und hat seinerseits EX16.1 überhaupt nicht bearbeitet.

Bei Teilnehmer 85 zeigt sich in EX16.1 mit dem Wachstumsfaktor 1,05 statt 1,5 scheinbar auch ein Problem in der Prozentrechnung, das sich zumindest in der komplexeren Stellenlogik von EL01.4 ebenfalls wiederfinden lassen könnte. Tatsächlich enthält seine Antwort 0,05 statt 1,005 neben dem Fehler bei der Bestandserhaltung auch einen Fehler in der Dezimalstelle. Andere Prozentrechnungen sind zwar richtig, aber generell scheint hier eine Unsicherheit zu bestehen. Bleibt man im Bereich der Prozentrechnung, so sieht man im Vergleich von W107.2 und EL01, dass der Teilnehmer 85 EL01 zwar richtig gelöst hat, beispielsweise die Teilnehmer 66, 130 und 135 aber alle 0,15 statt 1,15 angegeben haben und sich bei 66 und 130 dieser Fehlertyp in den EL01 Antworten wiederholt beziehungsweise auch noch mit Verminderung vermischt. Bei 135 ist nur die verminderte Prozentzahl falsch. In Abbildung 90 sind die Ergebnisse dargestellt, Teilnehmer 101 ist als Referenz mit den richtigen Ergebnissen hinzugefügt.

Insgesamt erkennt man die Schwächen bei elementaren Aufgaben in komplexeren Zusammenhängen, wo sie nur als Fähigkeit mit eingebracht werden müssen oft wieder. Es lohnt sich daher noch einmal nur die elementaren Aufgaben EL01.1 bis EL01.4 unter dem Aspekt zu betrachten, inwieweit bereits auf dieser Ebene Schlüsse gezogen werden könnten.

101	EL01.1	1,04	EL01.2	1,1	EL01.3	0,96	EL01.4	1,005
123	EL01.1	-9	EL01.2	-9	EL01.3	-9	EL01.4	-9
130	EL01.1	0,96	EL01.2	0,9	EL01.3	1,04	EL01.4	-0,095
135	EL01.1	1,04	EL01.2	1,1	EL01.3	-1,04	EL01.4	1,005
66	EL01.1	0,04	EL01.2	0,1	EL01.3	-0,04	EL01.4	0,005
84	EL01.1	-9	EL01.2	-9	EL01.3	-9	EL01.4	-9
85	EL01.1	1,04	EL01.2	1,1	EL01.3	0,6	EL01.4	0,05

Abbildung 90 Ausgewählte Teilnehmerergebnisse für EL01

22. Sie haben eine Kapital von 20000 Euro. Geben Sie jeweils den Faktor (die Zahl) an mit dem das Kapital multipliziert werden muss, wenn der angegebene Zinssatz gilt, um das Kapital im nächsten Jahr auszurechnen.
[ELO1]

- bei 4%
- bei 10%
- bei -4%
- bei 0,5%

Abbildung 91 ITEM EL01

In Abbildung 92 sind die Ergebnisse aller vier Items von EL01 für ausgewählte Assessment-Teilnehmer wiedergegeben.

Teilnehmer	EL01_1	EL01_1A	EL01_2	EL01_2A	EL01_3	EL01_3A	EL01_4	EL01_4A
101	EL01.1	1,04	EL01.2	1,1	EL01.3	0,96	EL01.4	1,005
107	EL01.1	0,4	EL01.2	1	EL01.3	-0,4	EL01.4	0,05
113	EL01.1	0,04	EL01.2	0,1	EL01.3	-0,04	EL01.4	0,005
117	EL01.1	50	EL01.2	20	EL01.3	-50	EL01.4	400
130	EL01.1	0,96	EL01.2	0,9	EL01.3	1,04	EL01.4	-0,095
135	EL01.1	1,04	EL01.2	1,1	EL01.3	-1,04	EL01.4	1,005
95	EL01.1	800	EL01.2	2000	EL01.3	-800	EL01.4	100

Abbildung 92 Ausgewählte Teilnehmerergebnisse für EL01

Neben dem richtigen Musterergebnis des Teilnehmers 101 findet man verschiedene systematische Fehler. So verwendet die Person 107 generell eine Dezimalstelle zu wenig. Der Teilnehmer 113 hat durchgehend auf den Erhalt des Kapitals verzichtet. Der Assessment-Teilnehmer 117 versucht jeweils das Kapital zu bestimmen. Er berechnet dazu ein Prozent und teilt dann noch einmal durch den Prozentsatz. Der Teilnehmer 130 verwechselt Abnahme und Zunahme (und macht noch einen Fehler am Ende). Der Teilnehmer 135 beherrscht scheinbar die Zunahme des Kapitals, aber nicht die Abnahme. Bei Teilnehmer 95 finden sich durchgängig richtig berechnete Werte, allerdings für die Zinsen statt für das neue Kapital.

Insgesamt ergibt sich also die Möglichkeit der Analyse möglicher Fehlerursachen, indem die Ergebnisse der einzelnen Assessment-Teilnehmer analysiert werden. Wesentlich ist, dass diese dann mit dem CONTENT und den FEEDBACK-Modellen verbunden sind.

B.4 Nutzung des Individualfeedbacks

B.4.1 Feedback für die Assessment-Teilnehmer

Wenn die Möglichkeit einer detaillierten Fehleranalyse existiert, ist damit die Basis geschaffen, gezielte Maßnahmen zur Behebung dieser Schwächen zu ergreifen.

Dazu wurde wieder beispielhaft auf den FACT 1107 „Bedeutung der Parameter beim exponentiellen Wachstum“ zurückgegriffen, bei dem sich im Beispiel in Abbildung 73 deutliche Unterschiede zwischen den beiden Assessment-Teilnehmern gezeigt haben. Ein mögliches Ergebnis ist in Abbildung 93 dargestellt.

Fakt	Ursache	Vorschlag
1107	Der Grundaufbau einer Exponentialfunktion ist nicht bekannt.	Aufgaben zum Aufbau einer Exponentialfunktion bearbeite
1107	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) = a \cdot b^x$ a den Anfangswert darste	Aufgaben, die den Anfangswert bei exponentiellem Wachs
1107	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) = a \cdot b^x$ b den Wachstumsfaktor c	Aufgaben, die den Wachstumsfaktor bei exponentiellem W
1107	Die Bedeutung des Exponenten ist unklar. Vermutlich wird Periode falsch eingerechnet.	Aufgaben zum Aufbau einer Exponentialfunktion, insbeson
1107	Die Bedeutung der Parameter a und b als Anfangswert und Wachstumsfaktor ist unklar	Aufgaben mit Angabe von Anfangswert und Wachstumsfakt
1107	Logischer Fehler bei der Bestimmung der Basis	Aufgaben zur Ermittlung der Parameter für exponentielles'

Abbildung 93 Ausschnitt der Ermittlung der Fehlerursachen und Maßnahmen für Fakt 1107

Man sieht, dass für Probleme beim Verständnis des Fakt 1107 einige Fehlertypen als Ursache ermittelt und entsprechende Feedback-Vorschläge mit Maßnahmen dokumentiert wurden.

In einer weiteren Detaillierung wird jetzt von der Ebene der Fakten auf die Ebene der Skills verfeinert. In Abbildung 94 wird wieder der Fakt 1107 betrachtet und der Operator „bestimmen“ gewählt, so dass die Schwachpunkte bezüglich des Skills „F1107b“ (Nummer 39) dargestellt werden.

Skill	Ursache	Vorschlag
39	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) = a \cdot b^x$ a den Anfangswert darste	Aufgaben, die den Anfangswert bei exponentiellem
39	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) = a \cdot b^x$ b den Wachstumsfaktor c	Aufgaben, die den Wachstumsfaktor bei exponentie
39	Die Bedeutung des Exponenten ist unklar. Vermutlich wird Periode falsch eingerechnet.	Aufgaben zum Aufbau einer Exponentialfunktion, in
39	Der Grundaufbau einer Exponentialfunktion ist nicht bekannt.	Aufgaben zum Aufbau einer Exponentialfunktion be

Abbildung 94 Ausschnitt der Ermittlung der Fehlerursachen und Maßnahmen für Fakt 1107 bestimmen

Je nach Tiefe der Gliederung der Fakten in Skills kann es dabei zu deutlich weniger oder ähnlichen Vorschlägen wie bei den Fakten kommen.

1. Ein Taubenzüchter hat einen Taubenschlag, der ganzjährig bewohnbar ist. Die Fütterung der Tauben erfolgt mit Mais und Weizenkörnern im Verhältnis 50:50. Seine Taubenzucht bestand vor drei Jahren aus 40 Tauben. Dieses Jahr hat er 135 Tauben gezählt, die jährlich etwa 200kg Futter fressen.
Geben Sie eine Funktionsgleichung an, die die Zunahme an Tauben richtig beschreibt.

f(x) =

Abbildung 95 Beispiel-Item EX16.1 mit offener Eingabe einer Funktion

Eine noch einmal deutlich detailliertere Analyse kann auf Ebene der einzelnen Fakten und ihrer **RESPONSE TYPES** erfolgen. Dafür wird eine Fehleranalyse der einzelnen Antworten auf die Items vorgenommen und versucht, die Antworten auf bestimmte Fehlertypen (**ERROR TYPES**) zurückzuführen. Je offener ein Item ist, desto vielfältiger sind die zu erwartenden Antworten. Hier ist beispielhaft das Item EX16.1 gewählt worden. Das Item ist in Abbildung 95 dargestellt, wie es im Fragebogen für die Assessment-Teilnehmer enthalten ist. Man sieht, dass die Frage offen gestellt ist, es ist keine Wachstumsart vorgegeben, desgleichen kein Funktionstyp. Außerdem enthält der Text eine Reihe nicht benötigter Angaben.

Entsprechend der offenen Fragestellung ergeben sich eine ganze Reihe von Antworten, die in Tabelle 38 dargestellt sind. Die richtige Antwort ist in der ersten Zeile enthalten. Sie enthält auch keinen Fehlertypcode. Man erkennt auch, dass in diesem Fall nur die richtige Antwort „predefined“ ist, während alle anderen Antworten den tatsächlich im Assessment gegebenen Antworten entsprechen. Die insgesamt 37 falschen Antworten sind klassifiziert und mit Fehlercodes versehen worden. Zunächst ist aus Übersichtgründen nur ein Fehlercode angegeben worden. Manchen Antworten sind auch mehrere Codes zugeordnet worden, sodass entsprechend mehrere Einträge in der Tabelle vorhanden sind.

Korrekt	Antwort	Outcoming Space	Fehlertyp
T	$40 \cdot 1,5^x$	predefined	
F	0	ex-post	unbestimmt
F	$135 \cdot 1,5^x$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$135 \cdot 1,88^{(x/3)}$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$135 \cdot 1,88^{(x/3)}$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$135 \cdot 7,0^t$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$135 \cdot 7,0^t$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$135+200^x$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$135+200^x$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$135+200^x$	ex-post	expfunk+statt*
F	$135+200^x$	ex-post	Aufgabentext nicht verstanden
F	$135x \cdot 200^x$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$135x \cdot 200^x$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$135x \cdot 200^x$	ex-post	Aufgabentext nicht verstanden
F	$200 \cdot 54^t$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$200 \cdot 54^t$	ex-post	Aufgabentext nicht verstanden
F	$200 \cdot 54^t$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$200/135 \cdot 40$	ex-post	Funktionstyp exponentielles Wachstum unklar
F	$200/135 \cdot 40$	ex-post	Aufgabentext nicht verstanden
F	$26,67 \cdot 1,05^x$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$26,67 \cdot 1,05^x$	ex-post	Prozentwert und Prozentfaktor nachkommastellen
F	$26,7 \cdot 1,05^x$	ex-post	Prozentwert und Prozentfaktor nachkommastellen
F	$26,7 \cdot 1,05^x$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$26,67 \cdot 1,5^x$	ex-post	Exp-Funktion Wachstum Parameter a unklar
F	$40 \times 1,50^3$	ex-post	Wert statt Variable in Funktion
F	40^*	ex-post	Exp-Wachstum falsche Basis
F	$40 \cdot (135+200)+40$	ex-post	Funktionstyp exponentielles Wachstum unklar
F	$40 \cdot (135+200)+40$	ex-post	Aufgabentext nicht verstanden
F	$40 \cdot (9/11)x+40$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$40 \cdot (9/11)x+40$	ex-post	Funktionstyp exponentielles Wachstum unklar
F	40^*t	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$40 \cdot 0,97^3$	ex-post	Exp-Wachstum falsche Basis
F	$40 \cdot 0,97^3$	ex-post	Exp-Funktion Wachstum Parameter vertauscht
F	$40 \cdot 0,97^3$	ex-post	Wert statt Variable in Funktion
F	$40 \cdot 0,97^3$	ex-post	Negativer Prozentwert und Prozentfaktor
F	$40^*1,$	ex-post	Rechenfehler Auflösung nach b
F	$40^*1,3$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$40^*1,3$	ex-post	Exp-Funktion Wachstum Parameter vertauscht
F	$40^*1,3$	ex-post	Exponentialzeichen fehlt
F	$40^*1,3^x$	ex-post	Exp-Wachstum falsche Basis
F	$40^*1,3^x$	ex-post	Exp-Funktion Wachstum Parameter vertauscht
F	$40^*1,5^3$	ex-post	Wert statt Variable in Funktion
F	$40^*1,54^x$	ex-post	Rechenfehler Auflösung nach b
F	$40^*1,55^3$	ex-post	Rechenfehler Auflösung nach b
F	$40^*1,55^3$	ex-post	Wert statt Variable in Funktion
F	$40^*1,5^3$	ex-post	Wert statt Variable in Funktion
F	$40^*2,9^x$	ex-post	Rechenfehler Auflösung nach b
F	$40^*2^{\wedge}135^x$	ex-post	Exp-Wachstum falsche Basis
F	$40^*200x+95$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$40^*200x+95$	ex-post	Funktionstyp exponentielles Wachstum unklar
F	$40^*200x+95$	ex-post	Aufgabentext nicht verstanden
F	$40^*3,35t$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$40^*3,35t$	ex-post	Funktionstyp exponentielles Wachstum unklar
F	$40^*3,35t$	ex-post	Aufgabentext nicht verstanden
F	40^*3^x	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$40^*337,5^t$	ex-post	Rechenfehler Auflösung nach b
F	40^*x^3	ex-post	Exp-Wachstum falsche Basis
F	$40+1,12^*x$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$40+1,12^*x$	ex-post	Funktionstyp exponentielles Wachstum unklar
F	$40+1,12^*x$	ex-post	expfunk+statt*
F	$40+1,5^x$	ex-post	expfunk+statt*
F	$40+31,67^*x$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	$40+31,67^*x$	ex-post	Funktionstyp exponentielles Wachstum unklar
F	$40+32x$	ex-post	Funktionstyp exponentielles Wachstum unklar
F	$40+95^x$	ex-post	expfunk+statt*
F	$40+95^x$	ex-post	Exp-Funktion Wachstum Parameter b unklar
F	-9	ex-post	unbestimmt

Tabelle 38 Response Types für Item 16.1

Die angegebenen Fehlertypcodes beinhalten eine Beschreibung des Fehlers und einen Verweis auf eine mögliche Ursache, die dann wiederum als Feedback dokumentiert ist. So lässt sich wie in Abbildung 96 zu sehen ist, eine Beschreibung der Fehler erstellen. Entsprechend können wieder Vorschläge für Maßnahmen ergänzt werden-

Antwort	Fehlertyp	Ursache
135+200^x	Der Anfangswert wird addiert statt multipliziert	Der Grundaufbau einer Exponentialfunktion ist nicht bekannt.
135+200^x	Der Parameter b wird nicht als Wachstumswert interpretiert.	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) a \cdot b^x$ b den Wa
135+200^x	Der Parameter a wird nicht als Anfangswert interpretiert.	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) a \cdot b^x$ a den Anf

Abbildung 96 Analyse der Antworten

So lässt sich dann auch für einzelne Assessment-Teilnehmer eine Vorschlagsliste möglicher Maßnahmen erstellen. Eine Auswertung nur für die Person 109 bezogen auf das hier angesprochene Item ergibt das Ergebnis in Tabelle 39.

Attribut	Wert
Person	109
Item	EX16.1
Antwort	$40 \cdot 3^X$
Korrekt	F
Fehlertyp	Der Parameter b wird nicht als Wachstumswert interpretiert.
Ursache	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) a \cdot b^x$ b den Wachstumsfaktor darstellt.
Vorschlag	Aufgaben, die den Wachstumsfaktor bei exponentiellem Wachstum thematisieren.

Tabelle 39 Einzelvorschlag für eine Person

Dies ist jetzt zur Vereinfachung stark reduziert. Aus der Gesamtzahl der Antworten ergeben sich entsprechend mehr Vorschläge. Diesen können dann durch Dokumentation in **FEEDBACK MATERIAL** auch Hilfsmaterialien wie beispielsweise Übungsaufgaben, Anleitungen und ähnliche Hilfsmittel zugeordnet werden.

B.4.2 Feedback für die Lehrperson

Ein anderer Aspekt ist, dass auch die Lehrperson aus den zugeordneten Fehlertypen Rückschlüsse ziehen kann. In Tabelle 40 ist wiederum der Übersichtlichkeit wegen nur für ein Item eine Übersicht erstellt, von wie vielen Personen ein bestimmter Fehlertyp im Assessment gemacht wurde. Man sieht, dass gerade beim Wachstumsfaktor b im exponentiellen Wachstum noch Probleme bestehen. Immerhin 16 Assessmentteilnehmer haben Antworten gegeben, die zumindest auch auf diesen Fehler hinweisen.

Außerdem scheint in der Grundlogik, dass eine Funktionsgleichung eine unabhängige Variable enthält noch eine Problematik zu liegen, da dies bei 11 Personen als mögliche Fehlerursache dokumentiert ist. Dies zeigt, dass hier Probleme deutlich werden können, die mit dem Kern des Domäneninhaltes Wachstum und Exponentialfunktion gar nichts direkt zu tun haben, sondern auf grundlegendere Defizite verweisen.

Anzahl Personen	Item	Fehler
16	EX16.1	Der Parameter b wird nicht als Wachstumswert interpretiert.
11	EX16.1	Unbestimmt
11	EX16.1	Bei der Erstellung einer Funktionsgleichung wird statt der Variablen ein Wert eingesetzt.
9	EX16.1	Es wird ein falscher Funktionstyp dem exponentiellen Wachstum zugeordnet.
7	EX16.1	Der Parameter a wird nicht als Anfangswert interpretiert.
6	EX16.1	Rechenfehler Auflösung nach b.
6	EX16.1	Logischer Fehler bei der Basis
6	EX16.1	Der Anfangswert wird addiert statt multipliziert
2	EX16.1	Es wird keine Exponentialfunktion angegeben.
1	EX16.1	Umrechnung (EL01) funktioniert, aber negative Veränderung nicht berücksichtigt.
1	EX16.1	Falsche Anzahl Nachkommastellen bei Umrechnung in den Prozentfaktor

Tabelle 40 Anzahl Personen, die einen bestimmten Fehlertyp machen

Betrachtet man statt des einen Items alle Items mit den zugeordneten Fehlerursachen, so lässt sich eine Gesamtanzahl von Personen ermitteln, bei denen die Fehleranalyse zumindest einmal auf einen bestimmten Fehlertyp verweist. Ein Ausschnitt einer solchen Auswertung ist in Abbildung 97 dargestellt.

Anzahl	Fehlertypcode	Ursache
60	expwachsfalschFunktyp	Es ist nicht klar, dass zur Beschreibung exponentiellen Wachstums eine Exponentialfunktion verwendet wird.
51	expwachsbedeutunga	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) a \cdot b^x$ a den Anfangswert darstellt.
35	linwachsfalschFunktyp	Es ist nicht klar, dass zur Beschreibung linearen Wachstums eine lineare Funktion verwendet wird.
28	prozent1	Fehler bei der Umrechnung eines Prozentsatzes in einen Faktor, typischerweise 0,05 statt 1,05 bei 5% Wachstum
27	expwachsbedeutungparam	Die Bedeutung der Parameter a und b als Anfangswert und Wachstumsfaktor ist unklar
20	expwachsbedeutungb	Es ist nicht klar, dass bei einer exponentiellen Funktion $f(x) a \cdot b^x$ b den Wachstumsfaktor darstellt.
16	linwachsbedeutungb	Es ist nicht klar, dass bei einer linearen Funktion $f(x) ax+b$ b den Anfangswert darstellt.
13	prozent3	Bei Prozentumrechnungen werden die Nachkommastellen falsch verwendet, beispielsweise 1,05 statt 1,5 bei 50%
11	gleichungstattfunktion	Es wird eine Gleichung statt einer Funktionsgleichung erstellt. Der Unterschied ist scheinbar nicht klar.
11	prozent2	Positive Veränderungen werden negativ, beispielsweise 0,97 statt 1,03 oder umgekehrt berücksichtigt.
10	expwachsnproportional	Proportionalität beim exponentiellen Wachstum wird nicht erkannt

Abbildung 97 Häufigkeit aufgetretener Fehlerursachen (Ausschnitt)

Eine solche Liste ist für eine Lehrperson eine Form von Feedback, die bei der weiteren Gestaltung des Bildungsprozesses hilfreich sein kann.

Hier exemplarisch eine Reihe weiterer Auswertungen und Beispiele aus dem konkreten Assessment heraus entwickelt worden. Für den grundsätzlichen Nachweis, dass auf Basis eines solchen Modells ein Feedback sowohl für die Assessment-Teilnehmer als auch für die Lehrpersonen möglich ist, erscheint dies ausreichend.

C. Abbildungsverzeichnis

Abbildung 1 Erste Übersicht Bildungsprozess.....	9
Abbildung 2 Bildungsprozess mit Bewertung und Beurteilung	11
Abbildung 3 Zentrale Begriffe zur Zielbeschreibung.....	12
Abbildung 4 Kompetenz und Performanz (Gnahs 2007, 23).....	13
Abbildung 5 Kompetenz - Performanz - Qualifikation.....	17
Abbildung 6 Kompetenzen und Qualifikationen als Teil des Individuums.....	23
Abbildung 7 Modell der Performanzmessung mittels konkreter Situationen in Assessments.....	29
Abbildung 8 Validität nach (Wilson 2004, 10).....	35
Abbildung 9 Beispiel eines Items mit Antworten auf verschiedenen Niveaus (Hardy, et al. 2010, 119)	39
Abbildung 10 Ursprüngliche und von Anderson/Krathwohl revidierte Bloom'sche Taxonomie (ETH - Zürich nach Bloom ... 2015)	40
Abbildung 11 Kompetenzniveaudimensionen (Universität Tübingen 2012, 2)	41
Abbildung 12 Struktur des DFG Schwerpunktprogramms 1293 (Fleischer, et al. 2013, 7)	42
Abbildung 13 ICC-Beispiel.....	45
Abbildung 14 Graphischer Test für zwei Splitkriterien (Strobl 2010, 41).....	47
Abbildung 15 Assessment Triangle (Pellegrino, Chudowsky und Glaser 2001).....	49
Abbildung 16 Four Building Blocks (Wilson 2004b, 17).....	52
Abbildung 17 Beobachtungs-, Bewertungs-, und Beurteilungsprozess.....	55
Abbildung 18 MOF Metaebenen für CSP-Informationsmodelle	66
Abbildung 19 Metamodellebenen des CSP-Modells	67
Abbildung 20 Die drei Ebenen des CSP-Modells.....	68
Abbildung 21 Allgemeiner Teil des Modells.....	73
Abbildung 22 Die 9 Teilmodelle des CORE-Modells	74
Abbildung 23 Semiotisches Dreieck [nach (Wikipedia - Semiotisches Dreieck 2015)]	79
Abbildung 24 Zwei Elemente mit ihren Qualifiern	82
Abbildung 25 Fakt nach NIAM	83
Abbildung 26 Fakt in graphischer Darstellung.....	84
Abbildung 27 Ausschnittweise Modellierung von Wachstumsprozessen.....	85
Abbildung 28 Ausschnittweise Modellierung von Exponentialfunktionen	86
Abbildung 29 Modellbereich Content - deklaratives Wissen.....	88
Abbildung 30 Mapping der Metatypen in gängigen Prozessbeschreibungssprachen [nach (Shazad, Elias und Johannesson 2009, 181)}.....	89
Abbildung 31 Wissenstyp PROCEDURE beispielhaft modelliert	90
Abbildung 32 Modellbereich Item	93
Abbildung 33 Competence Niveau Map (CNM) schematisch.....	96
Abbildung 34 Modellbereich Skill.....	96
Abbildung 35 Modellbereich Competence Specifications	98
Abbildung 36 Modellbereich Competence Dimension	103
Abbildung 37 Modellbereich Feedback.....	106
Abbildung 38 Bereiche des SITUATION-Modells	107
Abbildung 39 Modellbereich Assessment.....	109
Abbildung 40 Modellbereich Score.....	111
Abbildung 41 Modellbereich Statistik	113

Abbildung 42 PERSONAL-Modell.....	115
Abbildung 43 Grundstruktur der CSP-Assessment-Modellierung.....	118
Abbildung 44 Zusammenhang FACT - SNIPPET - ITEM (Ausschnitt).....	134
Abbildung 45 Muster Competence Niveau Map (CNM)	135
Abbildung 46 CNM Mathematik mittlerer Abschluss.....	135
Abbildung 47 CNMs zur Visualisierung des Zusammenhangs zwischen Kompetenz und Fakten.....	137
Abbildung 48 COMPETENCE NIVEAU MAP mit Operatoren	138
Abbildung 49 Beispiel für ein ASSESSMENT ITEM SET	140
Abbildung 50 ASSESSMENT ITEM SET EL01, mit EL01.1 bis EL01.4 als Items	142
Abbildung 51 Scoring mit PRO_PERS_PERSONAL_ITEM	149
Abbildung 52 Bewertung in PERSONAL ITEM SCORE	150
Abbildung 53 Abfrage zur Erzeugung der Gesamtscores PRO_PERS_PERSON_SCORE	151
Abbildung 54 Beispielscores in PERSONAL MANIFEST SCORE	152
Abbildung 55 Beispieldaten aus AGGREGATE	152
Abbildung 56 Aggregation mit PRO_PERS_AGGR_SKILL.....	153
Abbildung 57 Dynamische Ermittlung des Scores Variablen­gruppe	153
Abbildung 58 Ergebnis einer manifesten Analyse pro Person und Skill	154
Abbildung 59 Beispieldaten aus PERSONAL AGGREGATED SCORE.....	154
Abbildung 60 Analyse der Scores pro Skill auf materialisierten Scores.....	155
Abbildung 61 Ergebnis derselben Analyse auf Basis materialisierter Daten.....	155
Abbildung 62 Analyse pro Person und Fakt	155
Abbildung 63 Ergebnis der Analyse pro Fakt	156
Abbildung 64 Ziel des Schritts 10 im Informationsmodell.....	156
Abbildung 65 Zusammenhang zwischen Specific Competence Niveau und Feedback	160
Abbildung 66 Feedback-Analyse auf Faktenebene	160
Abbildung 67 Detaillierte Feedback-Analyse für einen RESPONSE TYPE	161
Abbildung 68 Struktur für die Ermittlung des Feedbacks für einen Assessment-Teilnehmer	162
Abbildung 69 Fehleranalyse für ein einzelnes Item	163
Abbildung 70 Übersicht über die aufgetretenen Fehlerursachen (Feedback)	163
Abbildung 71 Ergebnis der Bestimmung des Kompetenzniveaus.....	175
Abbildung 72 Scores (Anteile) der Assessment-Teilnehmer für Kompetenz C110	176
Abbildung 73 Scores pro Fact für zwei Assessment-Teilnehmer im Vergleich zum Durchschnitt.....	177
Abbildung 74 Ausschnitt der aus einer Kompetenz abgeleiteten Items	178
Abbildung 75 Lösungen pro Item für Kompetenzniveaus einer Kompetenz	178
Abbildung 76 Ausschnitt Zuordnung Niveau und Feedback	179
Abbildung 77 Bereich Qualification	188
Abbildung 78 Struktur eines CIP mit STEPs und STATES.....	191
Abbildung 79 Bereich CIP und GOALS	192
Abbildung 80 Graphischer Test	198
Abbildung 81 ICC-Plot für die verbliebenen Items	199
Abbildung 82 Graphischer Modelltest.....	203
Abbildung 83 ICC-Kurven der zweiten Analyse.....	204
Abbildung 84 Item EX15	230
Abbildung 85 Item W103	230
Abbildung 86 Item EX16.1	231
Abbildung 87 Ergebnisse für Item EX16.1 mit richtiger Antwort auf Item W103.1	231

Abbildung 88 Item W104.....	232
Abbildung 89 Item W107.....	233
Abbildung 90 Ausgewählte Teilnehmerergebnisse für EL01.....	236
Abbildung 91 ITEM EL01.....	237
Abbildung 92 Ausgewählte Teilnehmerergebnisse für EL01.....	237
Abbildung 93 Ausschnitt der Ermittlung der Fehlerursachen und Maßnahmen für Fakt 1107	238
Abbildung 94 Ausschnitt der Ermittlung der Fehlerursachen und Maßnahmen für Fakt1107 bestimmen	238
Abbildung 95 Beispiel-Item EX16.1 mit offener Eingabe einer Funktion	238
Abbildung 96 Analyse der Antworten	240
Abbildung 97 Häufigkeit aufgetretener Fehlerursachen (Ausschnitt).....	241

D. Assessment

(Aus technischen Gründen nicht in besserer Qualität verfügbar)

Fragebogen Seite 1 von 17



Variablenansicht

Die Variablenansicht zeigt alle Seiten des Fragebogens sowie die zugeordneten Variablen und Antwortcodes. Bitte beachten Sie, dass Filter und Platzhalter nicht korrekt wiedergegeben werden.

Für eine Übersicht aller Variablen im Befragungsprojekt verwenden Sie bitte die **Variablen-Übersicht**. Diese finden Sie in der Projektverwaltung im Menü auf der linken Seite.

Korrekturfahne Druckansicht

Seite 01
Start

1. Bitte geben Sie Ihren Nachnamen an: [PS01]

PS01_01 [01]
Offene Texteingabe

2. Bitte geben Sie Ihren Vornamen an [PS02]

PS02_01 [01]
Offene Texteingabe

3. Bitte wählen Sie Ihre Klasse aus [PS03]

BG12A

BG12B

BG12C

BG12D

PS03 Klasse
1 = BG12A
2 = BG12B
3 = BG12C
4 = BG12D
-9 = nicht beantwortet

Seite 02
W1

4. Ein Unternehmen, das im aktuellen Jahr 2010 (x=0) 150.000 Euro Umsatz macht, möchte in den nächsten Jahren seinen Umsatz um jährlich 6% steigern. [W101]

Entscheiden Sie für jede der folgenden Funktionsgleichungen, ob sie zur Problembeschreibung passt.

	ja	nein
$f(x) = 150000 * 0,06^x$	<input type="radio"/>	<input type="radio"/>
$f(x) = 1,06 * 150000^x$	<input type="radio"/>	<input type="radio"/>
$f(x) = 150000 * 1,06^x$	<input type="radio"/>	<input type="radio"/>

$$f(x) = 1,06x + 150000$$

W101_01 $f(x) = 150000 * 0,06^x$
W101_02 $f(x) = 1,06 * 150000^x$
W101_03 $f(x) = 150000 * 1,06^x$
W101_04 $f(x) = 1,06x + 150000$
 1 = nein
 2 = ja
 -9 = nicht beantwortet

5. Eine Algenart vermehrt sich durch Teilung ungeschlechtlich. Aus einer Alge entstehen so in einer Woche 4 Algen. In einem Teich werden zum Zeitpunkt $x=0$ genau 5 Algen gezählt. [EX10]

Entscheiden Sie für jede der folgenden Funktionsgleichungen, ob sie zu der gegebenen Fragestellung passt.

$$f(x) = 5 * 4^x$$

	ja	nein
$f(x) = 5 * 4^x$	<input type="radio"/>	<input type="radio"/>
$f(x) = 4 * 5^x$	<input type="radio"/>	<input type="radio"/>
$f(x) = 5x + 4$	<input type="radio"/>	<input type="radio"/>
$f(x) = 4x + 5$	<input type="radio"/>	<input type="radio"/>

$$f(x) = 4 * 5^x$$

$$f(x) = 5x + 4$$

$$f(x) = 4x + 5$$

EX10_01 $f(x) = 5 * 4^x$
EX10_02 $f(x) = 4 * 5^x$
EX10_03 $f(x) = 5x + 4$
EX10_04 $f(x) = 4x + 5$
 1 = nein
 2 = ja
 -9 = nicht beantwortet

6. Der Gewinn eines Unternehmens kann durch die Funktion $f(t) = 50000 + 2000t$ beschrieben werden. Dabei wird t in Jahren gemessen und es gilt für das Jahr 2010 $t=0$. [W102]

Der Gewinn für $t = 4$ ist 58000,-

	ja	nein
Der Gewinn für $t = 4$ ist 58000,-	<input type="radio"/>	<input type="radio"/>
Der Gewinn im Jahr 2013 ist 58000,-	<input type="radio"/>	<input type="radio"/>
Ein Umsatz von 64000,- wird im Jahr 2017 erreicht.	<input type="radio"/>	<input type="radio"/>
Es handelt sich um exponentielles Wachstum.	<input type="radio"/>	<input type="radio"/>

Der Gewinn im Jahr 2013 ist 58000,-

Ein Umsatz von 64000,- wird im Jahr 2017 erreicht.

Es handelt sich um exponentielles Wachstum.

W102_01 Der Gewinn für $t = 4$ ist 58000,-
W102_02 Der Gewinn im Jahr 2013 ist 58000,-
W102_03 Ein Umsatz von 64000,- wird im Jahr 2017 erreicht.
W102_04 Es handelt sich um exponentielles Wachstum.
 1 = nein
 2 = ja
 -9 = nicht beantwortet

7. Die Gewapa GmbH beschreibt ihren Umsatz mit der Gleichung $U(x) = 300000 * 1,08^x$. [W103]

Der Umsatz im Jahr 0 liegt bei 324000,-

	ja	nein
Der Umsatz im Jahr 0 liegt bei 324000,-	<input type="radio"/>	<input type="radio"/>
Das jährliche Wachstum liegt bei 8%.	<input type="radio"/>	<input type="radio"/>
Es kommt jährlich absolut gesehen immer derselbe Umsatz hinzu.	<input type="radio"/>	<input type="radio"/>
Es handelt sich um lineares Wachstum.	<input type="radio"/>	<input type="radio"/>

Das jährliche Wachstum liegt bei 8%.

Es kommt jährlich absolut gesehen immer derselbe Umsatz hinzu.

Es handelt sich um lineares Wachstum.

W103_01 Der Umsatz im Jahr 0 liegt bei 324000,-
W103_02 Das jährliche Wachstum liegt bei 8%.
W103_03 Es kommt jährlich absolut gesehen immer derselbe Umsatz hinzu.
W103_04 Es handelt sich um lineares Wachstum.
 1 = nein
 2 = ja
 -9 = nicht beantwortet

Seite 03
 w1

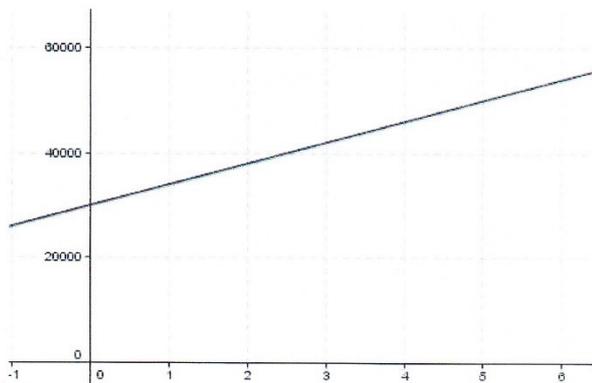
8. Die Gewapa GmbH beschreibt ihren Umsatz mit der Gleichung $U(x) = 300000 * 1,08^x$. x ist in Jahren mit $x=0$ für das Jahr 2010 angegeben.
 [W104]

Der Umsatz nach 4 Jahren beträgt (gerundet in ganzen Euro) Euro

Bestimmen Sie die Anzahl Jahre (ganzzahlig) nach denen sich der Umsatz um mindestens 50% erhöht hat. Jahre

Bestimmen Sie den Umsatz für 2009, wenn er sich von 2009 auf 2010 auch um 8% erhöht hätte (gerundet in ganzen Euro). Euro

W104_01 Der Umsatz nach 4 Jahren beträgt (gerundet in ganzen Euro). ... Euro
W104_02 Bestimmen Sie die Anzahl Jahre (ganzzahlig) nach denen sich der Umsatz um mindestens 50% erhöht hat. ... Jahre
W104_03 Bestimmen Sie den Umsatz für 2009, wenn er sich von 2009 auf 2010 auch um 8% erhöht hätte (gerundet in ganzen Euro). ... Euro
 Offene Texteingabe



9. [W105]

Hinweis: Gehen Sie davon aus, dass erkennbare Schnittpunkte jeweils in der Mitte der Beschriftungen liegen.

Bestimmen Sie den Anfangsbestand (in ganzen Euro) Euro

Bestimmen Sie das jährliche Wachstum (in ganzen Euro) Euro

W105_01 Bestimmen Sie den Anfangsbestand (in ganzen Euro) ... Euro
W105_02 Bestimmen Sie das jährliche Wachstum (in ganzen Euro) ... Euro
 Offene Texteingabe

10. Röntgenstrahlen kann man mit Bleiplatten abschirmen. Sie werden in Coulomb als Einheit gemessen (wie Strom in Ampere oder Wasser in Liter gemessen wird). Die Funktionsgleichung für die Strahlung lautet $f(x) = 0,95^x$. [W110]

	ja	nein
Die Strahlung nimmt je mm Bleidicke um 95% zu	<input type="radio"/>	<input type="radio"/>
Die Strahlung nimmt um 95 Coulomb je mm ab.	<input type="radio"/>	<input type="radio"/>
Die Strahlung nimmt um 5% je mm ab.	<input type="radio"/>	<input type="radio"/>
Die Strahlung nimmt um 0,5 Coulomb je mm ab.	<input type="radio"/>	<input type="radio"/>

W110_01 Die Strahlung nimmt je mm Bleidicke um 95% zu
W110_02 Die Strahlung nimmt um 95 Coulomb je mm ab.
W110_03 Die Strahlung nimmt um 5% je mm ab.
W110_04 Die Strahlung nimmt um 0,5 Coulomb je mm ab.
 1 = nein
 2 = ja
 -9 = nicht beantwortet

11. Im Jahr 2010 ($t=0$) betrug die Bevölkerung in Deutschland 83 Millionen Menschen. Im Jahr 2013 werden 76,221 Millionen Menschen gezählt. [W106]

Ergänzen Sie die Funktionsgleichung (ganzzahlig und bei Bedarf mit zwei Nachkommastellen)

$f(t) = \text{[]} * \text{[]} ^t$

W106_01 $f(t) = \dots *$
W106_02 $\dots ^t$
 Offene Texteingabe

12. Die Messung der Vermehrungsrate von Forellen hat folgenden

funktionalen Zusammenhang ergeben $f(t) = 110 \cdot 1,10^t$;
 [W107]

Bei der Übertragung der Formel auf einen Teich mit 15% Wachstumsrate und 200 Fischen Anfangsbestand ergäbe sich dann (ganzzahlig und bei Bedarf mit zwei Nachkommastellen):

$f(t) = \text{[]} * \text{[]} ^t$

W107_01 $f(t) = \dots *$
W107_02 $\dots ^t$
 Offene Texteingabe

13. Die Messung der Vermehrungsrate von Forellen hat folgenden

$$f(t) = 110 \cdot 1,10^t$$

funktionalen Zusammenhang ergeben
[W108]

Die prozentuale Vermehrung pro Periode beträgt (in ganzen Prozent)

Die prozentuale Zunahme beträgt (in ganzen %) %

W108_01 Die prozentuale Zunahme beträgt (in ganzen %) ... %
Offene Texteingabe

14. Ein Taubenzüchter hat einen Taubenschlag, der ganzjährig bewohnbar ist. Dir Fütterung der Tauben erfolgt mit Mais und Weizenkörnern im Verhältnis 50:50. Seine Taubenzucht bestand vor drei Jahren aus 40 Tauben. Dieses Jahr hat er 135 Tauben gezählt, die jährlich etwa 200kg Futter fressen. [EX16]

Geben Sie eine Funktionsgleichung an, die die Zunahme an Tauben richtig beschreibt.

f(x) =

EX16_01 f(x) = ...
Offene Texteingabe

15. Entscheiden Sie jeweils, ob es sich um exponentielles Wachstum (einschließlich exponentiellem Zerfall) handelt: [W109]

	ja	nein
In einem Wald wächst der Holzbestand jeweils um 2000 Raummeter pro Jahr an	<input type="radio"/>	<input type="radio"/>
In einem Unternehmen wird das Eigenkapital pro Jahr um 4% verzinst und die Zinsen bleiben im Unternehmen und erhöhen das Eigenkapital.	<input type="radio"/>	<input type="radio"/>
Von einem radioaktiven Stoff zerfällt jeweils die Hälfte der Menge pro Jahr.	<input type="radio"/>	<input type="radio"/>
Die gefahren km eines Pkw nehmen pro Monat um 2000 km zu, weil das Auto jeden Monat 2000km fährt.	<input type="radio"/>	<input type="radio"/>
Ein Jungvogel legt wöchentlich 30% seines Eigengewichts zu.	<input type="radio"/>	<input type="radio"/>
Von einem Kapital werden jährlich 5% der Ausgangssumme entnommen.	<input type="radio"/>	<input type="radio"/>

W109_01 In einem Wald wächst der Holzbestand jeweils um 2000 Raummeter pro Jahr an
W109_02 In einem Unternehmen wird das Eigenkapital pro Jahr um 4% verzinst und die Zinsen bleiben im Unternehmen und erhöhen das Eigenkapital.
W109_03 Von einem radioaktiven Stoff zerfällt jeweils die Hälfte der Menge pro Jahr.
W109_04 Die gefahren km eines Pkw nehmen pro Monat um 2000 km zu, weil das Auto jeden Monat 2000km fährt.
W109_05 Ein Jungvogel legt wöchentlich 30% seines Eigengewichts zu.
W109_06 Von einem Kapital werden jährlich 5% der Ausgangssumme entnommen.
 1 = nein
 2 = ja
 -9 = nicht beantwortet

16. Von einer Insektenart sind am Anfang 5 Exemplare vorhanden. Nach 3 Zeiteinheiten sind es 320 Exemplare. Das Wachstum soll durch eine Exponentialfunktion $f(x)=a*b^x$ beschrieben werden. Bestimmen Sie die Funktionsgleichung. [EX15]

$$f(x) = \square * \square ^x$$

EX15_01 $f(x)= \dots$
 EX15_02 $* \dots ^x$
 Offene Texteingabe

17. Entscheiden Sie welche der folgenden Aussagen richtig sind: [EX11]

	ja	nein
Beim linearen Wachstum kommt in jeder Zeiteinheit gleich viel hinzu.	<input type="radio"/>	<input type="radio"/>
Beim linearen Wachstum gilt eine Funktionsgleichung der Form $f(x)=ax+b$.	<input type="radio"/>	<input type="radio"/>
Beim linearen Wachstum gilt eine Funktionsgleichung der Form $f(x)=a*b^x$	<input type="radio"/>	<input type="radio"/>
Beim linearen Wachstum gibt in $f(x)=ax+b$ das b an, wie viel am Anfang vorhanden ist.	<input type="radio"/>	<input type="radio"/>
Beim linearen Wachstum gibt in $f(x) = a*b^x$ das a an, wie viel am Anfang vorhanden ist.	<input type="radio"/>	<input type="radio"/>
Beim linearen Wachstum kommt prozentual je Zeiteinheit gleich viel dazu.	<input type="radio"/>	<input type="radio"/>
Beim linearen Wachstum gilt, je mehr da ist, umso mehr kommt je Zeiteinheit hinzu.	<input type="radio"/>	<input type="radio"/>

EX11_01 Beim linearen Wachstum kommt in jeder Zeiteinheit gleich viel hinzu.
 EX11_02 Beim linearen Wachstum gilt eine Funktionsgleichung der Form $f(x)=ax+b$.
 EX11_03 Beim linearen Wachstum gilt eine Funktionsgleichung der Form $f(x)=a*b^x$
 EX11_04 Beim linearen Wachstum gibt in $f(x)=ax+b$ das b an, wie viel am Anfang vorhanden ist.
 EX11_05 Beim linearen Wachstum gibt in $f(x) = a*b^x$ das a an, wie viel am Anfang vorhanden ist.
 EX11_06 Beim linearen Wachstum kommt prozentual je Zeiteinheit gleich viel dazu.
 EX11_07 Beim linearen Wachstum gilt, je mehr da ist, umso mehr kommt je Zeiteinheit hinzu.
 1 = nein
 2 = ja
 -9 = nicht beantwortet

18. Entscheiden Sie welche der folgenden Aussagen richtig sind: [EX12]

	ja	nein
Beim exponentiellen Wachstum kommt in jeder Zeiteinheit gleich viel hinzu.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gilt eine Funktionsgleichung der Form $f(x)=ax+b$.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gilt eine Funktionsgleichung der Form $f(x)=a*b^x$	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gibt in $f(x)=ax+b$ das b an, wie viel am Anfang vorhanden ist.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gibt in $f(x) = a*b^x$ das a an, wie viel am Anfang vorhanden ist.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum kommt prozentual je Zeiteinheit gleich viel dazu.	<input type="radio"/>	<input type="radio"/>
Beim exponentiellen Wachstum gilt, je mehr da ist, umso mehr kommt je Zeiteinheit dazu.	<input type="radio"/>	<input type="radio"/>

EX12_01 Beim exponentiellen Wachstum kommt in jeder Zeiteinheit gleich viel hinzu.
EX12_02 Beim exponentiellen Wachstum gilt eine Funktionsgleichung der Form $f(x)=ax+b$.
EX12_03 Beim exponentiellen Wachstum gilt eine Funktionsgleichung der Form $f(x)=a*b^x$
EX12_04 Beim exponentiellen Wachstum gibt in $f(x)=ax+b$ das b an, wie viel am Anfang vorhanden ist.
EX12_05 Beim exponentiellen Wachstum gibt in $f(x) = a*b^x$ das a an, wie viel am Anfang vorhanden ist.
EX12_06 Beim exponentiellen Wachstum kommt prozentual je Zeiteinheit gleich viel dazu.
EX12_07 Beim exponentiellen Wachstum gilt, je mehr da ist, umso mehr kommt je Zeiteinheit dazu.
 1 = nein
 2 = ja
 -9 = nicht beantwortet

Seite 05
E

19. $5^x - 625 = 0$ [EX01]

Es handelt sich um eine Exponentialgleichung.
 Die Basis ist x.

ja nein

EX01_01 Es handelt sich um eine Exponentialgleichung.
EX01_02 Die Basis ist x.
 1 = nein
 2 = ja
 -9 = nicht beantwortet

20. Berechnen Sie jeweils (ganzzahlig, wenn das Ergebnis ganzzahlig ist, sonst auf zwei Stellen gerundet). [EX13]

$\ln(10) =$

$\log_4 20 =$

$\log_{12} 1 =$

$\log_a 1 =$

$\log_e e =$

EX13_01 [01]
EX13_02 [02]
EX13_03 [03]
EX13_04 [04]
EX13_05 [05]
 Offene Texteingabe

21. Bestimmen Sie x. [EX02]

$5^x - 625 = 0$

$4096 = 2^{x+10}$

$e^{2x} = e^{-x+6}$

$2^{2x} \cdot 2^{0,5x} = 32$

EX02_01 [01]
 EX02_02 [02]
 EX02_03 [03]
 EX02_04 [04]
 Offene Texteingabe

22. Sie haben eine Kapital von 20000 Euro. Geben Sie jeweils den Faktor (die Zahl) an mit dem das Kapital multipliziert werden muss, wenn der angegebene Zinssatz gilt, um das Kapital im nächsten Jahr auszurechnen. [ELO1]

bei 4% bei 10% bei -4% bei 0,5%

ELO1_01 bei 4% ...
 ELO1_02 bei 10% ...
 ELO1_03 bei -4% ...
 ELO1_04 bei 0,5% ...
 Offene Texteingabe

23. Gegeben ist die Funktion $f(x) = a \cdot b^x$ [EX07]

Beantworten Sie die folgenden Fragen

	ja	nein
Je größer a ist, desto mehr ist die Funktion nach rechts verschoben.	<input type="radio"/>	<input type="radio"/>
Je größer a ist, desto steiler verläuft der Graph.	<input type="radio"/>	<input type="radio"/>
Je größer a ist, desto mehr ist die Funktion gesteckt.	<input type="radio"/>	<input type="radio"/>
x ist der Stauchungsfaktor	<input type="radio"/>	<input type="radio"/>
a gibt den Funktionswert bei x=0 an	<input type="radio"/>	<input type="radio"/>
b ist die Basis der Exponentialfunktion	<input type="radio"/>	<input type="radio"/>
Je größer b ist, desto stärker steigt die Funktion an.	<input type="radio"/>	<input type="radio"/>
Wenn b kleiner 1 ist fällt die Funktion	<input type="radio"/>	<input type="radio"/>

EX07_07 Je größer a ist, desto mehr ist die Funktion nach rechts verschoben.
 EX07_08 Je größer a ist, desto steiler verläuft der Graph.
 EX07_06 Je größer a ist, desto mehr ist die Funktion gesteckt.
 EX07_02 x ist der Stauchungsfaktor
 EX07_11 a gibt den Funktionswert bei x=0 an
 EX07_01 b ist die Basis der Exponentialfunktion

EX07_10 Je größer b ist, desto stärker steigt die Funktion an.

EX07_12 Wenn b kleiner 1 ist fällt die Funktion

- 1 = nein
2 = ja
-9 = nicht beantwortet

24. Gegeben ist die Funktion $f(x) = a \cdot b^x + c$ [EX03]

Beantworten Sie die folgenden Fragen

	ja	nein
Je größer c ist, desto mehr wird der Graph nach oben verschoben	<input type="radio"/>	<input type="radio"/>
Je größer c ist desto mehr ist die Funktion gestreckt	<input type="radio"/>	<input type="radio"/>
Bei negativem c wird die Funktion an der y-Achse gespiegelt.	<input type="radio"/>	<input type="radio"/>
a+c gibt den Funktionswert bei x=0 an	<input type="radio"/>	<input type="radio"/>
Wenn a größer als c ist steigt die Funktion immer monoton an.	<input type="radio"/>	<input type="radio"/>

EX03_03 Je größer c ist, desto mehr wird der Graph nach oben verschoben

EX03_04 Je größer c ist desto mehr ist die Funktion gestreckt

EX03_05 Bei negativem c wird die Funktion an der y-Achse gespiegelt.

EX03_11 a+c gibt den Funktionswert bei x=0 an

EX03_12 Wenn a größer als c ist steigt die Funktion immer monoton an.

- 1 = nein
2 = ja
-9 = nicht beantwortet

25. Gegeben ist die Gleichung: $100 = 80 \cdot b^3$ [EX08]

Bestimmen Sie b auf zwei Nachkommstellen gerundet.

b=

EX08_01 b= ...
Offene Texteingabe

26. Gegeben ist die Gleichung: $20 = 5 \cdot b^4$ [EX09]

Bestimmen Sie b auf zwei Nachkommstellen gerundet.

b=

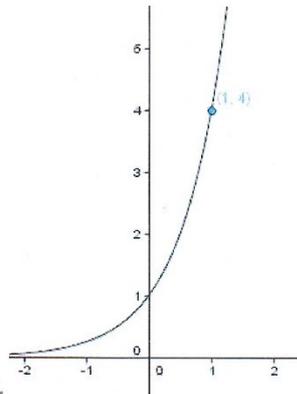
EX09_01 b= ...
Offene Texteingabe

Seite 06
EX

27. Eine Exponentialfunktion $f(x) = a \cdot b^x$ geht durch die Punkte (0|10) und (2|90). Bestimmen Sie a und b. [EX14]

a= b=

EX14_01 a= ...
EX14_02 b= ...
Offene Texteingabe



28.

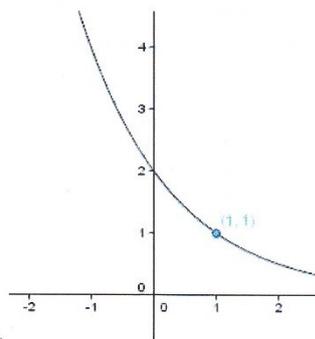
[EX04]

Geben Sie die Parameter der dargestellten Exponentialfunktion $f(x) = a * b^x$ an.

a =

b =

EX04_01 a =
EX04_02 b =
Offene Texteingabe



29.

[EX05]

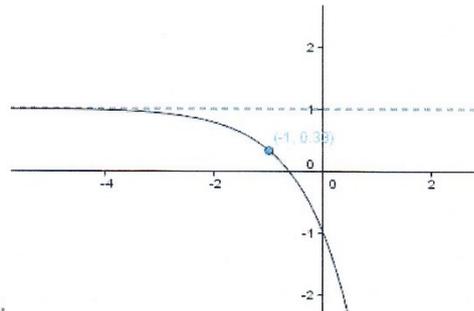
Geben Sie die Parameter der dargestellten Exponentialfunktion $f(x) = a * b^x$ an.

a =

b = (eine Nachkommastelle möglich)

EX05_01 a =
EX05_02 b = ... (eine Nachkommastelle möglich)

Offene Texteingabe



30.

[EX06]

Geben Sie die Parameter der dargestellten Exponentialfunktion $f(x) = a * b^x + c$ an.
Alle Angaben bei Bedarf ganzzahlig gerundet.

a =
b =
c =

EX06_01 a =
EX06_02 b =
EX06_03 c =
Offene Texteingabe

Seite 07
BO

31. Ein Hersteller von Solarautos hat seine Kostenfunktion ermittelt zu $K(x) = 0,25x^3 - 2x^2 + 6x + 12,5$ (alle Angaben in Tausend Euro). [BO01]

Tragen Sie die folgenden Werte jeweils ganzzahlig gerundet ein.

Die Menge des Betriebsoptimums (betriebsoptimale Menge) ist Stück (gerundet in ganzen Stück).

Die langfristige Preisuntergrenze ist Euro (gerundet in Tausend Euro, zwei Nachkommastellen).

Die Menge des Betriebsminimums (betriebsminimale Menge) ist Stück (gerundet in ganzen Stück).

Die kurzfristige Preisuntergrenze ist Euro (gerundet in ganzen Tausend Euro).

BO01_01 Die Menge des Betriebsoptimums (betriebsoptimale Menge) ist ... Stück (gerundet in ganzen Stück).
BO01_02 Die langfristige Preisuntergrenze ist ... Euro (gerundet in Tausend Euro, zwei Nachkommastellen).
BO01_03 Die Menge des Betriebsminimums (betriebsminimale Menge) ist ... Stück (gerundet in ganzen Stück).
BO01_04 Die kurzfristige Preisuntergrenze ist ... Euro (gerundet in ganzen Tausend Euro).
Offene Texteingabe

32. In einem Betrieb liegt die langfristige Preisuntergrenze bei 500,- Euro bei einer Menge von 2500 Stück. Die kurzfristige Preisuntergrenze liegt bei 300,- Euro bei 2000 Stück. [B002]

Entscheiden Sie für die folgenden Antworten jeweils welche richtig oder falsch ist.

	Ja	Nein
Bei Kosten von 300,- Euro pro Stück wird gerade im Betriebsminimum produziert.	<input type="radio"/>	<input type="radio"/>
Bei einem Preis von 500,- Euro pro Stück sind im Betriebsoptimum sowohl fixe als auch variable Kosten gedeckt.	<input type="radio"/>	<input type="radio"/>
Ein Preis von 400,- Euro ist dauerhaft kostendeckend.	<input type="radio"/>	<input type="radio"/>
Bei der Markteinführung kann das Produkt für 350,- Euro in einem begrenzten Aktionszeitraum angeboten werden.	<input type="radio"/>	<input type="radio"/>
Aus Stückkostensicht ist es am günstigsten 2500 Stück zu produzieren.	<input type="radio"/>	<input type="radio"/>
Die Geschäftsführung sollte einen Preis knapp unter 400,- Euro, beispielsweise 398,- Euro als Standardverkaufspreis festlegen.	<input type="radio"/>	<input type="radio"/>
Wenn die fixen Kosten gesenkt werden, sinkt auch die langfristige Preisuntergrenze.	<input type="radio"/>	<input type="radio"/>
Wenn die fixen Kosten erhöht werden, steigt auch die kurzfristige Preisuntergrenze.	<input type="radio"/>	<input type="radio"/>
Es gilt in jedem Fall: Je mehr produziert wird, umso geringer sind die Kosten pro Stück.	<input type="radio"/>	<input type="radio"/>
Der größtmögliche Gewinn kann in jedem Fall mit 2500 Stück erreicht werden.	<input type="radio"/>	<input type="radio"/>
Es gilt in jedem Fall: Die langfristige Preisuntergrenze ist größer als die kurzfristige Preisuntergrenze.	<input type="radio"/>	<input type="radio"/>
Es gilt in jedem Fall: Je mehr produziert wird, desto geringer sind die Fixkosten pro Stück.	<input type="radio"/>	<input type="radio"/>

BO02_06 Bei Kosten von 300,- Euro pro Stück wird gerade im Betriebsminimum produziert.
BO02_07 Bei einem Preis von 500,- Euro pro Stück sind im Betriebsoptimum sowohl fixe als auch variable Kosten gedeckt.
BO02_01 Ein Preis von 400,- Euro ist dauerhaft kostendeckend.
BO02_02 Bei der Markteinführung kann das Produkt für 350,- Euro in einem begrenzten Aktionszeitraum angeboten werden.
BO02_03 Aus Stückkostensicht ist es am günstigsten 2500 Stück zu produzieren.
BO02_05 Die Geschäftsführung sollte einen Preis knapp unter 400,- Euro, beispielsweise 398,- Euro als Standardverkaufspreis festlegen.
BO02_08 Wenn die fixen Kosten gesenkt werden, sinkt auch die langfristige Preisuntergrenze.
BO02_09 Wenn die fixen Kosten erhöht werden, steigt auch die kurzfristige Preisuntergrenze.
BO02_10 Es gilt in jedem Fall: Je mehr produziert wird, umso geringer sind die Kosten pro Stück.
BO02_04 Der größtmögliche Gewinn kann in jedem Fall mit 2500 Stück erreicht werden.
BO02_12 Es gilt in jedem Fall: Die langfristige Preisuntergrenze ist größer als die kurzfristige Preisuntergrenze.
BO02_11 Es gilt in jedem Fall: Je mehr produziert wird, desto geringer sind die Fixkosten pro Stück.
1 = Ja 2 = Nein -9 = nicht beantwortet

33. Das Betriebsoptimum ist [B004]

- Das Minimum der Gesamtkostenfunktion
 Das Minimum der Grenzkostenfunktion

- Das Minimum der Funktion der Stückkosten
- Das Minimum der Funktion der variablen Stückkosten
- Das Maximum der Gewinnfunktion
- Das Maximum der Erlösfunktion
- Die maximal zu produzierende Menge
- Der Schnittpunkt von Erlös- und Kostenfunktion

BO04 BO04

- 1 = Das Minimum der Gesamtkostenfunktion
- 2 = Das Minimum der Grenzkostenfunktion
- 3 = Das Minimum der Funktion der Stückkosten
- 4 = Das Minimum der Funktion der variablen Stückkosten
- 5 = Das Maximum der Gewinnfunktion
- 6 = Das Maximum der Erlösfunktion
- 7 = Die maximal zu produzierende Menge
- 8 = Der Schnittpunkt von Erlös- und Kostenfunktion
- 9 = nicht beantwortet

34. Das Betriebsminimum ist [BO05]

- Das Minimum der Gesamtkostenfunktion
- Das Minimum der Grenzkostenfunktion
- Das Minimum der Funktion der Stückkosten
- Das Minimum der Funktion der variablen Stückkosten
- Das Maximum der Gewinnfunktion
- Das Maximum der Erlösfunktion
- Die maximal zu produzierende Menge
- Der Schnittpunkt von Erlös- und Kostenfunktion

BO05 BO05

- 1 = Das Minimum der Gesamtkostenfunktion
- 2 = Das Minimum der Grenzkostenfunktion
- 3 = Das Minimum der Funktion der Stückkosten
- 4 = Das Minimum der Funktion der variablen Stückkosten
- 5 = Das Maximum der Gewinnfunktion
- 6 = Das Maximum der Erlösfunktion
- 7 = Die maximal zu produzierende Menge
- 8 = Der Schnittpunkt von Erlös- und Kostenfunktion
- 9 = nicht beantwortet

35. Gegeben ist Kostenfunktion $K(x) = x^3 - 6x^2 + 15x + 32$. [BO03]

Füllen Sie die folgenden Felder aus. Geben Sie jeweils das Vorzeichen mit an. Felder, die sie nicht benötigen füllen Sie mit einer 0.

Die Funktion der Stückkosten $k(x)$ lautet x^2 x $/x$.

Die Funktion der variablen Stückkosten lautet $kv(x)$ lautet x^2 x $/x$

Die Grenzkostenfunktion $K'(x)$ lautet x^2 x $/x$.

BO03_01 Die Funktion der Stückkosten $k(x)$ lautet ... x^2

BO03_02 ... x

BO03_03 [03]

BO03_04 ... $/x$.

BO03_05 Die Funktion der variablen Stückkosten lautet $kv(x)$ lautet ... x^2

BO03_06 ... x

BO03_07 [07]

BO03_08 ... /x
BO03_09 Die Grenzkostenfunktion $K'(x)$ lautet ... x^2
BO03_10 ... x
BO03_11 [11]
BO03_12 ... /x.
 Offene Texteingabe

Seite 08
PR

36. Sie sollen einen Hochpunkt/Tiefpunkt einer Funktion $f(x)$ bestimmen.
Gegeben sind folgende Schritte

- Zweite Ableitung bilden.
- Zweite Ableitung auf Null stellen und x -Werte ermitteln.
- Funktion auf Null stellen und x -Werte ermitteln.
- Erste Ableitung bilden.
- Ermittelte x -Werte in zweite Ableitung einsetzen.
- y - Wert durch Einsetzen der x -Werte in Ausgangsfunktion ermitteln.
- Erste Ableitung auf Null stellen und x -Werte ermitteln.
- Mit dritter Ableitung entscheiden <0 Tiefpunkt, >0 Hochpunkt.
- mit erster Ableitung entscheiden <0 Tiefpunkt, >0 Hochpunkt.
- mit zweiter Ableitung entscheiden <0 Hochpunkt, >0 Tiefpunkt.
- Dritte Ableitung bilden.
- y Wert durch Einsetzen in zweite Ableitung ermitteln.
- Ermittelte x -Werte in erste Ableitung einsetzen.
- Ermittelte x -Werte in dritte Ableitung einsetzen. [PR01]

Geben Sie die richtige Reihenfolge Ihrer Schritte an, beispielsweise c, a, d, ...

Es werden nicht alle Schritte benötigt.

Es sind mehrere Lösungen möglich, geben Sie nur eine Lösung an.

Reihenfolge:

PR01_01 Reihenfolge ...
 Offene Texteingabe

37. Sie sollen einen Wendepunkt einer Funktion $f(x)$ bestimmen.
Gegeben sind folgende Schritte

- Zweite Ableitung bilden.
- Zweite Ableitung auf Null stellen und x -Werte ermitteln.
- Funktion auf Null stellen und x -Werte ermitteln.
- Erste Ableitung bilden.
- Ermittelte x -Werte in zweite Ableitung einsetzen.
- y - Wert durch Einsetzen der x -Werte in Ausgangsfunktion ermitteln.
- Erste Ableitung auf Null stellen und x -Werte ermitteln.
- Mit dritter Ableitung entscheiden <0 Wendepunkt, >0 Wendepunkt.
- mit erster Ableitung entscheiden <0 Wendepunkt, >0 Wendepunkt.
- mit zweiter Ableitung entscheiden <0 Wendepunkt, >0 Wendepunkt.
- Dritte Ableitung bilden.
- y Wert durch Einsetzen in zweite Ableitung ermitteln.
- Ermittelte x -Werte in erste Ableitung einsetzen.
- Ermittelte x -Werte in dritte Ableitung einsetzen. [PR03]

Geben Sie die richtige Reihenfolge Ihrer Schritte an, beispielsweise c, a, d, ...

Es werden nicht alle Schritte benötigt.

Es sind mehrere Lösungen möglich, geben Sie nur eine Lösung an.

Reihenfolge:

PR03_01 Reihenfolge ...
Offene Texteingabe

Seite 09
SY

38. Eine Funktion dritten Grades besitzt an der Stelle 6 eine Nullstelle und an der Stelle 3 einen Wendepunkt mit der Steigung -3. Bestimmen Sie die Bedingungen und die Matrix aus denen sich die Funktion bestimmen lässt.
[SY01]

1) =
2) =
3) =
4) =

SY01_01 1) ...
SY01_02 [02]
SY01_03 [03]
SY01_04 [04]
SY01_05 = ...
SY01_06 2) ...
SY01_07 [07]
SY01_08 [08]
SY01_09 [09]
SY01_10 = ...
SY01_11 3) ...
SY01_12 [12]
SY01_13 [13]
SY01_14 [14]
SY01_15 = ...
SY01_16 4) ...
SY01_17 [17]
SY01_18 [18]
SY01_19 [19]
SY01_20 = ...
Offene Texteingabe

39. Eine Funktion 3. Grades hat in (-2|0) einen Hochpunkt und in (-1|-2) einen Wendepunkt. Bestimmen Sie die Funktionsgleichung. [SY03]

$f(x) = \text{[]} (a) \text{[]} (b) \text{[]} (c) \text{[]} (d)$
(bitte Koeffizienten mit Vorzeichen eingeben)

SY03_01 $f(x) = \dots (a)$
SY03_02 ... (b)
SY03_03 ... (c)
SY03_04 ... (d) (bitte Koeffizienten mit Vorzeichen eingeben)
Offene Texteingabe

40. Ein Betrieb weist an seiner Kapazitätsgrenze von 7 ME Kosten von 122 GE aus. Bei einer Produktion von 4 ME betragen die Grenzkosten 6 GE pro ME.

Das Betriebsminimum liegt im Punkt (4,5|9,75). Bestimmen Sie die Matrix zur Ermittlung der Kostenfunktion. [SY02]

1)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	=	<input type="text"/>
2)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	=	<input type="text"/>
3)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	=	<input type="text"/>
4)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	=	<input type="text"/>

SY02_01 1) ...
 SY02_02 [02]
 SY02_03 [03]
 SY02_04 [04]
 SY02_05 = ...
 SY02_06 2) ...
 SY02_07 [07]
 SY02_08 [08]
 SY02_09 [09]
 SY02_10 = ...
 SY02_11 3) ...
 SY02_12 [12]
 SY02_13 [13]
 SY02_14 [14]
 SY02_15 = ...
 SY02_16 4) ...
 SY02_17 [17]
 SY02_18 [18]
 SY02_19 [19]
 SY02_20 = ...
 Offene Texteingabe

Seite 10
FB

41. Es geht jetzt um Ihren Eindruck von diesem Test. [PS06]

Bitte setzen Sie den Schieberegler jeweils an die für Sie richtige Stelle

ist zeitlich zu knapp bemessen	ist zeitlich zu lang bemessen
passt nicht zum Unterricht	passt sehr gut zum Unterricht
fragt mein Wissen nicht wirklich ab	fragt mein Wissen gut ab
hat für mich nichts mit Mathematik zu tun	bildet mein Verständnis von Mathematik exakt ab
ich erwarte ein ganz anderes Ergebnis als in einer Klausur	ich erwarte das gleiche Ergebnis wie in einer Klausur

PS06_02 ist zeitlich zu knapp bemessen/ist zeitlich zu lang bemessen
 1 = ist zeitlich zu knapp bemessen
 100 = ist zeitlich zu lang bemessen
 -9 = nicht beantwortet

PS06_03 passt nicht zum Unterricht/passt sehr gut zum Unterricht
 1 = passt nicht zum Unterricht
 100 = passt sehr gut zum Unterricht
 -9 = nicht beantwortet

PS06_01 fragt mein Wissen nicht wirklich ab/fragt mein Wissen gut ab
 1 = fragt mein Wissen nicht wirklich ab
 100 = fragt mein Wissen gut ab
 -9 = nicht beantwortet

PS06_04 hat für mich nichts mit Mathematik zu tun/bildet mein Verständnis von Mathematik exakt ab

1 = hat für mich nichts mit Mathematik zu tun
100 = bildet mein Verständnis von Mathematik exakt ab
-9 = nicht beantwortet

PS06_05 ich erwarte ein ganz anderes Ergebnis als in einer Klausur/ich erwarte das gleiche Ergebnis wie in einer Klausur

1 = ich erwarte ein ganz anderes Ergebnis als in einer Klausur
100 = ich erwarte das gleiche Ergebnis wie in einer Klausur
-9 = nicht beantwortet

42. Ich finde diese Form generell besser oder schlechter gegenüber einer schriftlichen Klausur. [PS04]



PS04 PS04

1 =
2 =
3 =
4 =
5 =
-9 = nicht beantwortet

43. Meine Abschlusssanmerkungen [PS05]

PS05_01 [01]

Offene Texteingabe

Letzte Seite

Danke für Ihre Teilnahme!

Wir möchten uns ganz herzlich für Ihre Mithilfe bedanken.

Fenster schließen

J.-H. Wieken

E. Literaturverzeichnis

- Aalst, W. M. P., van der. *Patterns and XPDL: A Critical Evaluation of the XML Process Definition Language, Technical Report BPM-03-09*. BMPCenter.org, 2003.
- Achtenhagen, F., und E. Winther. „Wirtschaftspädagogische Forschung zur beruflichen Kompetenzentwicklung.“ In *Kompetenzerfassung in pädagogischen Handlungsfeldern*, von N. Jude, J. Hartig und E. Klieme, 117-140. Bonn/Berlin: BMBF, 2008.
- Achtenhagen, F., und M. Baethge. „Kompetenzdiagnostik als Large-Scale-Assessment im Bereich der beruflichen Aus- und Weiterbildung.“ *Zeitschrift für Erziehungswissenschaft, Sonderheft 8*, 2007: 51-70.
- Agel, V. *Valenztheorie*. Tübingen: Narr, 2000.
- Allweyer, T. *BPMN 2.0 Business Process Model and Notation*. Norderstedt: BoD, 2009.
- Anderson, J. *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum, 1993.
- Anderson, L. W., und D. R. Krathwohl. *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman, 2001.
- Balzert, H. *Lehrbuch der Software-Technik, Band 2*. Heidelberg/Berlin: Spektrum Akademischer Verlag, 1998.
- Baumgartner, P. *Taxonomie von Unterrichtsmethoden*. Münster: Waxmann, 2014.
- Bayreuth, Universität. *Beispiel: Zielsetzung Bachelorstudiengang*. 11. 09 2014. Universität Bayreuth, http://www.uni-bayreuth.de/universitaet/qualitaetsmanagement/Bologna_FAQ/Lernziele/index.html (Zugriff am 19. 10 2015).
- Bayrhuber, M., T. Leuders, R. Bruder, und M. Wirtz. „Repräsentationswechsel beim Umgang mit Funktionen - identifikation von Kompetenzprofilen auf der Basis eines Kompetenzstrukturmodells.“ In *Kompetenzmodellierung*, von E. Klieme, T. Leutner und M. Kenk, 28-39. Weinheim und Basel: Beltz, 2010.
- Beyen, W. „Kompetenzorientierung im Unterricht - Anmerkungen zu einem denkwürdigen Entwurf.“ *Wirtschaft & Erziehung*, 08 2013: 279-288.
- Biggs, J. B. *Teaching for quality learning at university*. Berkshire: Society for Research in Higher Education and Open University Press, 2003.
- Biggs, J. B., und K. F. Collis. *Evaluating the quality of learning: The SOLO-Taxonomy*. New York: Academic Press, 1982.
- „Bildungsstandards der Kultusministerkonferenz.“ *Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland*. München: Wolters Kluwer, 2005.
- Blerkom, M. L. van. *Measurement and Statistics for Teachers*. New York: Routledge, 2009.
- Bloom, B. S. *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim: Beltz, 1976.

- . *Taxonomy of educational objectives: The classification of educational goals: Handbook I, The cognitive domain*. New York: David McKay Company, 1956.
- BMBF. *DQR - Deutscher Qualifikationsrahmen*. Bundesministerium für Bildung und Forschung, 2011.
- . „Zur Entwicklung nationaler Bildungsstandards.“ Bonn/Berlin, 2009.
- . *Zur Entwicklung nationaler Bildungsstandards*. Bonn/Berlin, 2003.
- Bögeholz, S., und S. Eggert. „Welche Rolle spielt Kompetenzdiagnostik im Rahmen von Lehr-Lernprozessen?“ In *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen, Sonderheft ZfE, 18*, von D. Leutner, E. Klieme, J. Fleischer und H. Kuper, 59-64. Wiesbaden: Springer, 2013.
- Booch, G., J. Rumbaugh, und I. Jacobson. *Das UML Benutzerhandbuch*. München: Addison-Wesley, 2006.
- Brekke, H. E. *Semantik. Eine Einführung in die sprachwissenschaftliche Bedeutungslehre*. München: Fink, 1972.
- Büchter, A., und T. Leuders. *Mathematikaufgaben selbst entwickeln*. Berlin: Cornelsen, 2011.
- Burgess, B. H., und T. A. Hokel. *A Brief Introduction to the Zachman Framework*. Redondo Beach, CA: Framework Software, 1994.
- Carstensen, K.-U., C. Ebert, C. Endriss, S. Jekat, R. Klabunde, und H. Langer. *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Haltern am See: Spektrum Akademischer Verlag, 2001.
- Chen, P. „The entity-relationship-model - toward a unified view of data.“ *ACM Transactions on Database Systems*, 03 1976: 9-36.
- Chomsky, N. *Aspects of the Theory of Syntax*. Cambridge MA: MIT Press, 1965.
- Cimpian, E., S. Komazec, D. Lintner, C. Blamauer, und M. Evenson. *WFMC-D1.3 business process modeling ontology bpmo final version*. WFMC; SemBiz Deliverable, 2008.
- Cronbach, L. J. *Essentials of psychological testing*. New York: Harper, 1949.
- Csapo, B. „Goals of Learning and the Organization of Knowledge.“ In *Kompetentmodellierung*, von E. Klieme, D. Leutner und M. Kenk, 12-27. Weinheim und Basel: Beltz, 2010.
- Demars, C. *Item Response Theory*. Oxford: Oxford University Press, 2010.
- Deutscher Bildungsrat. *Empfehlungen der Bildungskommission zur Neuordnung der Sekundarstufe II*. Stuttgart, 1974.
- Edelmann, D., und R. Tippelt. „Kompetenzentwicklung in der beruflichen Bildung und Weiterbildung.“ In *Kompetenzdiagnostik, ZfE, Sonderheft 8*, von M. Prenzel, I. Gogolin und H.-H. Krüger, 129-146. Wiesbaden: VS Verlag für Sozialwissenschaften, 2007.

- Eicker, Stefan. *Enzyklopädie der Wirtschaftsinformatik - Systementwicklung*. 08. 05 2015.
<http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/is-management/Systementwicklung> (Zugriff am 15. 08 2015).
- Embretson, S. E., und S. P. Reise. *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2000.
- ETH - Zürich nach Bloom ... 2015. http://www.educ.ethz.ch/emmetten_2015/Schalk_Taxonomie
 (Zugriff am 13. 08 2015).
- European Commission. *Learning Opportunities and Qualifications in Europe*. 20. 08 2015.
<https://ec.europa.eu/ploteus/> (Zugriff am 20. 08 2015).
- Fleischer, J., K. Koeppen, M. Kenk, E. Klieme, und D. Leutner. „Kompetenzmodellierung: Struktur. Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms.“ In *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen*, von D. Leutner, E. Klieme, J. Fleischer und H. Kuper, 5-22. Wiesbaden: Springer VS, 2013.
- Frege, G. „Über Sinn und Bedeutung.“ *Zeitschrift über Philosophie und philosophische Kritik*, 1892: 25-50.
- Freundlich, R. *Einführung in die Semantik*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1972.
 —. *Sprachtheorie*. Wien/New York: Springer, 1970.
- Frey, A., und N.-N. Seitz. „Mehrdimensionale adaptive Kompetenzdiagnostik: Ergebnisse zur Messeffizienz.“ In *Kompetenzmodellierung*, von E. Klieme, D. Leutner und M. Kenk, 40-51. Weinheim und Basel: Beltz, 2010.
- Gartz, K. *Spiegel online: Klick, klick, Error - Uni-Klausuren am PC*. 2014.
<http://www.spiegel.de/unispiegel/studium/e-klausur-am-computer-uni-und-fh-schaetzen-pruefung-am-pc-a-983353.html> (Zugriff am 12. 08 2015).
- Gelman, R., und J. G. Greeno. „On the Nature of Competence: Principles for Understanding in a domain.“ In *Knowing, Learning and Instruction, Essays in Honor of Robert Glaser*, von L. B. Resnick, 125-186. Hillsdale, NJ: Lawrence Erlbaum, 1989.
- Gerrig, R. J., und P. G. Zimbardo. *Psychologie*. München: Pearson, 2008.
- Gnahn, D. *Kompetenzen - Erwerb, Erfassung, Instrumente*. Bonn: Deutsches Institut für Erwachsenenbildung, 2007.
- Gronau, N., und J. Fröming. „KMDL - Eine semiformale Beschreibungssprache zur Modellierung von Wissenskonzersionen.“ *Wirtschaftsinformatik*, 5 2006: 349-360.
- Gruber, T. R. „A Translation Approach to Portable Ontologies.“ *Knowledge Acquisition*, 06 1993: 199-220.
- Hammer, M., und J. Champy. *Reengineering the Corporation*. New York: HarperCollins, 1993.

- Hardy, I., et al. „Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter.“ In *Kompetenzmodellierung*, von E. Klieme, D. Leutner und M. Kenk, 115-125. Weinheim und Basel: Beltz, 2010.
- Hartig, J., und A. Frey. „Sind Modelle der Item-Response-Theory (IRT) das "Mittel der Wahl" für die Modellierung von Kompetenzen?“ In *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen*, von D. Leutner, E. Klieme, J. Fleischer und H. Kuper, 47-51. Wiesbaden: Springer VS, 2013.
- Hattie, J. *Lernen sichtbar machen*. Baltmannsweiler: Schneider, 2013.
- Hausser, K. *Identitätspsychologie*. Berlin: Springer, 1995.
- Heid, H. „Erziehung.“ In *Erziehungswissenschaft, Ein Grundkurs*, von D. Lenzen, 43-68. Reinbek, 1994.
- Hellmig, Lutz, und andere. *Universität Rostock, Polyvalente Aufgaben*. 25. 11 2015.
<http://www.mathe-mv.de/publikationen/sekundarstufe-i/polyvalente-aufgaben/> (Zugriff am 20. 03 2016).
- Hermes, C., und P. Vaßen. *Entwicklung kompetenzorientierter Aufgaben für dem Mathematikunterricht*. Berlin: Cornelsen Verlag, 2012.
- Hogrebe, F., N. Gehrke, und M. Nüttgens. „Gebrauchstauglichkeit semiformaler Modellierungssprachen für das Anforderungsmanagement.“ In *Modellierung 2010*, von G. Engels, D. Karagiannis und H. C. Mayr, 31-46. Klagenfurt: GI-Edition, 2010.
- Hsieh, Hsiu-Fang, und Sarah E. Shannon. „Three Approaches to Qualitative Content Analysis.“ *Qualitative Health Research* 15, 2015: 1277-1288.
- Initiative Fortbildung e.V. „www.initiativefortbildung.de.“ 2013.
http://www.initiativefortbildung.de/pdf/2007/Ausbilden4/Bewerten_Beurteilen.pdf (Zugriff am 06. 08 2015).
- IQB. „Kompetenzstufenmodell zu den Bildungsstandards im Fach Deutsch.“ 24. 03 2015.
https://www.iqb.hu-berlin.de/bista/ksm/KSM_Deutsch_Spra.pdf (Zugriff am 21. 08 2015).
- Jahn, D. „Durch das praktische Gesalteln von didaktischen Designs nützliche Erkenntnisse gewinnen: Eine Einführung in die Destaltungsforschung.“ *Wirtschaft & Erziehung*, 1 2014: 3-15.
- Jurecka, A. „Introduction to the Computer-based Assessment of Competencies.“ In *Assessment of Competencies in Educational Contexts*, von J. Hartig, E. Klieme und D. Leutner, 193-213. Göttingen: Hogrefe, 2008.
- Kaufhold, M. *Kompetenz und Kompetenzerfassung*. Wiesbaden: VS Verlag für Sozialwissenschaften, 2006.
- Klieme, E., A. Bürgermeister, B. Harks, W. Blum, D. Leiß, und K. Racoczy. „Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht.“ In *Kompetenzmodellierung*, von E. Klieme, D. Leutner und M. Kenk, 64-74. Weinheim und Basel: Beltz, 2010.

- Klieme, E., J. Hartig, und D. Rauch. „The Concept of Competence in Educational Contexts.“ In *Assessment of Competencies in Educational Contexts*, von J. Hartig, E. Klieme und D. Leutner, 3-22. Cambridge und Göttingen: Hogrefe & Huber, 2008.
- Klieme, E., und D. Leutner. „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen.“ *Zeitschrift für Pädagogik*, 6 2006: 876-903.
- Klieme, E., und J. Hartig. „Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs.“ In *Kompetenzdiagnostik, ZfE, Sonderheft 8*, von M. Prenzel, I. Gogolin und H.-H. Krüger, 11-29. Wiesbaden: VS Verlag für Sozialwissenschaften, 2007.
- Klinger, U. „Mit Bildungsstandards Unterrichts- und Schulqualität entwickeln.“ In *Friedrich Jahresheft*, 130-143. Seelze: Erhard Friedrich Verlag, 2005.
- Kreschnak, H. *Computergestützte Analysen von Schülerleistungen*. Berlin: Volk und Wissen, 1985.
- Kuhn, J., und C. Rohrer. „Approaching ambiguity in real-life sentences - the Application of an Optimality Theory-inspired constraint ranking in a large scale LFG grammar.“ 6. *Fachtagung der Sektion Computerlinguistik DGfS-CL 1997*. 1997. <http://elib.uni-stuttgart.de/opus/volltexte/1999/518/> (Zugriff am 17. 08 2015).
- Kultusministerium, Niedersächsisches. *Kerncurriculum das Gymnasium, Gesamtschule, Fachgymnasium, Abendgymnasium, Kolleg - Mathematik*. Hannover: Niedersächsisches Kultusministerium, 2009.
- Kultusministerkonferenz. „Einheitliche Prüfungsanforderungen in der Abiturprüfung Deutsch.“ *Beschluss der Kultusministerkonferenz vom 1.12.1989 in der Fassung vom 24.05.2002*. KMK, 2002.
- . *Handreichungen für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe*. Bonn, 1999.
- . „KMK - Bildungsstandards Abitur Mathematik.“ 18. 10 2012. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf (Zugriff am 20. 08 2015).
- Kuropka, D. *Modelle zur Repräsentation natürlichsprachlicher Dokumente*. Berlin: Logos, 2004.
- Lehner, F. *Wissensmanagement*. München/Wien: Carl Hanser, 2009.
- Leutner, D., E. Klieme, H. Kuper, und J. Fleischer. „Editorial: Kompetenzmodelle zur erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen.“ In *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen, Sonderheft ZfE, 18*, von D., Klieme, E., Fleischer, J., Kupwe, H. (Hrsg.) Leutner, 1-4. Wiesbaden: Springer, 2013.
- Leutner, D., J. Fleischer, C. Spoden, und J. Wirth. „Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik.“ In *Kompetenzdiagnostik, ZfE, Sonderheft 8*,

- von M. Prenzel, I. Gogolin und H.-H. Krüger, 149-167. Wiesbaden: VS Verlag für Sozialwissenschaften, 2007.
- Meyer, H. *Was ist guter Unterricht?* Berlin: Cornelsen, 2004.
- Mislevy, R. J., und G. D. Haertel. „Implications of Evidence-Centered Design for Educational Testing.“ *Educational Measurement: Issues and Practice*, 4 2006: 6-20.
- Mislevy, R. J., und M. M. Riconscente. *Evidence-Centered Assessment Design: Layers, Structures and Terminology*. PADI Technical Report, Menlo Park: SRI International, 2005.
- Neubrand, M., E. Klieme, O. Lüdtke, und J. Neubrand. „Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung.“ *Unterrichtswissenschaft*, 1 2002: 100-119.
- „Niedersächsisches Schulgesetz.“ Niedersächsisches Kultusministerium, 2015.
- Nijssen, G. M., und T. A. Halpin. *Conceptual Schema and Relational Database Design*. Victoria: Prentice-Hall, 1989.
- Nonaka, I., und H. Takeuchi. *The Knowledge-Creating Company*. New York: Oxford Press, 1995.
- Object Management Group (OMG). *Meta Object Facility (MOF) CORE (ISO/IEC 19508)*. ISO/IEC, 2014.
- Oldenburg, Universität. *Interview*. 05. 04 2004. <http://www.uni-oldenburg.de/fileadmin/us...tatt/download/Interview.do> (Zugriff am 14. 08 2015).
- Online, Spiegel. *Faire Noten - Software soll Schulaufsätze bewerten*. 2010. Spiegel-online, <http://www.spiegel.de/schulspiegel/wissen/0,1518,714844,00.html> (Zugriff am 19. 10 2015).
- Orth, H. *Schlüsselqualifikationen an deutschen Hochschulen. Konzepte, Standpunkte und Perspektiven*. Neuwied: Luchterhand, 1999.
- Pant, H. A. „Wer hat einen Nutzen von Kompetenzmodellen?“ In *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen*, von D. Leutner, E. Klieme, J. Fleischer und H. Kuper, 71-79. Wiesbaden: Springer VS, 2013.
- Pellegrino, J. W., N. Chudowsky, und R. Glaser. *Knowing what students know: The science and design of educational assessment*. Washington: National Academic Press, 2001.
- Perels, Franziska. „Modellierung und Messung fächerübergreifender Kompetenzen und ihre Bedeutung für die Bildungsforschung.“ In *Kompetenzmodellierung, Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*, *ZfE*, 56. Beiheft, von E. Klieme, D. Leutner und M. Kenk, 270-273. Weinheim und Basel: Beltz, 2010.
- phpBB. *Social Sciences Survey*. 01. 12 2012. <https://www.socisurvey.de/admin/index.php>.
- Poloczek, J. *Kompetenzorientierter Informatikunterricht*. 2007. http://www.informatik.uni-frankfurt.de/~poloczek/MNU-Tagung-2007-09-27/kompetenzorientierter_informatikunterricht.pdf (Zugriff am 10. 10 2014).

- Pressman, R. S. *Software Engineering*. Hamburg: McGraw-Hill, 1989.
- Reiss, K., A. Heinze, und R. Pekrun. „Mathematische Kompetenz und ihre Entwicklung in der Grundschule.“ In *Kompetenzdiagnostik, ZfE, Sonderheft 8*, von M. Prenzel, I. Gogolin und H.-H. Krüger, 107-127. Wiesbaden: VS Verlag für Sozialwissenschaften, 2007.
- Rost, J. *Dorsch - Lexikon der Psychologie - Reliabilität*. 2013. <https://portal.hogrefe.com/dorsch/rasch-reliabilitaet/> (Zugriff am 12. 08 2015).
- . *Dorsch - Lexikon der Psychologie - Validität*. 2013. <https://portal.hogrefe.com/dorsch/validitaet/> (Zugriff am 2015. 08 12).
- Roth, H. *Pädagogische Anthropologie*. Hannover: Schroedel, 1971.
- Saldern, M. von. *Schulleistung 2.0: Von der Note zum Kompetenzraster*. Norderstedt: BoD, 2011.
- Salomaa, A.K. *Formale Sprachen*. New York: Academic Press, 1973.
- Schalles, C., M. Rebstock, und J. Creagh. „Ein generischer Ansatz zur Messung der Benutzerfreundlichkeit von Modellierungssprachen.“ In *Modellierung 2010*, von G. Engels, D. Karagiannis und H. C. Mayr, 15-30. Klagenfurt: GI-Edition, 2010.
- Schott, F. *Lehrstoffanalyse*. Düsseldorf: Pädagogischer Verlag Schwann, 1975.
- Schott, F., und S. A. Ghanbari. *Kompetenzdiagnostik, Kompetenzmodelle, kompetenzorientierter Unterricht*. Münster: Waxmann, 2008.
- Schwarze, J. *Grundlagen der Statistik I*. Herne/Berlin: Neue Wirtschaftsbriefe, 2001.
- Shazad, K., M. Elias, und P. Johannesson. „Towards Cross Language Process Model Reuse - A Language Independent Representation of Process Models.“ In *The Practice of Enterprise Modeling IFIP, Second IFIP 8.1 Working Conference, PoEM 2009*, von U. Frank, P. Loucopoulos, O. Pastor und I. Petrounias, 176-190. Stockholm/Heidelberg: Springer, 2009.
- Soffer, P., und Y. Wand. „On the notion of soft-goals in business process modeling.“ *Business Process Management Journal*, 2005: 663-679.
- Stock, W.G., und M. Stock. *Wissensrepräsentation*. München: Oldenbourg, 2008.
- Strobl, C. *Das Rasch-Modell*. München und Mering: Rainer Hampp Verlag, 2010.
- Stuckenschmidt, H. *Ontologien*. Berlin/Heidelberg: Springer, 2009.
- Teachsam. *KMK-Bildungsstandards*. 2010.
http://www.teachsam.de/bildungsstandards/Deutschland/d-bildungsstandards_2.htm#Operatoren (Zugriff am 20. 08 2015).
- Thomas, O. *Management von Referenzmodellen: Entwurf und Realisierung eines Informationssystems zur Entwicklung und Anwendung von Referenzmodellen*. Berlin: Logos, 2006.
- Toulmin, S. E. *The uses of Argument*. Cambridge MA: Cambridge University Press, 1958.

- Universität Tübingen. *Beschreibung des Kategoriensystems*. 27. 08 2012. <http://www.uni-tuebingen.de/fakultaeten/wirtschafts-und-sozialwissenschaftliche-fakultaet/faecher/erziehungswissenschaft/abteilungen/schulpaedagogik/aufgabenanalyse/beschreibung-des-kategoriensystems.html> (Zugriff am 20. 08 2015).
- Weinert, F. E. „Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit.“ In *Leistungsmessungen in Schulen*, von F. E. Weinert, 17-31. Weinheim und Basel: Beltz, 2001.
- WFMC. *Workflow Process Definition Interface - XML Process Definition Language (XPDL)*, WfMC-TC-1025. Lighthouse Point, FL: Workflow Management Coalition, 2002.
- White, R. „Motivation reconsidered: The concept of competence.“ *Psychological Review*, 1959: 297-333.
- Wieken, J.-H. „Towards a meta-model to utilize the measuring of competence.“ In *ECCE 2011 - European Conference on Cognitive Ergonomics*, von A. Dittmar und P. Forbrig, 245-248. Rostock: Universitätsdruckerei Rostock, 2011.
- Wieken, J.-H., und P. Forbrig. „Individual Education Feedback using Competence Structure Models.“ In *Proceedings of the 2014 European Conference on Cognitive Ergonomics ECCE 2014*, von C. Sary, 10:1-10:4. Wien: ACM, 2014.
- . „Operationalisierung von Kompetenzen mit einem Metamodell.“ *Delfi 2013 - Bremen*. 2013. <http://workshop-learning-analytics.f4.htw-berlin.de/wp-content/uploads/2013/09/WiekenForbrigKurz.pdf> 20.03.2014 (Zugriff am 01. 10 2014).
- Wikipedia - *Kritischer Realismus*. 06. 05 2015. [http://de.wikipedia.org/wiki/Realismus_\(Philosophie\)#Kritischer_Realismus](http://de.wikipedia.org/wiki/Realismus_(Philosophie)#Kritischer_Realismus) (Zugriff am 17. 08 2015).
- Wikipedia - *Persönlichkeitseigenschaft*. 14. 03 2015. <https://de.wikipedia.org/wiki/Pers%C3%B6nlichkeitseigenschaft> (Zugriff am 10. 08 2015).
- Wikipedia - *Semiotisches Dreieck*. 11. 06 2015. http://de.wikipedia.org/wiki/Semiotisches_Dreieck (Zugriff am 17. 08 2015).
- Wilhelm, O., und R. Nickolaus. „Was grenzt das Kompetenzkonzept von etablierten Kategorien wie Fähigkeit, Fertigkeit oder Intelligenz ab?“ In *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen, Sonderheft ZFE, 18*, von D. Leutner, E. Klieme, J. Fleischer und H. Kupwe, 23-26. Wiesbaden: Springer, 2013.
- Wilson, M. „Assessment, Accountability and the Classroom: A community of Judgement.“ In *Towards Cherences between Classroom Assessment and Accountability 103rd Yearbook of the national Society for the study of Education*, von M. Wilson, 1-19. Chicago: The University of Chicago Press, 2004.
- . *Constructing measures: an item response modeling approach*. Mahwah, New Jersey: Lawrence Erlabum Associates, 2004b.

Winter, F. *Leistungsbewertung - Eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen*. Hohengehren: Schneider, 2014.

Winther, E. *Kompetenzmessung in der beruflichen Bildung*. Bielefeld: Bertelsmann, 2010.

Zeuch, N., H. Geerlings, H. Holling, W. J. van der Linden, und J. P. Bertling. „Projekt Regelgeleitete Itementwicklung - Regelgeleitete Konstruktion von statistischen Textaufgaben: Anwendung von linear logistischen Testmodellen und Aufgabencloning.“ In *Kompetenzmodellierung*, von E. Klieme, D. Leutner und M. Kenk, 52-63. Weinheim und Basel: Beltz, 2010.

Thesen

1) Die angestrebten Ergebnisse eines Bildungsprozesses (Ziele, Kompetenzen, Inhalte) können in Modellform externalisiert werden.

Das in dieser Arbeit entwickelte Core-Situation-Personal-Modell (CSP) bildet die relevanten Inhalte mit ihren Verbindungen als Meta-Modell in UML ab. Das Modell ist zusätzlich in einem Datenbankmanagementsystem implementiert und wird beispielhaft für eine Domäne angewendet.

2) Domäneninhalte können in einer semiformalen, an natürliche Sprache angelehnten Form visualisiert werden.

Für die Inhaltsmodellierung ist eine FACT-orientierte Beschreibung eingeführt, wobei die FACTs als „Sätze“ mit mehreren Valenzen aufgefasst werden. Die gesamte Struktur kann direkt im Metamodell abgebildet werden.

3) Die Externalisierung in Modellform erlaubt die Konstruktion und Durchführung von Assessments mit Erklärungsfähigkeit.

Das Modell verbindet die drei Ebenen Domäneninhalt, Assessment und tatsächliche Resultate. Durch die sowohl innerhalb als auch zwischen diesen drei Ebenen miteinander verbundenen Strukturen können alle Ergebnisse sowohl hinsichtlich der Konstruktion aus diesen Inhalten abgeleitet, als auch hinsichtlich der Ergebnisse der Analyse nachverfolgt werden.

4) E-Assessments auf der Basis eines derartigen Modells führen summativ zu mit Ergebnissen aus klassischen Klausuren signifikant korrelierten Bewertungen und stochastischen Traits.

Die testweise Anwendung des CSP-Modells ergibt eine signifikante Korrelation sowohl der manifesten Bewertung (Score) als auch der stochastischen Kompetenzschätzung (Trait) mit einer vergleichbaren paper-and-pencil-Klausur.

5) Das Modell erlaubt die Gruppierung nach inhaltlichen und kompetenzorientierten Aspekten, die im Zusammenspiel mit psychometrischen Modellen auch bei kleinen Datenmengen detaillierte Aussagen über diese Aspekte zulassen.

Basierend auf der Modellstruktur konnten Items des Assessments zu Variablen gruppiert werden, für die ein manifester Score und eine stochastische Kompetenzschätzung möglich sind. So ist auch eine Unterstützung des Scores und letztlich eine detaillierte Aussage zu den einzelnen inhaltlichen Aspekten, beispielsweise zu „Parameter von Exponentialfunktionen“ möglich.

6) Feedback auf Individualebene und auf Lerngruppenebene im Sinne formativer Assessments können aus einem derartigen Modell abgeleitet werden.

Durch die Gruppierung zu Aspekten ist sowohl ein individuelles Feedback wie auch ein Feedback auf Gruppenebene hinsichtlich bestimmter Inhalte und Kompetenzen möglich. Inhalte lassen sich dabei auch auf Kompetenzen abbilden. Außerdem kann eine Verbindung zu einer weiteren Steuerung des Bildungsprozesses erfolgen.

7) Die Nutzung eines solchen Modells insbesondere im Zusammenspiel mit e-Assessments bietet die Möglichkeit der Einsparung erheblicher Aufwände.

Die Dokumentation erfordert einen deutlichen Initialaufwand, der durch Wiederverwendung und Mehrfachverwendung in verschiedenen Gruppen ausgeglichen werden sollte. Bei der Auswertung des Assessments ergeben sich selbst bei einer relativ kleinen Gruppe von 76 Personen bereits geschätzte Einsparungen von bis zu 90%.

Zusammenfassung

Sowohl das exponentiell wachsende verfügbare Wissen als auch die zunehmende Volatilität haben zu einer wachsenden Umorientierung der Zielsetzung von Bildungsprozessen in Richtung auf die Kompetenzorientierung geführt. Daher sind die Messung und die Bewertung der erreichten Kompetenzen eine wesentliche Grundlage für die Bewertung des Erfolges eines Bildungsprozesses. Ein fundiertes Feedback ist ein kritischer Erfolgsfaktor für die Justierung des gesamten Prozesses. Er erfordert eine Externalisierung von Zielen und Inhalten und die Orientierung von Assessments an dieser externalisierten Beschreibung.

Dafür wird ein dreistufiges Informationsmodell als Meta-Modell entwickelt, das als Core-Situation-Personal-Modell (CSP) bezeichnet und in der Unified Modeling Language (UML) beschrieben wird. Das Modell ist domänenorientiert, allerdings nicht auf eine bestimmte Domäne beschränkt. Das CSP-Informationsmodell besitzt die drei Ebenen CORE (Ziele, Kompetenzen, Domäneninhalte, Item-Prototypen), SITUATION (Assessments, Items, Response Types, manifeste Variablen, stochastische Variablen mit psychometrischen Modellen) und PERSONAL (individuelle Ergebnisse, ursprüngliche und typisierte Antworten der Assessment-Teilnehmer). Für die konkrete Anwendung wird komplementär zum CSP-Informationsmodell ein zielorientiertes Vorgehensmodell zur Erarbeitung der Modellinhalte, vom Domäneninhalt über die Erstellung des Assessments bis zu dessen Auswertung mit insgesamt 15 Schritten vorgeschlagen.

Für die Visualisierung des Domäneninhaltes wird eine graphische, semiformale Darstellung entwickelt, deren Elemente unmittelbar auf das Informationsmodell zurückführbar sind. Außerdem wird analog eine graphische Darstellung für die Beschreibung der Kompetenzermittlung und –messung aus den Ergebnissen des Assessments auf Basis der Domäneninhaltsbeschreibung entwickelt. Beispielhaft wird das Modell in eine relationale Datenbank übertragen, am Beispiel von Inhalten der Mathematikdomäne „Lineares und exponentielles Wachstum“ gefüllt und in einem e-Assessment mit 76 Teilnehmern angewendet.

Die Gesamtanalyse der Assessment-Ergebnisse zeigt eine hohe Korrelation zwischen den Bewertungen des e-Assessments im CSP und der parallelen klassischen Klausur und eine signifikante Korrelation der mittels des RASCH-Modells stochastisch ermittelten Kompetenzen (Traits) mit den manifesten Ergebnissen des e-Assessments (Scores). Das gilt auch für Teilkompetenzen und Inhalte. Damit besteht eine den klassischen Klausuren ähnliche, aber differenziertere Bewertungsmöglichkeit, die außerdem basierend auf den Modellinhalten eine Erklärungsfähigkeit und Lenkungsfunktion besitzt. Es sind sowohl formative als auch summative Aussagen basierend auf demselben Assessment und demselben Modell möglich.

Hinsichtlich des zeitlichen Aufwands stehen Vorarbeiten bei der Externalisierung des Domäneninhaltes den Einsparungen durch die erhöhte Transparenz und Wiederverwendbarkeit von Assessments als Ganzes oder von einzelnen Teilen gegenüber. Bei der Auswertung der Ergebnisse ergeben sich bereits bei einer Gruppe von 76 Personen geschätzte Einsparungen von 90%. Das Potenzial wächst hier mit der Größe der Gruppen und der Zentralisierung.

Das Modell sollte auch in anderen Domänen getestet werden, die für eine strukturierte Beschreibung weniger geeignet erscheinen als die Mathematik, was aber nicht im Bereich dieser Arbeit liegt.

Abstract

Since knowledge is growing exponentially and relevant parts undergo frequent changes, increasing emphasis is put on the term “competence” up to the point that it has become the main goal in educational processes. Therefore, the assessment of competencies is a prerequisite to evaluate the success of an educational process. It is essential to provide sound feedback to both the teachers as well as learners. So the objectives and domains contents need a description in a structured, traceable and explanatory model that can provide evidence-based feedback on an individual as well as on classroom level.

The development of the so called Core-Situation-Personal (CSP) information-model incorporates these factors. It is a meta-model using the Unified Modeling Language (UML) and is implemented in a database management system. Based on requirements such as flexibility to model different domains, traceability to provide explanations for results, the integration of psychometric models, an understandable language-oriented content description, reusability and an economy of resources, the three-tier CSP Information Model is developed. The three tiers are CORE (external and internal goals, competencies and competence hierarchies, competence levels, domain content, facts and objects, item prototypes), SITUATION (assessments, assessment structures, items, response types, manifest scoring model, psychometric model) and PERSONAL (actual responses, typed responses, personal data). Competencies may be modeled in different hierarchies and combined in matrices with themselves as well as with competence levels in taxonomies.

A complementary process model with 15 steps is added to facilitate the application of the model. Each step defines fundamental goals in terms of information structures and gives hints to necessary actions. In particular, to facilitate the content modeling a graphical language-oriented approach is proposed to visualize the meaning of the content in a pseudo-lingual manner. Another critical piece, namely the mapping of manifest response types to competence-specific competence levels has been visualized to provide transparency.

In a first test the model is applied to a mathematical domain with 76 learners in a 75-minute e-assessment. A first group of results show a significant correlation between manifest results of learners in the CSP-based e-assessment and the parallel paper-and-pencil-Assessment designed by teachers themselves, as well as a significant correlation between the estimated trait and e-Assessment results. As expected from their design, parts of the assessment items are grouped according to specific domain contents or competencies. Their analysis shows that through the description in the CSP it becomes possible to give further teaching and learning advice on specific domain subjects or expertise. Thus, based on the CSP it is possible to get a manifest score, an estimated trait of competence and formative feedback from the same assessment.

While initially the domain description is an additional time consuming input the reusability of the assessment as a whole or at least part of it will be worth the increased effort. Even more interesting is the effort necessary for correction which is remarkably less than for a classic paper-and-pencil-Assessment. Here a reduction of 90% in groups like in the example seems not to be unrealistic. Effort reductions is expected to increase or decrease with a respective change in group size.

Though the results are promising, the model has to be tested with other domains which are possibly not as suitable for a structured domain content description as mathematics. Here, further research is needed.

