

Esther Winther



Kompetenzmessung in der beruflichen Bildung

Habilitation



Esther Winther

Kompetenzmessung in der beruflichen Bildung

© W. Bertelsmann Verlag GmbH & Co. KG
Bielefeld 2010

Gesamtherstellung:
W. Bertelsmann Verlag, Bielefeld
wbv.de

Umschlagabbildung: Shutterstock



Bestell Nr.: 6004148
ISBN: 978-3-7639-4726-3 (Print)
ISBN: 978-3-7639-4727-0 (E-Book)
Printed in Germany

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Insbesondere darf kein Teil dieses Werkes ohne vorherige schriftliche Genehmigung des Verlages in irgendeiner Form (unter Verwendung elektronischer Systeme oder als Ausdruck, Fotokopie oder unter Nutzung eines anderen Vervielfältigungsverfahrens) über den persönlichen Gebrauch hinaus verarbeitet, vervielfältigt oder verbreitet werden.

Für alle in diesem Werk verwendeten Warennamen sowie Firmen- und Markenbezeichnungen können Schutzrechte bestehen, auch wenn diese nicht als solche gekennzeichnet sind. Deren Verwendung in diesem Werk berechtigt nicht zu der Annahme, dass diese frei verfügbar seien.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Inhaltsverzeichnis

1 Kompetenzmodellierung und Kompetenzmessung in der beruflichen Bildung: Ansatz, Forschungsdesiderat und Vision	5
1.1 Bildungspolitische Revisionsprozesse	6
1.1.1 <i>Revision des Steuerungssystems der beruflichen Bildung</i>	<i>6</i>
1.1.2 <i>Neufassung der europäischen Zertifizierung von Bildungsergebnissen</i>	<i>8</i>
1.1.3 <i>Revision der Input-Parameter der beruflichen Bildung</i>	<i>10</i>
1.2 Internationale Methoden und Auswertungsstandards	12
1.2.1 <i>„Measuring“ als Leitidee der Kompetenzdiskussion</i>	<i>12</i>
1.2.2 <i>Probabilistische Testtheorie</i>	<i>13</i>
1.3 Inhalt und Struktur der Arbeit	14
 2 Konstrukt und Modell: Theoretische Arbeiten der Kompetenzforschung . 17	
2.1 Kompetenzen: Ergebnisse von Bildungsprozessen?	17
2.1.1 <i>Behaviouristische Auffassung von Kompetenz</i>	<i>17</i>
2.1.2 <i>Kompetenz als Sammlung von allgemeinen Fähigkeiten (generic skills)</i>	<i>18</i>
2.1.3 <i>Kognitivistische Auffassung von Kompetenz</i>	<i>20</i>
2.2 Referenzrahmen zur Definition von Kompetenzen	22
2.2.1 <i>Referenzrahmen zur Analyse der kognitiven Struktur von Kompetenzen</i>	<i>23</i>
2.2.2 <i>Referenzrahmen zur Analyse der Reichweite von Kompetenzen</i>	<i>30</i>
2.3 Kompetenzmodelle: Zusammenhang von kognitiver Struktur und Reichweite	37
2.3.1 <i>Struktur, Graduierung und Entwicklung</i>	<i>37</i>
2.3.2 <i>Modellierung von Kompetenzen in den internationalen Vergleichsstudien</i>	<i>43</i>
2.4 Kompetenzdefinitionen und Kompetenzmodelle im Bereich der beruflichen Bildung	47
2.4.1 <i>Beispiele für Kompetenzmodelle der Berufs- und Wirtschaftspädagogik</i>	<i>48</i>
2.4.2 <i>Kompetenzstrukturmodell für die kaufmännische Bildung</i>	<i>53</i>
 3 Constructing Measures: Instrumentenentwicklung und deren Gestaltungskriterien	59
3.1 Assessment-Modelle	59
3.1.1 <i>Assessment Triangle</i>	<i>60</i>
3.1.2 <i>Evidence-centered Assessment Design (ECD)</i>	<i>64</i>
3.1.3 <i>Four Building Blocks</i>	<i>73</i>
3.2 Gestaltungskriterium 1: Das konzeptionelle Verständnis der Domäne	79
3.2.1 <i>Hypothese von der Domänenspezifität des Lehrens und Lernens</i>	<i>79</i>
3.2.2 <i>Domänen- vs. Fachverständnis</i>	<i>82</i>
3.2.3 <i>Domänenmodell für die kaufmännische Bildung</i>	<i>84</i>
3.3 Gestaltungskriterium 2: Die Konstruktrepräsentation	99
3.3.1 <i>Das Prinzip der Funktionalen Modellierung</i>	<i>100</i>
3.3.2 <i>Das Prinzip der Inhaltlichen Komplexität</i>	<i>104</i>

3.3.3	<i>Das Prinzip der Kognitiven Taxonomie</i>	108
3.3.4	<i>Modell der Instrumentenentwicklung für die kaufmännische Bildung</i>	117
4	Analysing Measures: Einführung in die Item Response-Theorie	119
4.1	Personenmerkmal, Testitem und Messwert: Zur Aussagekraft von Tests	120
4.1.1	<i>Validierung des Testverhaltens</i>	121
4.1.2	<i>Kurve des Testverhaltens</i>	125
4.2	Modelle der Item Response-Theorie	130
4.2.1	<i>IRT-Modelle für nominalskalierte Daten</i>	131
4.2.2	<i>IRT-Modelle für ordinalskalierte Daten</i>	136
4.3	Schätzung der Modellparameter	145
4.3.1	<i>Maximum Likelihood-Schätzung</i>	145
4.3.2	<i>Schätzung der Itemparameter</i>	149
4.3.3	<i>Schätzung der Personenparameter</i>	155
5	Interpreting Measures: Studien zur Messung kaufmännischer Kompetenz	163
5.1	Studie I: Kaufmännische Kompetenz in vollzeitschulischen Ausbildungsgängen	164
5.1.1	<i>Kaufmännisches Vorwissen als Kategorie des Domänenwissens</i>	165
5.1.2	<i>Literacy-Konzeption und curriculare Validität</i>	167
5.1.3	<i>Re-Analyse des kaufmännischen Vorwissenstests auf Basis der IRT</i>	175
5.2	Studie II: Kaufmännische Kompetenz in Ausbildungsgängen des Dualen Systems	196
5.2.1	<i>Definition und Umfang berufsfachlicher Kompetenz</i>	197
5.2.2	<i>Ausbildungsreale Authentizität und tätigkeitsspezifische Validität</i>	204
5.2.3	<i>Testitems berufsfachlicher Kompetenz</i>	216
5.2.4	<i>Ergebnisse</i>	222
6	Entwicklungsperspektiven der Kompetenzmodellierung und Kompetenzmessung in der beruflichen Bildung	249
6.1	Forschungsdesiderate der beruflichen Kompetenzmodellierung	250
6.1.1	<i>Struktur beruflicher Kompetenz</i>	250
6.1.2	<i>Graduierung beruflicher Kompetenz</i>	254
6.1.3	<i>Entwicklung beruflicher Kompetenz</i>	257
6.2	Forschungsdesiderate der beruflichen Kompetenzmessung	260
6.2.1	<i>Entwicklung und Validierung von Kompetenzmessverfahren</i>	260
6.2.2	<i>Blue Prints zur Konstruktion von beruflichen Assessments</i>	263
6.3	Kompetenzorientierung in Lehr- und Lernprozessen	264
6.3.1	<i>Kompetenzorientierte Unterrichtsgestaltung</i>	264
6.3.2	<i>Befunde zur Lehrprofessionalität</i>	269
	Literatur	273

1 Kompetenzmodellierung und Kompetenzmessung in der beruflichen Bildung: Ansatz, Forschungsdesiderat und Vision

Die Beschäftigung mit der Thematik Kompetenzmodellierung und Kompetenzmessung ist vorrangig durch zwei Entwicklungstendenzen motiviert. Zum einen liegt ein Interesse darin zu prüfen, inwieweit Methoden und empirische Auswertungsstandards, wie sie in den großen internationalen Schulleistungsstudien zur Anwendung gekommen sind und dabei permanent weiterentwickelt werden, für den Bereich der Berufsbildung fruchtbar gemacht werden können. Dieses ist ein zentrales Forschungsdesiderat, das erst in den letzten Jahren verstärkt in den Blick genommen wurde. Verbunden damit ist auch die Frage, in welchem Umfang die Berufsbildungsforschung mit ihrer starken Betonung einer Handlungsorientierung sowie der Notwendigkeit, Wissen ständig zu aktualisieren, ihrerseits die Kompetenzdiskussion anregen kann.

Zum anderen zeigen sich in der Berufsbildung Tendenzen, die eine Fokussierung auf kompetenzorientierte Forschung und Entwicklung begünstigen, wenn nicht sogar konsequent nach sich ziehen. Ausgangspunkt hierfür sind Ergebnisse von Analysen zur Arbeitsmarkt-, Sozial- und Berufsbildungspolitik, die schwerwiegende Probleme für das deutsche System der beruflichen Aus- und Weiterbildung aufdecken, die sowohl die individuellen persönlichen Entwicklungschancen für einen Großteil der Jugendlichen und (jungen) Erwachsenen als auch die Sicherung der Humanressourcen des Produktions- und Dienstleistungsstandortes Deutschland gefährden. Ersichtlich wird dies in der anhaltenden Forderung nach effizienten Bildungsprogrammen (vgl. die Website des BMBF). In diesem Zusammenhang kann die Einführung routinemäßiger Kontrollen der Systemeffizienz des deutschen Bildungswesens als „erstaunliche Wende“ (Leschinsky & Cortina, 2008, S. 22) bewertet werden. Diese Form outputorientierter Schul- und Bildungsevaluation ist dabei eng mit dem Kompetenzbegriff verknüpft. In den wissenschaftlichen und politischen Diskussionen – und in auf diesen folgenden Strukturprogrammen – wird *berufliche Kompetenz als Grundlage beruflicher Mobilität und eigenständiger Lebensführung* gesehen.

Diese Feststellung gilt nicht nur für Deutschland; so kann für den beruflichen Bereich auf die Entwicklung des Europäischen Qualifikationsrahmens (EQF) und des darauf beziehbaren Leistungspunktesystems (ECVET) verwiesen werden, die mehr Transparenz für die Berufsbildung in Europa schaffen und die Gleichstellung von beruflichen Abschlüssen (auch in Bezug auf allgemeine Abschlüsse) regeln sollen (aktuelle Informationen sind über die Websites des BMBF und des BIBB abrufbar; vgl. auch Achtenhagen, 2006; Sloane, 2008). Gleichstellungsregelungen stellen eine notwendige normative Definition zur Klassifikation von Ausbildungsgängen und -abschlüssen dar. Sollen sie ihren Hauptzweck, nämlich Glaubwürdigkeit und Vertrauen bei den Ausbildungseinrichtungen sowie institutionellen und individuellen Arbeitsmarktakteuren zu schaffen (Commission of the European Communities, 2005, pp. 5/6), erfüllen, muss bekannt sein, welche Kompetenzen in den unterschiedlichen europäischen Berufsbildungsprozessen

tatsächlich erworben werden. Dies ist insofern von zunehmender Wichtigkeit, als die Vorstellung von der einen, über das Erwerbsleben andauernden beruflichen Beschäftigung Illusion zu sein scheint; vielmehr zeigen Studien, dass Wechsel innerhalb, aber auch zwischen verschiedenen Berufsdomänen zu den charakteristischeren Karrierewegen gehören. Wolter (2007) legt beispielsweise deutlich dar, wie schwach der Zusammenhang zwischen Ausbildungsmarkt und Arbeitsmarkt ausgeprägt ist und welche Auswirkungen dies auf die berufliche Stabilität hat (vgl. auch Heidegger & Petersen, 2004). Die variierenden berufsbiographischen Stationen setzen im Sinne der Förderung eines lebenslangen Lernens Konzepte voraus, mit deren Hilfe sich unterschiedliche Aspekte der individuellen Entwicklung aufgreifen, festhalten und funktional verarbeiten lassen. Mit den im Rahmen dieser Arbeit vorgeschlagenen Ansätzen zur Kompetenzmodellierung und Kompetenzmessung im beruflichen Bereich wird versucht, einen Beitrag zur gezielten Förderung von Bildungsmobilität, die die Berufsmobilität umfasst, zu leisten.

1.1 Bildungspolitische Revisionsprozesse

Im Bereich der beruflichen Bildung können zumindest drei zentrale Entwicklungen bestimmt werden, die eine Diskussion um Kompetenzen und Kompetenzmodelle notwendig werden lassen: Revisionen des Steuerungssystems der beruflichen Bildung, Neufassungen der europäischen Zertifizierung von Bildungsergebnissen sowie die Revision der Input-Parameter der beruflichen Bildung.

1.1.1 *Revision des Steuerungssystems der beruflichen Bildung*

Revisionen des Steuerungssystems der beruflichen Bildung entsprechen dem Sachverhalt, dass sich die Arbeitsplatz- und Ausbildungsstrukturen in den letzten zwei Jahrzehnten grundlegend gewandelt haben: Die Ausbildungszahlen für das Duale System der Berufsausbildung gehen auf ca. 43 % zurück, das Vollzeitschulsystem stagniert bei ca. 17 %, während die Zahl der Jugendlichen, die weder eine qualifizierte Schulbildung noch eine qualifizierte Berufsbildung erhalten, sondern sich in so genannten „Maßnahmen“ befinden, auf 40 % zunehmen wird (Baethge & Wieck, 2006; Baethge, Solga & Wieck, 2007). Vergleichbare Strukturen lassen sich für alle europäischen Berufsbildungssysteme nachweisen (Dietrich, 2003). Für Deutschland gilt, dass im berufsbildenden Bereich zwar bislang keine zentralen strukturellen Veränderungen zu beobachten sind, aber vielfache (und noch immer unzureichende) soziale Ergänzungsmaßnahmen gefördert werden, durch die ein Netz außerbetrieblicher Bildungsinstitutionen geschaffen wird (Lerschinsky & Cortina, 2008). Das zentrale Problem liegt darin, dass durch diese Maßnahmen das Übergangssystem von der schulischen Ausbildung in eine berufliche Beschäftigung verlängert wird. Dieses Übergangssystem erfasst sowohl die Transitionen von der schulischen Ausbildung bis zur Aufnahme einer qualifizierenden Berufsausbildung als auch die Übergänge in ein Beschäftigungsverhältnis. Die Erforschung sowie die Gestaltung von Übergangssystemen ist ein zentrales Forschungsdesiderat der Berufsbildungsforschung sowie der Bildungspolitik. Es

ist offensichtlich, dass Programme einer Effizienzsteigerung der Übergangssysteme bereits deutlich vor dem eigentlichen Einstieg in die Ausbildungs- bzw. Beschäftigungslaufbahn ansetzen müssen. Die empirischen Befunde innerhalb Europas zeigen, dass Jugendarbeitslosigkeit und die damit verbundene Perspektivlosigkeit die zentrale gesellschaftliche Herausforderung darstellen. In Europa sind im Mittel nahezu 15 Prozent der Jugendlichen zwischen 15 und 19 Jahren ohne Beschäftigung; die Dunkelzahlen liegen deutlich darüber (European State Secretariat for the Economy, 2005). Das hat zur Konsequenz, dass sich formal kein Zuwachs an Qualifikationen für diese Jugendlichen ergibt. Die Zahlen zeigen auch, dass die Angst vor einem Nichteinfinden in das Beschäftigungssystem lange vor dem Eintritt in den eigentlichen Übergangsprozess präsent ist und das Lernen und Leisten behindert (Warner Weil, Wildermeersch, Jansen & Percy-Smith, 2005).

Theoretisch-konzeptionelle Forschungen zu Fragen der Transitionen von der Erstausbildung in das Erwerbsleben befassen sich mit Grundannahmen eines meritokratischen Bildungssystems (u. a. Fend, 1981; Raffe, 2008), in dem die Leistungen und Fähigkeiten eines Lernenden über Entwicklungschancen und den sozialen Status entscheiden. In diesem Zusammenhang befasst sich u. a. die Schweizer TREE-Studie (Transitionen von der Erstausbildung ins Erwerbsleben, TREE, 2008) neben der Problematik der Bewältigung von Übergängen und den zentralen Fragen der schulischen und beruflichen Sozialisation insbesondere mit soziologischen Theorien der Statusreproduktion sowie den primären und sekundären Effekten sozialer Ungleichheit (vgl. hierzu auch Baumert & Schümer, 2002). Arbeitspsychologische Theorien (u. a. Baethge & Bartelheimer, 2005) sowie transition-from-school-to-work-Modelle (OECD, 2000) zielen vor diesem Hintergrund auf ein Umdenken im Hinblick auf die Einstellungspolitik der Betriebe und verweisen ausdrücklich auf eine Vielzahl von Übergängen innerhalb beruflicher Karrieren. Damit ist die Vorstellung einer lebenslangen festen Beschäftigung als eher unwahrscheinlich anzusehen (Stamm, 2006). Dies hat weitreichende, zum Teil kaum überschaubare Folgen zum einen für die Sozialsysteme der einzelnen Staaten und zum anderen für die derzeit bestehenden Berufsbildungssysteme. Hierbei ist von großer politischer Bedeutung, dass die anfallenden Kosten für ineffiziente Übergangssysteme derzeit nicht verlässlich kalkuliert werden können (Descy & Tessaring, 2001; auch Mühlemann, Wolter, Fuhrer & Wüest, 2007). Insbesondere die makro- und mikroökonomischen Kohärenzen lassen sich über die vielfältigen Indikatoren der verschiedenen europäischen Organisationen und/oder Studien (OECD; Cedefop; EUROSTAT) bis jetzt nicht umfassend aufklären (vgl. auch Hannan & Werquin, 2000).

Angeichts eines qualitativ wie quantitativ nicht hinreichend ausdifferenzier- ten Ausbildungsplatzangebots und der unzureichenden Ergebnisse der verschiedenen „Maßnahmen“-Programme der europäischen Berufsbildungslandschaft geht es für die deutsche Entwicklung der Berufsbildung darum sicherzustellen, dass aufgrund der knappen Ausbildungsgelegenheiten im Dualen System vor allem für die beruflichen Vollzeitschulen und die „Maßnahmen“-Programme Kon-

zepte entwickelt werden, die tragfähig genug sind, berufliche Beschäftigungschancen zu eröffnen. Aus pädagogischer Perspektive liegen Anknüpfungspunkte vor allem darin, Übergangssysteme als Qualifikationssysteme zu gestalten. Für den kaufmännisch-verwaltenden Bereich, der in dieser Arbeit behandelt wird, geht es vor allem darum, dass die hauptsächlichen Mängel der Ausbildung, wie die Abstraktheit, die Linearisierung, die Zerstückelung und Parzellierung der Ziele und Inhalte der Lehrgänge und Richtlinien – und damit des Unterrichts und der Ausbildung – überwunden werden. Hinzu tritt, dass infolge der Nutzung von unternehmensweiten Informationssystemen, wie z. B. SAP, sehr viele interne Unternehmensprozesse nicht mehr direkt erfahrbar sind. Dies erschwert, aufgrund der lediglich in geringem Maße oder gar nicht gegebenen Anschaulichkeit, das Lernen, aber auch das spätere aktive, verständnisgeleitete und zugleich verantwortungsbewusste Handeln am Arbeitsplatz. Hier gewinnen kompetenzbasierte Überlegungen für den Ausbildungs-, aber auch für den Weiterbildungsbereich an Bedeutung. Mit der Bestimmung der Elemente eines Kompetenzkonzeptes für die kaufmännische berufliche Bildung sollen Grundlagen dafür geschaffen werden, dass Kompetenzen in den verschiedenen Institutionen der kaufmännisch-verwaltenden Ausbildung zielgerichtet und effektiv aufgebaut werden können. Hierfür bietet sich im Rahmen der zu wählenden Kompetenzkonzeption vor allem die Förderung des Entscheidungsverhaltens an.

1.1.2 Neufassung der europäischen Zertifizierung von Bildungsergebnissen

Mit Hilfe des „European Qualifications Framework“ (EQF) sollen im Hinblick auf die Anerkennung bzw. Zertifizierung von Bildungsergebnissen erreichte Lernergebnisse als Qualifikationen über die Kategorien „Wissen“, „Fähigkeiten“ und „Kompetenzen“ klassifiziert werden (u. a. Sellin, 2008) – von der Stufe 1: un- und angelernte Fähigkeiten, bis zur Stufe 8: Graduiertenwissen. Entscheidend für die Realisierung dieses Vorschlags sind zwei Sachverhalte: Zum einen sind Qualifikationen, die im allgemeinbildenden Bereich und in der beruflichen Aus- und Weiterbildung erworben werden, mit ihren Niveaus von Teilanforderungen und Kompetenzprofilen innerhalb eines Klassifikationsschemas aufeinander zu beziehen; zum anderen sollen hier auch Qualifikationen eingeordnet werden, die als Ergebnisse informellen Lernens (d. h. eines Lernens außerhalb von hierfür ausgewiesenen Institutionen und Lernorten) zu gelten haben (vgl. Baethge, Achtenhagen, Arends, Babic, Baethge-Kinsky & Weber, 2006; Björnåvold, 2000). So heißt es in der BiBB-Expertengruppe:

„Im Bildungssystem sollen sich mit einem kompetenzorientierten Referenzrahmen Bildungsgänge und Bildungsangebote europaweit abschlussneutral darstellen und aufeinander beziehen lassen. Lernergebnisse (auch informell erworbene) sollen sich qualifikations- und bildungsbereichsübergreifend anrechnen und Lernprozesse lernort-unabhängig und grenzüberschreitend organisieren lassen“ (Hanf et al., 2007, S. 1).

Für einen konstruktiven Umgang mit dem European Qualifications Framework sind neben den unzweifelhaft gegebenen Limitationen vor allem dessen prinzipielle Möglichkeiten zu analysieren; die Erwartungen des Council of Education Ministers sind in dieser Beziehung sehr hoch¹: (1) Der EQF ist ein Instrument, das die Vergleichbarkeit von Qualifikationen sichert; (2) er ist entworfen worden, um Transparenz zu sichern und nationale Qualifikationen und Systeme im internationalen Austausch verständlich zu machen; (3) er erleichtert die Durchsetzung einer auf EUROSTAT basierenden komparativen Statistik, und er soll (4) als Input verstanden werden, nationale Bildungs- und Qualifikationssysteme zu sichten und zu bewerten. In der Summe fällt die aktuelle Betrachtung des EQF jedoch nüchterner aus. Zweifellos ist ein besonderes Merkmal, dass Lernergebnisse basierend auf den Kategorien „Wissen“, „Fähigkeiten“ und „Kompetenzen“ ohne Bezug auf die Ausbildungsinstitution, die Ausbildungsdauer und die Ausbildungsform klassifiziert werden (sollen). Es zeigt sich jedoch, dass dies ohne sektor- bzw. bereichsspezifische Referenzrahmen für die einzelnen Berufe kaum plausibel zu realisieren ist (Sellin, 2008); denn es sind zumindest drei Perspektiven in die Betrachtung zu integrieren: (1) die hierarchische Struktur des Bildungssystems, (2) die zu erwartenden beruflichen Anforderungen und Funktionen sowie (3) der zu beschreibende Verlauf des Erwerbs von „Wissen“, „Fähigkeiten“ und „Kompetenzen“ (in Anlehnung an Markowitsch & Luomi-Messerer, 2008). Seit der ersten Vorstellung des EQF im Jahr 2005 sind zwar wenige, dafür aber sehr zentrale Elemente revidiert worden. Als Meilenstein kann die Consultation to Recommendation Conference (Budapest, Februar 2006) angesehen werden,² bei der folgende Beschlüsse gefasst wurden: (1) Die EQF-Tabelle, die die Definitionen der verschiedenen Klassifizierungslevel enthält, ist neu zu gestalten, wobei insbesondere eine präzisere Auswahl der Deskriptoren anzustreben ist. (2) Die Lernergebnisse sollen als Kompetenzen definiert werden. Die Kompetenzdefinition richtet sich an dem Verständnis einer zielbewussten Handlungsfähigkeit in beruflichen und sozialen Kontexten aus; (3) die Äquivalenz von beruflicher und schulischer Kompetenz ist deutlicher und gesicherter herauszustellen, was insbesondere eine Revision der „akademischen“ Klassifizierungslevel (Level 6 bis 8) nach sich zieht.

Als Grundlage für künftige Revisionen scheinen angemessene Operationalisierungsprozesse notwendig zu sein, für deren Umsetzung eine theoretisch und empirisch gesicherte Modellierung gerade von Kompetenz von besonderer Wichtigkeit ist. Der im European Qualifications Framework verwendete Kompetenzbegriff muss auch nach der Neuformulierung aus dem Jahr 2006 als nicht stimmig operationalisiert angesehen werden. Er erschöpft sich in der Aufzählung von Kompetenzkomponenten, die sich nach wie vor nicht stringent auf die Ausarbeitungen des Framework beziehen lassen (u. a. Böhlinger, 2008).

¹ Der Vorschlag des Council of Education Ministers von November 2006 liegt als Download vor: http://ec.europa.eu/education/policies/educ/eqf/com_2006_0479_en.pdf

² Eine Zusammenfassung der Konferenzergebnisse liegt in digitaler Form zum Download vor: http://ec.europa.eu/education/policies/educ/eqf/back_en.html

„Competence means the proven ability to use knowledge, skills and personal, social and/or methodological abilities, in work or study situations and in professional and personal development. In the context of the European Qualifications Framework, competence is described in terms of responsibility and autonomy“ (European Commission, 2008, p. 11).

Für Operationalisierungen kommt nach Sichtung der einschlägigen Literatur erschwerend hinzu, dass trotz Bemühungen der OECD (vgl. Rychen & Salganik, 2001) auch im internationalen Vergleich kein einheitlicher, allgemein akzeptierter Kompetenzbegriff existiert (vgl. auch Klieme & Hartig, 2008; detailliert Kapitel 2.1). So modifizieren beispielsweise Winterton, Delamare-Le Deist und Stringfellow in angelsächsischer Tradition das System des EQF wie folgt (vgl. auch Raffe, 2005):

„Knowledge (and understanding) is captured by cognitive competence; skills are captured by functional competence and „competence“ (behavioural and attitudinal, including meta-competencies) is captured by social competence“ (2004, p. 19).

In der deutschen Tradition wird an den Begriffen der Fachkompetenz, Methodenkompetenz, Personalkompetenz und Sozialkompetenz festgehalten; das in den französischsprachigen Ländern vertretene Kompetenzverständnis betont den Gleichklang von *savoir*, *savoir-faire* und *savoir-être* und zielt damit auf ein umfassendes Konzept im Sinne einer Einstellung ab, das sich schwer kategorisieren lässt.

Die zentrale Frage im Zuge der Kompetenzdebatte bleibt, wie Kompetenzmodelle aussehen und genutzt werden sollten, um – im Sinne eines einheitlichen European Qualifications Framework und entsprechend korrespondierender National Qualifications Frameworks – die Durchlässigkeit zwischen den nationalen Bildungssystemen zu fördern und die Anschlussfähigkeit der deutschen Berufe im europäischen Kontext sicherzustellen.

1.1.3 Revision der Input-Parameter der beruflichen Bildung

Die curricularen Input-Parameter im Bereich der kaufmännischen Aus- und Weiterbildung befinden sich in einem Revisionsprozess. So sehen beispielsweise die curricularen Vorgaben für den Ausbildungsberuf Industriekaufmann/Industriekauffrau sowie die Rahmenrichtlinien für das niedersächsische Fachgymnasium Wirtschaft die Integration der Unterrichtsfächer Betriebswirtschaft und Rechnungswesen vor und entsprechen damit der Forderung nach einer stärker systemischen Betrachtungsweise einzel- und gesamtwirtschaftlicher Zusammenhänge (vgl. u. a. Niedersächsisches Kultusministerium, 2001, S. 3; vgl. auch KMK, 2000a). Diese Ziel- und Inhaltsvorgaben, die über das Gemeinsame Ergebnisprotokoll mit der Wirtschaft und den Gewerkschaften abgestimmt sind, bilden zusammen mit den systematischen Anforderungen am Ausbildungs-/Arbeitsplatz die formalen Grundlagen für die Entwicklung von Kompetenzkonzepten. Mit den

zu leistenden Operationalisierungen helfen sie, „normativ“ die Ziele des jeweiligen Kompetenzkonzepts festzulegen und die Approximationsleistungen zur Erreichung dieser Ziele zu bestimmen. Diese Entwicklungen gehen einher mit einem Wandel der Systemsteuerung weg von einer Input-Orientierung hin zu einer Output-Steuerung.

Erst dieser Wechsel lässt in Bildungskontexten die Frage aufkommen, was tatsächlich als Lernresultat von den Schülern/Auszubildenden erreicht wird. Dabei ist eine Betrachtung, die auf der Inputseite ein Kompetenzkonzept auf das Subjekt bezieht und auf der Outputseite auf den Inhaltsbereich des Kompetenzerwerbs fokussiert, abzulehnen. Derartige Vorschläge (vgl. u. a. Lindemann, 2008; Dehnbostel & Lindemann, 2006) führen zu einer Zersplitterung, die sich nachhaltig schädlich auf die Curriculumentwicklung im Sinne neu zu gestaltender kompetenzbasierter Curricula auswirken dürfte. Oder technisch formuliert: Kompetenzerwerb und -entwicklung sind stets kontextbezogen und damit inhaltlich fixiert und an ein (lernendes) Subjekt gebunden.

Die internationalen Schulleistungsvergleiche – hier namentlich TIMSS/III und PISA – haben für die in diesen Studien erfolgreichen Nationen gezeigt, dass eine regelmäßige Kontrolle des realen Lernstandes mit entsprechendem zeitlichen Vorlauf schließlich auch eine individualisierte Optimierung des Lernprozesses auf Basis von Kompetenzstrukturmodellen möglich macht (Subjekt-Objekt-Bezug). Kompetenzstrukturmodelle erlauben unter dieser Perspektive einer empirisch fundierten Diagnose von Kompetenzen der Lernenden eine wirksame Modifizierung der Lerngelegenheiten (detailliert vgl. Kapitel 2.3). Jedoch muss an dieser Stelle betont werden, dass diese Output-Steuerung nur dann sinnvoll ist, wenn nicht nur im Hinblick auf die Abschlüsse der jeweiligen Ausbildungsgänge auf den Lernstand geschaut wird, sondern auch – im Sinne einer Verknüpfung von summativen und formativen Assessments (u. a. Shavelson, 2008) – auf davor liegende Lehr-Lernprozesse. Im Umkehrschluss bedeutet dies, dass Überprüfungen, die nur auf der Basis von Abschlussprüfungen erfolgen, diesem Steuerungsprinzip nicht entsprechen können (vgl. Schwiippert, 2005). Diese Argumentation steht im scheinbaren Widerspruch zu der im deutschsprachigen Raum vertretenen Ansicht einer Input-orientierten Kompetenzdiskussion, die sich als Opposition zu einer Output/Outcome-orientierten Debatte versteht (zu einem Input-orientierten Kompetenzverständnis in der Berufsbildung, vgl. Clement, 2003). In dem ausufernden Diskurs über Input-, Prozess- oder Output-organisierte Bildungssysteme werden jedoch eher paradigmatische Argumente gewählt, als – über empirische Forschung gestützt – konkrete Fragen zum Kompetenzerwerb bzw. zur Kompetenzentwicklung in den unterschiedlich akzentuierten Systemen beantwortet.

Die Revisionen des Input-Systems legen letztlich eine Neugestaltung (1) der beabsichtigten Lernergebnisse, (2) der diesen zugrunde liegenden Lehr-Lernprozesse und (3) der notwendigen Curricula im Sinne einer Neugestaltung mit Hilfe von Kompetenzmodellierungen nahe. Dabei ist zu fragen, welche Anforderungen ein Absolvent bewältigen können sollte, welche Kompetenz(en) zur

Anforderungsbewältigung notwendig wären und wie er diese Kompetenz(en) erlangen könnte und sollte.

1.2 Internationale Methoden und Auswertungsstandards

Die Ergebnisse der PISA-Studien (Programme for International Student Assessment), die seit ihrer ersten Runde (OECD, 2001; Baumert, Klieme, Neubrand, Prenzel, Schiefele, Schneider, Stanat, Tillmann & Weiß, 2001) sowohl bei politischen Verantwortungsträgern als auch in der breiten Öffentlichkeit große Beachtung finden, sind ein eindrucksvolles Beispiel dafür, wie empirische Befunde zu Veränderungen der Bildungslandschaft, der wissenschaftlichen Diskussion und von Reformbestrebungen beitragen können. Nachdem sich der „Bildungsrückstand“ der deutschen Schülerinnen und Schüler in harten Zahlen ausdrücken ließ, reagierten die Kultusminister der Länder und die Bundesregierung mit Programmen und Initiativen zur Verbesserung des Bildungssystems (KMK, 2001; BMBF, 2002); dabei gewinnen Fragen der Kompetenzmodellierung und Kompetenzmessung auch für die berufliche Bildung zunehmend an Bedeutung (vgl. Baethge, Achtenhagen, Arends, Babic, Baethge-Kinsky & Weber, 2006).

1.2.1 „Measuring“ als Leitidee der Kompetenzdiskussion

Das Ziel, berufliche Kompetenz zu messen, impliziert, dass das Ziel der Testanwendung möglichst die Quantifizierung entsprechender Merkmalsausprägungen sein sollte (Lienert, 1969). Die zentrale Frage ist, über welche Kompetenz/über welches Kompetenzprofil ein einzelner Proband verfügt. Die Beantwortung dieser Frage setzt voraus, dass Konstruktion, Anwendung, Auswertung und die Analyse der Testergebnisse aufeinander bezogen sein müssen, um die Aussagekraft der erzielten Ergebnisse sicherzustellen (Mummendey, 2003; Rost, 2004).

In der englischsprachigen Literatur wird dieses Ziel mit der Verbform „measuring“ in deutlicher Abgrenzung zum Substantiv „measurement“ beschrieben. *Measuring* charakterisiert dabei eine zentrale Einstellung des Testers zum Test, die sich (1) auf die theoretisch-konzeptionelle Herleitung des zu messenden Konstrukts, (2) auf das theoretisch begründete Antwortverhalten der Probanden, (3) auf die Konstruktion von Items, die sich auf das theoretische Konstrukt sowie auf das angenommene Antwortverhalten beziehen, und (4) letztlich auf die Auswertung und Analyse der Testergebnisse konzentriert (u. a. Wilson, 2005; 2008).

In der vorliegenden Arbeit werden die Fragen der beruflichen Kompetenzmessung vor diesem Hintergrund als vierstufiger geschlossener Prozess behandelt: (1) Wie lässt sich berufliche Kompetenz definieren, welche Operationalisierungen sind zielführend? Hierbei sind zwei Aspekte von zentraler Bedeutung: zum einen die Domäne und damit der Inhaltsbereich von beruflicher Kompetenz und zum anderen die Person als Merkmalsträger von Kompetenz. Auch hier kann die PISA-Studie Vorbild sein. Anders als bei vergleichbaren internationalen Schulleistungsstudien (TIMSS; Third International Mathematics and Science Study; vgl. Baumert, Bos & Lehmann, 2000) wird nicht auf curriculare Übereinstimmung ab-

gestellt, sondern im Sinne des Literacy-Konzepts auf funktionale Aspekte der Bildung. Der Fokus liegt damit auf Wissensbeständen und Fähigkeiten, von denen angenommen werden kann, dass sie für eine persönlich und wirtschaftlich zufriedenstellende Beteiligung des Individuums an der Gesellschaft benötigt werden. Dieses Konzept ist meines Erachtens für den Bereich der beruflichen Bildung besonders tragfähig. Die Anforderungen an Absolventen der beruflichen Bildung sind vorrangig am Arbeitsmarkt orientiert. Dies schließt eine persönliche Weiterentwicklung implizit mit ein, da nur hierdurch die Anschlussfähigkeit an sich verändernde Arbeits- und Lebensbedingungen ermöglicht wird. Eng damit verbunden sind (2) die Fragen, welche Aktionsradien, welche Handlungs- und Entscheidungsspielräume Absolventen beruflicher Bildung auf dem Arbeitsmarkt haben, und damit nach typischen Anforderungsfeldern, anhand derer sich „kompetentes“ Verhalten ablesen lässt, sowie (3) die Frage nach Testsituationen und Testitems, die berufliches Verhalten angemessen abbilden und dadurch einer Kompetenzmessung zugänglich machen. (4) Die Auswertung und Analyse erfolgen rückgebunden an die Anforderungsprofile des Arbeitsmarktes und die Handlungsparameter der Beschäftigten.

Ziel des *Measuring*-Prozesses ist es, die Eigenschaften sowohl der Testsituation als auch der Testperson möglichst fehlerfrei aufeinander zu beziehen. Hierbei stellt insbesondere die Analyse der Personeneigenschaften nach wie vor eine entscheidende Herausforderung dar, die aber gemeistert werden muss, wenn auf Basis der erzeugten Daten Schlussfolgerungen bezüglich der Unterschiede zwischen Personen bzw. Personengruppen gezogen werden sollen (Walter, 2005). Eine detaillierte Analyse der personenbezogenen Parameter der Testsituationen eröffnet hier weitere Optionen. Dieser Ansatz eignet sich im besonderen Maße für die Anwendung im Bereich der beruflichen Bildung, da – anders als im allgemeinbildenden Bereich – berufliche Handlungssituationen nicht künstlich erzeugt werden müssen, sondern bereits als beobachtbare Interaktionsprozesse vorliegen.

1.2.2 Probabilistische Testtheorie

Die mit PISA erzeugten Daten stammen aus psychometrischen Tests und wurden mit Verfahren der probabilistischen Testtheorie ausgewertet. Diese Auswertungsverfahren haben ein Berichtswesen, wie es PISA aufgezeigt hat, überhaupt erst ermöglicht und die Gestaltung dieses Large-Scale Assessment im intendierten Sinne realisierbar werden lassen. Da in der empirisch ausgerichteten Berufs- und Wirtschaftspädagogik, aber auch in anderen empirisch orientierten Forschungsfeldern in Deutschland die so genannte klassische Testtheorie dominiert, war die Rezeption der erhaltenen Ergebnisse im Hinblick auf ihre Herkunft und Akkumulation mit einigen Schwierigkeiten verbunden.

Nicht erst mit der im Jahr 2004 gestarteten Initiative eines international vergleichendes Large-Scale Assessment für den Bereich der beruflichen Bildung (VET-LSA; vgl. Baethge, Achtenhagen, Arends, Babic, Bathge-Kinsky & Weber, 2006) sind Fragen der probabilistischen Testtheorie auch für die Berufs- und

Wirtschaftspädagogik bedeutsam. Auch wenn sich die empirische Forschung im beruflichen Bereich seit langem mit Fragen der Leistungsbewertung (u. a. Beck, 2001), der Verlinkung von formativen und summativen Assessments (u. a. Wintner, 2006), der Gestaltung von Testumgebungen (u. a. Sembill & Seifried, 2007) und der diagnostischen Evidenz von Leistungstests (u. a. Wuttke & Wolf, 2007; Nickolaus, Knöll & Gschwendtner, 2007) beschäftigt, waren diese Studien jedoch bislang in der Regel von geringen Fallzahlen geprägt, und die eingesetzten Erhebungsinstrumente zeigten eine geringere Anzahl an Items auf. Für derartige Konstellationen werden aber die Vorteile, die die probabilistische Testtheorie im Vergleich zur klassischen Testtheorie aufweist, nicht hinreichend deutlich. Diese Vorteile kommen dagegen in den groß angelegten internationalen Studien wie PISA oder dem geplanten VET-LSA zur Geltung: Hier geht es darum, eine möglichst große Zahl von Testaufgaben, die verschiedene Personeneigenschaften im Sinne von Kompetenzprofilen erfassen sollen, möglichst vielen Testpersonen vorzulegen. Aufgrund der Testparameter – Längsschnittdesign, Verlinkung der Teilstichproben sowie Effekte auf die Aufgabenschwierigkeiten – wird eine Auswertung der Daten mit Hilfe der probabilistischen Testtheorie präferiert (Rost, 2006). Der zentrale Vorteil ist, dass dabei nicht alle Testpersonen alle Testitems bearbeiten müssen und dennoch eine alle Informationen ausschöpfende Auswertung möglich ist (Walter, 2005).

Ein Fokus dieser Arbeit liegt auf der Instrumentenkonstruktion zur Messung beruflicher Kompetenz. Um insbesondere zu einer genauen Berechnung von Differenzwerten der Kompetenz sowie zu genaueren Schätzungen der Testreliabilität zu gelangen, wird auf Verfahren der probabilistischen Testtheorie zurückgegriffen. In die Wahl der Testtheorie gehen auch Vorstellungen darüber ein, wie berufliche Kompetenz beschaffen ist: aus welchen Dimensionen sie besteht, welche Graduierungen sich zeigen etc. Der Rückgriff auf die probabilistische Testtheorie ermöglicht es, die Test-Items so zu gruppieren, dass sie jeweils eine eigene Dimension erfassen, und Personengruppen zu unterscheiden, in denen die jeweils gemessene Kompetenz über unterschiedliche Lösungsprofile abgebildet wird (im Detail vgl. Kapitel 4 sowie Kapitel 5).

1.3 Inhalt und Struktur der Arbeit

Das in dieser Arbeit vertretene Kompetenzverständnis ist strikt kontextbezogen und inhaltlich fixiert, es ist subjekt-orientiert und zielt damit auf konkrete Handlungen eines Individuums ab, und es ist ferner entwicklungsbasiert, da es Kompetenzen als erlernbar definiert.

Im Kapitel 2 werden Kompetenzen und Kompetenzmodelle vorgestellt. Dabei werden, ausgehend vom Kompetenzverständnis des allgemein bildenden Bereichs, auf Basis kognitiver Theorien und entsprechender Architekturvorstellungen die Struktur und die Reichweite von Kompetenzen in kaufmännisch-beruflichen Anforderungssituationen beschrieben. Die theoretische Ausarbeitung schafft die Grundlagen dafür, dass Aussagen zur Generalisierbarkeit und Über-

tragbarkeit von Kompetenzdefinitionen und -modellen im Bereich der beruflichen Bildung getroffen werden können.

Im Rahmen des dritten Kapitels werden Fragen der Instrumentenentwicklung bearbeitet. Bislang ist unbeantwortet, wie sich berufliche Kompetenzen messen lassen und wie Instrumente der beruflichen Kompetenzmessung zu gestalten wären, um die für berufliche Anforderungs- und Arbeitssituationen charakteristischen Handlungen und Argumentationsmuster einer systematischen Erfassung zugänglich zu machen. Die Debatte wird insbesondere dadurch erschwert, dass das auf dem Konzept der beruflichen Handlungskompetenz basierende, weitestgehend akzeptierte Kompetenzverständnis in der beruflichen Bildung unzureichend operationalisiert ist und generalisiert auf unterschiedliche Berufe und Berufsbereiche bezogen wird. Von daher werden zum einen Inhalte, Handlungsmuster und Arbeitsmethoden der kaufmännischen Domäne identifiziert und in Form eines kaufmännischen Domänenmodells als Grundlagen für eine Testentwicklung ausgearbeitet. Zum anderen wird die Ausgestaltung der Testbereiche im Hinblick auf eine authentische Abbildung komplexer beruflicher Anforderungssituationen erläutert. Dies erfolgt anhand einer Festlegung allgemeiner Konstruktions- und Gestaltungskriterien. Berufliche Anforderungssituationen sind demnach durch drei Gestaltungskriterien repräsentiert: durch den Inhalt, der die berufliche (Test-)Situation beschreibt, durch die kognitiven Verarbeitungsprozesse, die Auszubildende zur Bewältigung der beruflichen (Test-)Situation einzusetzen haben, sowie durch die Modellierungsleistungen der Auszubildenden, die die Grundlage dafür darstellen, dass die (Test-)Situationen auf beruflich relevante und zu bearbeitende Probleme bezogen werden können. Es wird erwartet, dass diese Gestaltungskriterien zentralen Einfluss auf die Komplexität und das Anspruchsniveau der (Test-)Situationen haben und dass Testentwickler diese Kriterien dazu nutzen können, Anforderungssituationen systematisch zu variieren und an die Leistungsfähigkeit der Probanden anzupassen. Für die Nutzung der Gestaltungskriterien als Blue Print für die Testentwicklung werden auf Basis der probabilistischen Testtheorie zwei Studien zur Messung kaufmännischer Kompetenz präsentiert.

Der Darstellung der Studien ist im Kapitel 4 eine Einführung in die probabilistische Testtheorie vorangestellt. Im Rahmen dieser Einführung werden die Grundannahmen der Item Response-Theorie präsentiert sowie ausgewählte Modelle einschließlich der Schätzung der Modellparameter vorgestellt. Der Schwerpunkt liegt insbesondere darauf aufzuzeigen, welche Modelle für eine gegebene Datenstruktur geeignet sind, wie Berechnungen mit Hilfe der Software *ConQuest* (Wu, Adams & Wilson, 1997; Wu, Adams, Wilson & Haldane, 2007) vorgenommen werden können und wie sich der erzeugte Output lesen und interpretieren lässt.

Im Kapitel 5 werden verschiedene Verfahren der Kompetenzmessung und Adaptationen der beruflichen Kompetenzmodellierung für einen vollzeitschulischen und einen dualen Bildungsgang präsentiert. Für den vollzeitschulischen Bildungsgang erfolgt die Messung kaufmännisch beruflicher Kompetenzen auf Basis eines Vorwissenstests, der an niedersächsischen Fachgymnasien Wirtschaft

eingesetzt wurde. Die Ergebnisse zeigen, dass über den Vorwissenstest domänenverbundene konzeptuale und prozedurale Fähigkeitsstrukturen von Auszubildenden erfasst werden können und dass es gelingt, Tests zu konstruieren, die die Heterogenität der Auszubildenden in beruflichen Bildungsgängen angemessen berücksichtigen.³ Für den dualen Ausbildungsgang werden mit Hilfe einer computerbasierten Unternehmenssimulation berufliche Kompetenzen von Auszubildenden zum Industriekaufmann/zur Industriekauffrau erfasst. Die Studie zeigt, dass mit authentischen Simulationen Prozesse und Handlungen abgebildet werden können, wie sie für kaufmännische Angestellte, und damit auch für kaufmännische Auszubildende, berufliche Realität sind. Mit der Unternehmenssimulation gelingt es, handlungsbasierte und verstehensbasierte Fähigkeitsstrukturen von Auszubildenden zu erfassen und auf konkrete berufliche Arbeits- und Geschäftsprozesse zu beziehen. Damit liegt ein Testinstrumentarium vor, für das der Anspruch erhoben werden kann, das Kompetenzverständnis der beruflichen Bildung in authentische, berufsbezogene Anforderungssituationen umgesetzt und verschiedene Facetten der beruflichen Kompetenz abgebildet zu haben.⁴ Die Ergebnisse der Studie zeigen ein differenziertes Bild der Leistungsfähigkeiten von Auszubildenden: Während die handlungsbasierten Fähigkeitsstrukturen der Auszubildenden stark ausgeprägt sind, zeigen sich im Hinblick auf verstehensbasierte Fähigkeitsstrukturen Defizite, die sich vergleichbar in nationalen und internationalen Schulleistungsstudien ablesen lassen.

Die Arbeit schließt in Kapitel 6 mit einem Ausblick, in dem Forschungsdesiderate skizziert und Entwicklungen der Kompetenzmodellierung und Kompetenzmessung in der beruflichen Bildung insbesondere im Hinblick auf Modelle der Kompetenzentwicklung sowie Instrumente der Kompetenzdiagnostik analysiert werden. Übergreifend werden darüber hinaus Impulse der beruflichen Kompetenzforschung im Hinblick auf Fragen der Kompetenzorientierung in schulischen und betrieblichen Lehr-Lernprozessen sowie der Lehreraus- und -weiterbildung aufgezeigt und diskutiert. Zwei Aspekte sind in diesem Zusammenhang von Bedeutung: (1) Aussagen zum Erwerb und Ausbau von beruflichen Kompetenzen müssen auf fachdidaktisch akzentuierten Kompetenzannahmen basieren, die unmittelbar in die Gestaltung von Lehr-Lernprozessen einfließen können, und (2) die fachdidaktische Kompetenz der Lehrenden ist als wesentlicher Einflussfaktor auf die Leistungen der Lernenden in Konzeptionen der Lehrerbildung besonders zu betonen.

³ Der Vorwissenstest wurde im Rahmen des DFG-Projekts „Integrierte Kompetenzentwicklung in den berufsbezogenen Fächern des Wirtschaftsgymnasiums“ konstruiert und im Rahmen des DFG-Projekts „Systemisches Verstehen von Geschäftsprozessen als kaufmännische Kompetenz“ auf Basis der probabilistischen Testtheorie re-analysiert.

⁴ Für das BMBF-Projekt „Konstruktvalidität von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz“ wurde die Unternehmenssimulation ALUSIM mit dem Ziel entwickelt, innovative Testverfahren für einen internationalen Vergleich in der beruflichen Bildung bereitzustellen sowie diese zugleich zu validieren und zu prüfen.

2 Konstrukt und Modell: Theoretische Arbeiten der Kompetenzforschung

2.1 Kompetenzen: Ergebnisse von Bildungsprozessen?

Die Diskussion um die Definition und den Bedeutungsgehalt des Kompetenzbegriffes sowie Fragen der Messung von Kompetenzen sind mit den Ergebnissen der großen Schulleistungsstudien wieder verstärkt in den Blick geraten; sie ist jedoch – auch in der Berufsbildung – keine neue Erscheinung. So erschien beispielsweise im Jahr 1922 im Rahmen der Reihe „Experimental Studies in Psychology and Pedagogy“ eine Arbeit von Morris Simon Viteles mit dem Titel „Job specifications and diagnostic tests of job competency designed for the auditing division of a railway company (a psychological study in industrial guidance)“. In dieser Schrift werden Fähigkeiten, Kenntnisse und besondere Eigenschaften (Vertrauenswürdigkeit, Gebundenheit an eine Familie) von Arbeitern im Eisenbahnbau beschrieben und Möglichkeiten aufgezeigt, wie sich diese durch Befragung ermitteln lassen (Viteles, 1922). Auch bereits in den frühen Lehrbüchern der Entwicklungspsychologie (u. a. Kretzschmar, 1912; Hoffmann, 1922) lassen sich Definitionen von Kompetenzen und dazugehörige statistische Verfahren ihrer Erfassung finden.

Die Sichtung der einschlägigen Literatur zeigt, dass insbesondere im internationalen Vergleich kein einheitlicher, allgemein akzeptierter Kompetenzbegriff existiert (vgl. Westera, 2001; Klieme & Hartig, 2008). Norris (1991) formuliert das wissenschaftliche Bestreben nach einer Kompetenzdefinition wie folgt:

„As tacit understandings of [the word competence has] been overtaken by the need to define precisely and operationalise concepts, the practical has become shrouded in theoretical confusion and the apparently simple has become profoundly complicated“ (Norris, 1991, p. 331).

Er identifiziert jedoch drei zentrale Forschungstraditionen, in die sich Kompetenzdefinitionen einordnen ließen: (1) Eine in den USA vertretene stark behaviouristische Auffassung, die auf der Ebene der Beobachtung von Performanz agiert; (2) eine in Großbritannien sowie Australien präferierte, vor allem auf Managementprobleme hin ausgerichtete Orientierung auf generic skills, die vergleichbar mit der in Deutschland geführten Diskussion um Schlüsselqualifikationen ist (u. a. Mertens, 1974); sowie (3) die insbesondere im OECD-Kontext vertretene kognitive Tradition.

2.1.1 Behaviouristische Auffassung von Kompetenz

Kompetenz als stark behaviouristisch akzentuiertes Konstrukt ist performanzbezogen und wird damit einer direkten Beobachtung zugänglich. Hierbei sind jedoch folgende Bedingungen zu erfüllen (vgl. Gagné & Bricks, 1984, p. 79): Das Verhalten und die das Verhalten auslösende Situation müssen möglichst exakt beschrieben werden. Darüber hinaus sind Aussagen über die Grenzen der stimulierenden Situation und damit die Grenzen der beobachtbaren Performanz zu

treffen. Die entscheidende Bedingung jedoch ist, dass Klarheit über die Fähigkeit (Art, Umfang, Ausprägung) bestehen muss, auf die durch Beobachtung der Performanz zurückgeschlossen werden soll.

Generell ist allen behaviouristischen Auffassungen von Kompetenz gemein, dass sie auf operationalisierbare Outcomes von Tests abstellen. Hierzu zählen das in Testsituationen direkt beobachtbare Wissen (knowledge) und Können (ability). Wissen und Können sind eng verbunden mit der Repräsentation von Fakten, Prozeduren, Prinzipien und Theorien bezogen auf eine spezifische Domäne (declarative knowledge). In diesem Zusammenhang stellen beispielweise Scheerer und Hager (1994) heraus, dass Kompetenz weniger als die Verfügbarkeit von Wissen und Können, sondern mehr als die effektive Anwendung des verfügbaren Wissens und Könnens in spezifischen Situationen zu verstehen ist. Damit ist gemeint, dass eine breite Wissens- und Könnensbasis nicht allein für eine erfolgreiche Performanz in – gewöhnlich – komplexen Situationen garantieren kann. Vielmehr zeichnet sich Kompetenz dadurch aus, dass eine Person aus dem ihr verfügbaren Repertoire das für eine spezifische Situation notwendige Wissen und Können so selektieren kann, dass effizientes und effektives Verhalten in der Situation resultiert.

2.1.2 Kompetenz als Sammlung von allgemeinen Fähigkeiten (generic skills)

Das Konzept der generic skills zeichnet sich dadurch aus, dass Kompetenz sich aus einer Vielzahl allgemeiner Fähigkeiten zusammensetzt, von denen angenommen wird, dass ihre Verfügbarkeit die Bewältigung von vornehmlich komplexen und arbeitsweltorientierten Situationen begünstigt. In der englischsprachigen Literatur werden über generic skills umfassende Konzepte und zusammenhängende Systeme von Kenntnissen, Fertigkeiten, Fähigkeiten und Einstellungen beschrieben, über die eine Person bei der Ausübung beruflicher Tätigkeiten verfügen sollte (Norris, 1991). Die Interpretationsrichtung ist dabei im Gegensatz zu der behaviouristischen Kompetenzauffassung nicht subjekt- sondern objektorientiert: (Berufliche) Anforderungen werden so zerlegt, dass sie sich als Bündel allgemeiner Kenntnisse, Fertigkeiten und Einstellungen beschreiben lassen. Skills im Sinne von Fähigkeiten werden als komplexe Konstrukte definiert:

„Skill is the ability to do something, but the word also connotes a dimension of increasing ability. Thus, while skill is synonymous with competence, it also evokes images of expertise, mastery, and excellence“ (Attwell, 1990, p. 433).

In der Folge entstanden unterschiedliche Zusammenstellungen von arbeitsbezogenen allgemeinen Fähigkeiten: Reich (1991, p. 84) stellt beispielsweise fest, dass sich in beruflichen Kontexten drei wesentliche Anforderungsprofile zeigen: *routine production services*, *in-person services* und *symbolic-analytic services*. Korrespondierend dazu identifizierten Carnevale, Gainer und Meltzer (1990, pp. 17–36) sieben Fähigkeitsgruppen, über die Arbeitnehmer verfügen sollten.

- Learning to learn
- Competence in reading, writing, and computation
- Communication skills: listening and oral communication
- Adaptability: creative thinking and problem solving
- Personal management: self-esteem, motivation and goal setting, and employability and career development
- Group effectiveness: interpersonal skills, negotiation, and teamwork
- Influence: organizational effectiveness and leadership

Andere vergleichbare Zusammenstellungen finden sich u. a. bei Chickering und Reisser (1993) oder auch im Rahmen des kanadischen MTA-Projekts (Making the Match between University Graduates and Corporate Employers; Evers, Rush & Berdrow, 1998).

In der deutschsprachigen Literatur findet diese Forschungstradition in der Debatte um Schlüsselqualifikationen ihre Entsprechung. Im Zuge der Bildungsexpansion in den 1970er Jahren wurden vermehrt Forderungen nach einer stärkeren Ausrichtung der Berufsbildung auf den Arbeitsmarkt laut. Nach Mertens (1974) sind Schlüsselqualifikationen solche Fähigkeits- und Einstellungsbündel, die es Individuen erlauben, sich einer ständig verändernden wirtschaftlichen, gesellschaftlichen und sozialen Umwelt anzupassen. Zur Sicherung der Employability, im Sinne der Beschäftigungsfähigkeit in der Arbeitswelt von heute und unter Berücksichtigung der Arbeitswelt von morgen, unterscheidet Mertens vier Gruppen:

- *Basisqualifikationen* sind „Qualifikationen höherer Ordnung“, zu denen beispielsweise logisches, kritisches und analytisches Denken, Kreativität oder das Erkennen von Zusammenhängen gehören.
- *Horizontalqualifikationen* erlauben es, auf Informationen zugreifen und diese möglichst effizient nutzen zu können. Dazu zählen vor allem solche Fähigkeiten, die sich auf die Suche, das Identifizieren, das Verstehen und das Verarbeiten von relevanten Informationen beziehen lassen.
- *Breitenelemente* bezeichnen allgemeine Kenntnisse und Fertigkeiten, die für wiederkehrende berufliche Tätigkeiten am Arbeitsplatz benötigt werden und universell in verschiedenen Situationen einsetzbar sind, wie beispielsweise Grundfähigkeiten (Lesen, Schreiben, Rechnen), EDV- und Sprachkenntnisse, Wissen über Arbeitsschutzregelungen oder Arbeitstechniken.
- *Vintagefaktoren* stellen auf Wissens- und Könnensunterschiede zwischen den Generationen ab, die es zu verringern gilt.

Obgleich die Ansicht, Schlüsselqualifikationen stellten das Individuum unter das Diktat der Innovation (v. Hentig, 2002), nicht vorbehaltlos zu akzeptieren ist, zeigt sich doch bereits in den vier Qualifikationsbereichen Mertens', an welchen Stellen Kompetenzkonzepte, die auf der Ansammlung allgemeiner Fähigkeiten aufbauen, Limitationen aufweisen (müssen): (1) Die Universalitätsannahme, die sicherstellen soll, dass dieses Konzept über verschiedene Berufsbereiche einsetzbar ist (vgl. hierzu Elliott, 1989), hat fast logisch zwingend zur Folge, dass sich

eine Messung kaum realisieren lässt (Norris, 1991). Dies zeigt sich auch daran, dass beispielsweise im Rahmen von Lehrerweiterbildungsmaßnahmen oder auch Trainingskonzepten in der Wirtschaft Kategorisierungen bzw. Katalogisierungen anstelle messrelevanter Operationalisierungen vorgenommen werden. (2) Die Generalitätsannahme, die sich auf den Abstraktionsgrad der Schlüsselqualifikationen bzw. generic skills bezieht, stellt deren Vermittelbarkeit bzw. Erlernbarkeit in Frage. Allgemeine Fähigkeiten lassen sich schwerlich abstrakt vermitteln. Nach Schelten (2004) können Schlüsselqualifikationen ohne Bezug auf spezifische Inhalte ihre Transferwirkung nicht einlösen und wären damit nicht generalisierbar. Ein zu starker Inhaltsbezug – und hier ist das Dilemma begründet – bedeutete jedoch, dass der Anspruch, über Schlüsselqualifikationen übergreifende Qualifikationen großer Reichweite zu erfassen, nicht eingehalten werden könnte – denn dies wäre ein Widerspruch der Universalitätsannahme.

2.1.3 Kognitivistische Auffassung von Kompetenz

Charakteristisch für kognitivistisch akzentuierte Kompetenzauffassungen ist die Performanz-Kompetenz-Abgrenzung, die sich auf Chomsky (1965) zurückführen lässt (vgl. auch Evans, Newstead & Byrne, 1993; Langford & Hunting, 1994). Nach Chomsky wird (linguistische) Kompetenz über kognitive Strukturen und Regeln repräsentiert, die notwendig sind, um spezifische Fähigkeiten (Sprache) zu erzeugen. Im Gegensatz dazu repräsentiert (linguistische) Performanz die beobachtbare Fähigkeit (Sprache) in der praktischen Anwendung. Verallgemeinert formulieren Sternberg und Ben-Zeev (2001):

„Competence is [...] emphasizing what people are ideally able to do“ (Sternberg & Ben-Zeev, 2001, p. 360). „Performance is [...] emphasizing what people typically do“ (ebd., p. 373).

Eine vergleichbare Auffassung findet sich auch bei Anderson (1992). Hier wird zwischen Kompetenz und Fähigkeiten (statt Performanz) unterschieden. Fähigkeiten repräsentieren den operativen Outcome und Kompetenzen die zugrundeliegende kognitive Funktion. Der Bedeutungsumfang der kognitiven Strukturen und Regeln ist dabei zunächst nicht fixiert. Gagné (1977) und auch Barnett (1994) verweisen beispielsweise auf strategische Fähigkeiten und Regeln. Demnach ist die Bewältigung neuer und auch unerwarteter Situationen ein wesentlicher Ankerpunkt innerhalb kognitiver Kompetenzkonzepte. Kompetenz ist demnach auch als Fähigkeit zu definieren, in spezifischen Settings oder Situationen effektive Entscheidungen treffen zu können (Kirschner, v. Vilsteren, Hummel & Wigman, 1997; zusammenfassend vgl. Westera, 2001).

In der Feasibility Study für ein international vergleichendes Large-Scale Assessment werden eine Reihe aktueller kognitiv geprägter Kompetenzdefinitionen vorgestellt (vgl. Baethge, Achtenhagen, Arends, Babic, Baethge-Kinsky & Weber, 2006, S. 18), die die von Weinert (2001a) angemahnte „konzeptuelle Inflation“ des Kompetenzbegriffes sehr schön illustrieren können (Abbildung 2.1).

Abbildung 2.1 Kompetenzdefinitionen

„The proven/demonstrated - and individual - capacity to use know-how, skills, qualifications or knowledge in order to meet usual - and changing - occupational requirements“ (Bjørnavold, 2000, S. 208).

„The ability to successfully meet complex demands in a particular context through the mobilization of psychosocial prerequisites (including both cognitive and non-cognitive aspects“ (Rychen & Salganik, 2003, S. 43).

„An ability that extends beyond the possession of knowledge and skills. It includes: i) cognitive competence involving the use of theory and concepts, as well as informal tacit knowledge gained experientially; ii) functional competence (skills or know-how), those things that a person should be able to do when they work in a given area; iii) personal competence involving knowing how to conduct oneself in a specific situation; and iv) ethical competence involving the possession of certain personal and professional values“ (Lisbon-to-Copenhagen-to-Maastrich Consortium Partners, 2004, S. 89)

„The ability to apply knowledge, know-how and skills in an habitual and/or changing work situation“ (CEDEFOP, 2003, S. 39).

Auch wenn in der einschlägigen Fachliteratur der Schluss gezogen wird, dass es unmöglich sei, eine einheitliche Theorie oder Definition zu formulieren, welche die Vielfältigkeit unterschiedlicher Ansätze in Einklang bringen könne (vgl. u. a. Norris, 1991; Ellström, 1997; Klieme, 2004), so ist es doch für den wissenschaftlichen Gebrauch in jedem Fall notwendig, für die entsprechende Studie eine präzisierende Definition vorzunehmen. Eine solche findet sich bei Weinert. Kompetenzen sind demnach

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

In seinem Gutachten für die OECD verweist Weinert darauf, dass Kompetenzen vor dem Hintergrund des angestrebten Anwendungs- bzw. Verwendungszwecks zu definieren seien. Hiermit wird eine inhaltliche Anbindung festgeschrieben, die vorrangig durch die Fachdidaktik zu leisten ist (vgl. auch Weinert, 1999). Neben dieser inhaltlichen Ausrichtung stellt Weinert heraus, dass die individuelle Ausprägung von Kompetenz durch die Facetten Fähigkeit, Wissen, Ver-

stehen, Können, Handeln, Erfahrung und Motivation bestimmt wird (Weinert, 1999; 2001a). Vor diesem Hintergrund bringt er die unterschiedlichen Bedeutungsumfänge kognitiver Kompetenzauffassungen in folgenden Katalog (hierzu Hartig & Klieme, 2006, S. 128):

- „Kompetenzen als generelle kognitive Leistungsdispositionen, die Personen befähigen, sehr unterschiedliche Aufgaben zu bewältigen,
- Kompetenzen als kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf bestimmte Klassen von Situationen und Anforderungen beziehen. Diese spezifischen Leistungsdispositionen lassen sich auch als Kenntnisse, Fertigkeiten und Routinen charakterisieren,
- Kompetenzen im Sinne der für die Bewältigung von anspruchsvollen Aufgaben nötigen motivationalen Orientierungen,
- Handlungskompetenzen als eine Integration der drei erstgenannten Konzepte, bezogen auf die Anforderungen eines spezifischen Handlungsfeldes, wie z. B. eines Berufes,
- Metakompetenzen als das Wissen, die Strategien oder die Motivation, welche sowohl den Erwerb als auch die Anwendung spezifischer Kompetenzen erleichtern,
- Schlüsselkompetenzen als Kompetenzen im funktionalen Sinn, die für einen relativ breiten Bereich von Situationen und Anforderungen relevant sind. Hierzu gehören z. B. muttersprachliche oder mathematische Kenntnisse.“

Als Gemeinsamkeit zwischen den verschiedenen Konzepten lässt sich festhalten, dass Kompetenzen, mit welchem Bedeutungsumfang sie auch immer definiert werden, in einem gewissen Maße über ähnliche Situationen generalisierbar sind. Diese wesentliche Eigenschaft von mehr oder weniger eng definierten Leistungsdispositionen hat eine lange Tradition in der Kognitionsforschung und stellt auch das verbindende Element der beiden im Folgenden zu beschreibenden Referenzrahmen dar. Mit Hilfe dieser Referenzrahmen soll eine Präzisierung des Kompetenzbegriffs vorgenommen werden.

2.2 Referenzrahmen zur Definition von Kompetenzen

Die im Folgenden zu führende Abgrenzung unterschiedlicher Kompetenzvorstellungen und Bedeutungsumfänge ist als Konstruktbildung zu verstehen. Hierbei findet zum einen eine Orientierung an Anderson (ACT; 1983; 1987; 1993; 1996; vgl. auch Anderson, Simon & Reder, 1997) statt, der eine Auseinandersetzung mit theoretischen Kompetenzkonzepten auf der Basis mentaler Repräsentationen und der kognitiven Architektur von Leistungen führt. Die ACT-Theorie (Adaptive Character of Thought) gibt Hinweise auf den Erwerb und die Struktur von kognitiven Leistungsdispositionen. Der Kompetenzrahmen, der sich daraus ableiten lässt, ermöglicht es, typische Eigenschaften von Kompetenzen in einem funktionalen Sinne herauszustellen. Zum anderen wird auf die Arbeiten von Greeno und Gelman (1989; vgl. auch Greeno, Riley & Gelman, 1984; Greeno, 1998) Bezug ge-

nommen. In Weiterentwicklung des linguistischen Kompetenzverständnisses (Chomsky, 1965; 1968; 1977; 1981) haben sich diese Autoren um eine kognitive Abstraktion des Kompetenzbegriffes bemüht und ein entwicklungspsychologisches Modell der „Nature of Competence“ entwickelt, das sich m. E. sehr gut in die aktuelle Kompetenzdebatte einfügt und Inhalte sowie Differenzen innerhalb dieser sowohl angemessen darstellen als auch begründet auflösen kann, da dieses Modell auf den Kontextbezug des Kompetenzaufbaus und -erwerbs fokussiert.

Beide hier gewählten Referenzrahmen stehen für verschiedene Auffassungen der Kognitionsforschung: Anderson als Vertreter der sog. micro-cognition; Greeno als Anhänger der situative cognition (vgl. u. a. Sternberg & Ben-Zeev, 2001; Schraagen, Militello, Ormerod & Lipshitz, 2008). Sie unterscheiden sich zentral in folgenden Punkten:

„On the one hand, the cognitivist camp argues that cognitive psychology can contribute greatly to the understanding of educational practice by examining people’s problem-solving strategies and representations. [...] On the other hand, the situative camp claims that cognitive psychology assumes that learning resides in the mind of the individual, instead of in the interaction between the individual and his or her environment. [...] Furthermore, the situative view argues that learning is tied to specific contexts...“ (Sternberg & Ben-Zeev, 2001, p. 330).

Die Unterschiede zwischen beiden kognitionspsychologischen Perspektiven sind insbesondere für den Bedeutungsumfang des Kompetenzverständnisses von zentraler Bedeutung. In der Berufs- und Wirtschaftspädagogik schlägt sich die These der Situativität von Lernen und Leisten vor allem in der Domänendiskussion nieder (vgl. u. a. Straka & Macke, 2009). Es wird zu zeigen sein, dass der Umfang des Domänenverständnisses zentralen Einfluss auf die Definition von Kompetenz und in Folge auch auf die Messung von Kompetenz hat (vgl. dazu Kapitel 3.2).

2.2.1 Referenzrahmen zur Analyse der kognitiven Struktur von Kompetenzen

Im Folgenden soll kein umfassender Einblick in die maßgeblich von Anderson (1983; 1993) entwickelte ACT-Theorie (Adaptive Character of Thought) gegeben werden. Das Ziel ist vielmehr, an diesem konkreten Beispiel aufzuzeigen, welche Eigenschaften Kompetenzen auszeichnen und wie sich diese begründen lassen. Die Auswahl der ACT-Theorie als Referenzrahmen ist vorrangig durch zwei Argumente gestützt:

- Die ACT-Theorie gibt zum einen Hinweise darauf, wie sich Routinen herausbilden. In der beruflichen Bildung sind Arbeitsroutinen als Teil der Kompetenzdiskussion ein zentraler Bestandteil. Anderson beschreibt, wie Routinen zunehmend ohne Aktivierung des deklarativen Wissens genutzt werden können (u. a. Anderson, 1994; hierzu auch Reif, 2008). Von besonderem Interesse ist, wie sich im Zusammenspiel von deklarativem und pro-

zeduralem Wissen diese prozeduralen Fähigkeiten entwickeln, die als Handlungskomponenten eine berufliche Anforderungssituation zu charakterisieren helfen.

- Zum anderen haben kognitive Architekturvorstellungen den zentralen Vorteil, dass sie für eine psychologische Theoriebildung begründete Einschränkungen der Modellierungsfreiheitsgrade liefern. Sie bieten damit einen einheitlichen konzeptuellen und technischen Rahmen, in den sich unterschiedliche Anwendungen einpassen lassen; damit wird eine Vergleichbarkeit verschiedener Modelle erst ermöglicht (vgl. u. a. Opwis, 2006, S. 729).

Kognitive Architektur und Bereichsspezifität

Kognitive Architekturmodelle, wie beispielsweise das ACT-Modell von Anderson, spezifizieren Dispositionen als überdauernde Eigenschaften des menschlichen kognitiven Systems. Ursprünglich ging Anderson (1983) von einem deklarativen Langzeitgedächtnis für Basiswissen aus, dessen Einzelfakten (= *chunks*) in einem semantischen Netz miteinander verbunden sind. Parallel dazu postuliert er ein zweites, prozedurales Langzeitgedächtnis, das Produktionsregeln der deklarativen Wissensanwendung enthält (= *productions*) und sich als Funktion der Anwendungshäufigkeit darstellen lässt. Jedes Element im Langzeitgedächtnis besitzt nach dieser Argumentation eine eigene Stärke, die als Funktion der Häufigkeit des Gebrauchs variiert (hierzu auch Fersch, 2006).

„There are three essential theoretical commitments one makes in the ACT knowledge representation. One is that there are two long-term repositories of knowledge: a declarative memory and a procedural memory. The second is that the *chunk* is the basic unit of knowledge in declarative memory. The third is that the *production* is the basic unit of knowledge in procedural memory“ (Anderson, 1993, p. 17).

Gerade die Beziehung zwischen deklarativem und prozeduralem Wissen wurde im Laufe der Zeit neu strukturiert und in der Weiterentwicklung der Theorie (ACT-R; Adaptive Character of Thought - Rational) festgeschrieben (Anderson, 1993). Insgesamt sind zwei Veränderungen bemerkenswert: Die erste betrifft die Frage, wie sich prozedurales Wissen aus dem deklarativen Gedächtnis entwickelt. Ursprünglich lag die Betonung darauf, wie deklarative Wissensbestände vermittelt werden, um prozedurales Wissen zu stimulieren. In der Neuordnung der Theorie stehen hier jedoch (Anwendungs-)Beispiele im Fokus, die aufzeigen, wie Prozeduren konkret auszuführen wären, und die darauf bezogenen deklarativen Repräsentationen.

„Originally, the emphasis had been on declarative memory for instruction, but now the emphasis is on declarative memory for examples of how the procedures should be executed“ (Anderson, 1994, p. 1322).

Der Erwerb prozeduraler Fähigkeiten hat sich damit von einem Lernen durch Analogien zu einem Lernen durch (situationsspezifische) Beispiele geändert. Anderson betont, dass dies nicht zwangsläufig, aber in Einzelfällen doch mit einer Veränderung der deklarativen Wissensbestände einhergeht/einhergehen muss. Auslöser dieser Sichtweise, die auch einen Paradigmenwechsel vom input- zum outputgesteuerten Wissenserwerb einschließt, waren Forschungen zu allgemeinen Problemlösestrategien und dem Transfer von Wissen (u. a. Chi, Bassok, Lewis, Reimann & Glaser, 1989; Collins, Brown & Newman, 1989; Lave & Wenger, 1990). Nach Thorndike (1906) ist der Transfer von einer Situation auf eine andere als Funktion davon zu verstehen, wie Wissen repräsentiert ist (zit. n. Singley & Anderson, 1989). Übertragen auf die ACT-R-Theorie bedeutet dies, dass Transfer von Wissen von einer Situation auf eine andere nichts anderes ist als eine Funktion des als Repräsentation zur Verfügung stehenden Wissens. Dabei ist ein wesentliches Transfermerkmal darin zu sehen, wie gut die korrespondierenden Prozeduren aufeinander abgestimmt sind.

Übertragen auf die verschiedenen Bedeutungsumfänge von Kompetenz lässt sich aus diesen theoretischen Annahmen Folgendes schließen (*Merksatz I*):

Merksatz I

Je bereichsunspezifischer eine Kompetenz ist (beispielsweise Schlüsselkompetenzen), desto einfacher lässt sie sich in wechselnden Situationen generalisieren. Dabei lassen sich die Bereichsunspezifität respektive die Bereichsspezifität nach der ACT-R-Theorie dadurch bestimmen, in welchem Umfang die die Kompetenz definierenden deklarativen und prozeduralen Wissensbestände Ähnlichkeiten aufweisen. Oder anders formuliert: Eine Kompetenz ist umso bereichsunspezifischer, je stärker sie sich in den zu betrachtenden Situationen über gleiche bzw. ähnliche *chunks* als Untereinheiten des deklarativen Wissens bzw. über gleiche bzw. ähnliche *productions* als Untereinheiten des prozeduralen Wissens definieren lässt.

Ein anschauliches Beispiel für eine nach der ACT-R-Theorie als bereichsunspezifisch einzuordnende Schlüsselkompetenz wäre Kreativität. Interessanterweise konnte die Forschung zahlreiche Einflussfaktoren auf Prozesskomponenten menschlicher Kreativität identifizieren; der situationale Kontext zählt allerdings nicht zu ihnen. Im Hinblick auf die kognitiven Faktoren zeigen sich zwar die Art und Weise, wie zuvor bei einer anderen Aufgabe gedacht und gearbeitet wurde, als prädiktiv. Jedoch beziehen sich die hier angesprochenen situativen Aspekte nicht auf die in der ACT-R-Theorie gewählte Form von Transfer, der sich auf den Erwerb von Fähigkeiten abstellt und zunächst einmal nicht danach fragt, in welchem Umfang in der Vergangenheit gemachte Erfahrungen auf aktuelle Situationen wirken (Anderson, 1993, p. 183). Stärker als kognitive Faktoren ist Kreativität durch affektive und motivationale Faktoren beeinflusst (Sternberg,

1988; Isen, 2000). Wird der ACT-R-Theorie gefolgt, wäre das Entstehen einer kreativen Lösung ein bereichsunspezifischer Prozess, da ein Vergleich der *chunks* und *productions* im hohen Maße unabhängig von der gegebenen Situation starke Ähnlichkeiten aufwiese; der Kreativitätsprozess ließe sich folglich bereichsübergreifend über idealtypische Abläufe abbilden. Auf der Stufe der Vorbereitung wären neben dem Faktenwissen das Wissen um kreativitätsrelevante Denkweisen und Heuristiken für die Problemlösung entscheidend (Amabile, 1996). Denkweisen und Heuristiken gelten gemeinhin als generalisierbar; eine Auflösung der Kompetenz in *chunks* und *productions* würde in verschiedenen Situationen vergleichbare Lösungen erbringen (vgl. u. a. Singley & Anderson, 1989). Die Stufe der Inkubation bewertet Kreativität als stochastisches Modell des Zufalls, der von der Breite des Wissens begünstigt wird (Simonton, 2003). Auch hier ließe sich die Relation von *chunks* und *productions* als Ähnlichkeitsmaß unabhängig von der spezifischen Situation nutzen. Auf den Stufen der Bewertung und Ausarbeitung des Problemlöseprozesses kommen wiederum allgemeine Strategien und Heuristiken, also Strukturierungen des Suchprozesses, anstelle bereichsabhängiger Algorithmen zum Tragen. Der Reiz, möglichst bereichsunspezifische Kompetenzen zum Ziel der Erfassung zu machen, liegt in ihrer Generalisierbarkeit. Ließe sich eine Kompetenz über ihre kognitive Architektur theoretisch als bereichsunspezifisch definieren, wären Induktionen möglich; Erkenntnisse, die in speziellen Szenarios gewonnen wurden, ließen sich theoretisch begründet auf andere Szenarien übertragen. Je breiter allerdings ein Bereich oder eine Situation für eine Kompetenz definiert wird, umso kritischer wird die Operationalisierung.

Kontextualisierung und Erlernbarkeit: Knowledge Deployment

Die zweite zentrale Änderung hinsichtlich des Übergangs vom deklarativen zum prozeduralen Wissen bezieht sich auf den ursprünglich angenommenen Langzeitcharakter des deklarativen Gedächtnisses. Zur Erfüllung von spezifischen Aufgaben ist jedoch weniger eine permanente konkrete Wissensspeicherung notwendig als vielmehr ein aktives Zugreifen auf notwendige Wissensbestände. Empirische Befunde zeigen, dass Prozeduren ausgeführt werden können, ohne dass eine entsprechende deklarative Repräsentation des Wissens vorliegen muss (vgl. dazu Anderson, 1994). Als Konsequenz wird folglich nicht mehr von prozeduralem Wissen, sondern von prozeduralen Fähigkeiten (*procedural skills*, vgl. u. a. Anderson, 1996) gesprochen. Erst diese Erweiterung der Argumentation innerhalb der ACT-R-Theorie macht den Erwerb von Routinen plausibel, da nicht für jede Prozedur eine korrespondierende Wissensrepräsentation vorliegen muss; Routinen, wie sie u. a. in Arbeitsprozessen erworben werden, können sich im Gegenteil auch ohne aktiven Rückgriff auf deklarative Wissensbestände zeigen. Diese Annahmen stimmen auch mit den Ergebnissen der Cognitive Load Theory überein, die aufzeigen, dass kognitive Auslastung insbesondere aus Prozessen der Schemakonstruktion resultiert (deklarative Repräsentation), eine erfolgreiche Schemaanwendung (Prozeduralisierung) die kognitiven Prozesse hingegen erleichtert (Sweller, 1988; Sweller, van Merriënboer & Paas, 1998). Die

hier angesprochenen Fragen sind: Wie lassen sich prozedurale Fähigkeiten effektivieren (= Routinen bilden) und wie lässt sich korrespondierendes relevantes Wissen schnell identifizieren, um das kognitive System zu entlasten? Anderson bezeichnet diesen Aspekt der ACT-R-Theorie als „Knowledge Deployment“ (Anderson, 1996, p. 360). Eine zentrale Idee der ACT-R-Theorie ist, dass Wissensstrukturen ein Ergebnis der Enkodierung einer spezifischen Situation sind; Kompetenzen sind folglich nichts anderes als eine Auflistung von einzelnen Wissenskomponenten, die sich aus der Anforderungssituation ableiten lassen:

„The theory implies that acquiring competence is very much labor-intensive business in which one must acquire one-by-one all the knowledge components. [...] However, it would be misrepresenting the matter to say that competence is just a matter of getting all the knowledge units right. There is the very serious matter of deploying the right units at the right time, [...]“ (Anderson, 1996, pp. 359-360).

Die ACT-R-Theorie greift hierfür auf einen zweistufigen, parallel ablaufenden Prozess zurück: Zum einen sind jene Wissensstrukturen (chunks und productions) zu identifizieren, die charakteristisch für die Anforderungssituation und von besonderer Bedeutung für den konkreten Kontext sind. Über diese Wissensstrukturen wird dann in einem zweiten Schritt die (beobachtbare) Leistung in der Anforderungssituation definiert. Dies bedeutet letztlich, dass Anforderungen dann bewältigt werden können, wenn das zur Lösung relevante Wissen (deklaratives Wissen und/oder prozedurale Fähigkeit) in eine beobachtbare Leistung transformiert werden kann. Oder anders formuliert: Lösungen geben das Enkodierungsmuster für eine beobachtbare Leistung vor, wobei das für eine Leistung relevante Wissen zunächst erst einmal aktiv verfügbar sein muss. Die Verfügbarkeit des Wissens ist dabei abhängig von der (a posteriori) Wahrscheinlichkeit, mit der es in der Situation eingesetzt und/oder gebraucht wird. Zur Abschätzung der Wahrscheinlichkeiten greift das ACT-R-Modell auf Bayes-Inferenzen zurück (Anderson, 1996, p. 360):

$$\frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(H)}{P(\bar{H})} * \frac{P(E|H)}{P(E|\bar{H})}$$

Transformiert in Logarithmen

$\text{Log}(\text{posterior odds}) = \text{Log}(\text{prior odds}) + \text{Log}(\text{likelihood ratio})$

$\text{Log}(P(H|E)) = \text{Log}(P(H)) + \text{Log}(\text{likelihood ratio})$

mit: $P(H|E)$ = a posteriori Wahrscheinlichkeit für ein Ereignis H unter der Bedingung E und $P(H)$ = a priori Wahrscheinlichkeit für ein Ereignis H

Über das Bayes-Theorem gelingt es im Rahmen der ACT-R-Theorie, die logarithmierte a posteriori-Wahrscheinlichkeit zu bestimmen, mit der Wissen in einem spezifischen Kontext aktiviert wird („activation-level“). Diese Wahrscheinlichkeit entspricht der logarithmierten Wahrscheinlichkeit, dass das Wissen bereits in der Vergangenheit in einer vergleichbaren Situation zur Anwendung gebracht wurde („base-level“) plus einem Schätzer, dass es auch in zukünftigen vergleichbaren Kontexten zur Erbringung der Leistung hilfreich sein wird („contextual-priming“). Es wird folglich unterstellt, dass für die Lösung von spezifischen Anforderungssituationen ein Abgleich bisher zu bewältigender Anforderungen erbracht wird und dieser Rückschlüsse darüber anregt, welches Wissen in der aktuellen Situation zu aktivieren ist. Die in vergangenen Situationen gemachten Erfahrungen lassen sich als Lerneffekt interpretieren. Anderson bringt diese Inferenz in die folgende Gleichung (Anderson, 1996, p. 360)

$$\text{Activation-Level} = \text{Base-Level} + \text{Contextual-Priming}$$

Bei der Ausbildung von Routinen gilt, dass sich der Bias (likelihood ratio) dem Wert 1 annähert. Da bei der Ausführung von Routinen keine Variation des Kontexts angenommen wird, geht die kontextuale Färbung der Anforderungssituation (contextual-priming) folglich gegen 0. Im Zentrum steht jene Wahrscheinlichkeit, die Auskunft darüber gibt, in welchem Maße relevante Wissensstrukturen in vertrauten Anwendungsräumen aktiviert werden. Diese Wahrscheinlichkeit entspricht bei einer Routine in hohem Maße der a priori-Wahrscheinlichkeit. Bei einer Routine genügt zur Enkodierung des Kontextes die Aktivierung der prozeduralen Fähigkeit. Ein Rückgriff auf die korrespondierende deklarative Repräsentation ist damit nicht mehr notwendig (vgl. auch Reif, 2008). Werden die oben ausgeführten Annahmen vor dem Hintergrund eines Problemlöseprozesses geprüft, ergeben sich folgende Argumentationen: Anders als bei der Ausbildung von Routinen treten in einem Problemlöseprozess Variationen des Kontextes auf, die sich beispielsweise darin zeigen, dass es unterschiedliche Strategien der Zielerreichung gibt. Als anschauliches Beispiel sei hier auf Luchins' „waterjug problem“ (Luchins, 1942) und dessen Weiterentwicklungen verwiesen. Beim Wassermüllproblem sollen Testpersonen eine festgelegte Menge Wasser abfüllen. Ihnen stehen drei Krüge mit Wasser zur Verfügung, mit denen sie durch die geschickte Kombination aus Addition und Subtraktion den als Ziel vorgegebenen Wasserstand nachbilden sollen. Die Ergebnisse Luchins' zeigen, dass eine einmal gefundene, als erfolgreich und generalisierbar bewertete Lösungsstrategie auch dann nicht aufgegeben wird, wenn einfachere Lösungen möglich und teilweise offenkundig sind.

In einem ähnlichen Versuch von Lovett (1994, zit. n. Anderson, 1996, p. 363), in dem Testpersonen aus Bausteinen eine festgelegte Länge eines vorgegebenen Zielbausteins nachbilden sollen, werden diese Ergebnisse bestätigt. Lovett erwei-

terte ihr Experiment, indem sie zwei Lösungsstrategien vorgibt: Die „Overshoot-Strategie“ geht davon aus, dass der Ausgangsbaustein größer als der Zielbaustein ist und demnach kombinatorisch verkleinert werden muss; die „Undershoot-Strategie“ startet hingegen mit einem kleineren Ausgangsbaustein, der auf die Größe des Zielbausteins zu bringen ist. Zusätzlich wird den Testpersonen die Information gegeben, welche der beiden Strategien in vergangenen Tests die erfolgreichere war. Es zeigten sich folgende Ergebnisse: (1) Die Testpersonen zeigten eine starke Tendenz, die (scheinbar) erfolgreichere Strategie zu wählen, auch wenn eine andere Herangehensweise einfacher gewesen wäre. (2) Es konnte gezeigt werden, dass die Wahrscheinlichkeit, eine der beiden Strategien auszuwählen, sich aus der Kombination des (als angenommen geltenden) Erfolgs und einem Schätzer, der sich als Abweichung vom Erfolg beschreiben lässt, ergibt. Die ACT-R-Theorie schreibt dieses Ergebnis wie folgt (Anderson, 1996, p. 364): Die logarithmierte Wahrscheinlichkeit, dass die gewählte Strategie erfolgreich sein wird, ergibt sich aus der Erfolgswahrscheinlichkeit der Strategie in der Vergangenheit und einem variablen Fehlerwert, welcher ein Schätzer der Abweichung vom Ideal ist. Interessant ist, dass sich die a priori-Wahrscheinlichkeit als Funktion der von der Testperson während des Tests wahrgenommenen (Erfolgs-)Erfahrungen und der von anderen Testpersonen gemachten (angenommenen) Erfahrung darstellt (suggestive Erfolgsgewissheit).

$$\text{Log(odds operator)} = \text{Log(prior odds)} + \text{context appropriateness}$$

Anders als bei der Anwendung von Routinen kommt der Likelihood Ratio bei der Problemlösung eine entscheidende Bedeutung zu. Der Quotient enthält Annahmen über Quantität und Qualität der Abweichung von der Ideallösung, die sich nicht immer wie in den oben beschriebenen Experimenten über Milliliter oder Zentimeter ausdrücken lassen und stark von der gegebenen Anforderungssituation abhängen. Es kann unter Bezugnahme auf die ACT-R geschlussfolgert werden, dass (1) der contextual-priming-Parameter (likelihood ratio) an den Vertrautheitsgrad der Situation gekoppelt ist und (2) die a posteriori-Wahrscheinlichkeiten des aktiven Wissenszugriffs stark erfahrungsbasiert sind und sich folglich auf Lerneffekte zurückführen lassen. Übertragen auf die verschiedenen Bedeutungsumfänge von Kompetenz lässt sich aus diesen theoretischen Annahmen Folgendes schließen (*Merksatz II*):

Merksatz II

Kompetenzen sind (1) im Gegensatz zu anderen Leistungsdispositionen (z. B. Intelligenz) stark an spezifische Kontexte gebunden, die das Enkodieren der beobachtbaren Performanz stützen (contextual-priming-Effekt). Kontextspezifi-

fische Kompetenzen setzen Erfahrungen mit den jeweiligen Kontexten voraus; sie lassen sich folglich (2) als erlernbar charakterisieren (prior odds-Effekt).

Zusammenfassend bietet die ACT-Theorie und insbesondere deren Weiterentwicklung, die ACT-R, über allgemeine kognitionstheoretische Annahmen die Möglichkeit, den Bedeutungsumfang von Kompetenzen einzugrenzen und damit die Definition von Kompetenz zu präzisieren. Als gültige Kriterien für Kompetenz lassen sich (1) die Kontextualisierung und (2) die Erlernbarkeit identifizieren. Über die Analyse der kognitiven Architektur ergibt sich ferner, dass Kompetenzen Kombinationen aus Wissenskomponenten sind. Diese Wissenskomponenten – deklarative Repräsentationen und prozedurale Fähigkeiten – werden jedoch erst durch einen aktiven Zugriff als Performanz in einer spezifischen Situation beobachtbar. Es lässt sich darüber hinaus festhalten, dass der Umfang an Bereichsspezifität einer Kompetenz sich dadurch charakterisieren lässt, wie verschieden die zugrundeliegenden Wissensstrukturen sich in spezifischen Kontexten oder Situationen zeigen.

2.2.2 Referenzrahmen zur Analyse der Reichweite von Kompetenzen

Die Strukturierung von Kompetenzen richtet sich nach den zugrunde liegenden Anforderungen und nach den kognitiven Prozessen, die zur Bewältigung dieser Anforderungen notwendig sind (vgl. u. a. Prenzel, Walter & Frey, 2007). Dieses sind Kriterien der Reichweite von Kompetenzen, die insbesondere dann eine zunehmende Bedeutung erfahren, wenn die Definition von Kompetenzen um funktionale oder soziale Aspekte erweitert wird (vgl. u. a. Winterton, Delamare-Le Deist & Stringfellow, 2006; Business Council of Australia, 2002). In diesen Erweiterungen des kognitiven Kompetenzbegriffs zeichnet sich ein Dilemma gegenwärtiger Kompetenzdefinitionen deutlich ab: Die kontextspezifischen Leistungsdispositionen und die die Leistung bedingenden funktionalen und personalen Handlungsfähigkeiten werden übergreifend isoliert betrachtet und selten aufeinander bezogen. Der kognitive ‚Wert‘ der Kompetenz ist vor dem Hintergrund adäquater Handlungen in einer spezifischen Anforderungssituation jedoch nur mit Hilfe relevanter Vermittlungsleistungen wirklich aufzuschlüsseln: Nicht jede Anforderungssituation fordert alle Ressourcen eines Lernenden (vgl. die Herausbildung von Routinen, Kap. 2.2.1). Vor diesem Hintergrund wurden und werden im Rahmen bekannter Schulleistungstudien (IGLU, PISA, DESI) zusätzlich bereichsübergreifende Kompetenzen (u. a. Problemlösekompetenz; Prenzel, Baumert, Blum, Lehmann, Leutner, Neubrand, Pekrun, Rost & Schiefele, 2005) und fächerübergreifende Faktoren (u. a. Selbstkonzept, Motivation; Klieme, Neubrand & Lüdtke, 2001, S. 184) erhoben. Greeno, Riley und Gelman (1984; modifiziert: Gelman & Greeno, 1989) greifen in ihrem entwicklungspsychologischen Kompetenzmodell mit der Unterscheidung von *conceptual*, *procedural* und *utilizational competence*

- die Beziehung zwischen den fachlich spezifischen und übergreifenden kognitiven Komponenten sowie den handlungsregulativen Faktoren zur Planung und Ausführung kompetenter Handlungen in einer Anforderungssituation auf und betonen in besonderer Weise die Notwendigkeit der regulativen Auseinandersetzung mit der Anforderungssituation für das Erbringen der Leistung (Gelman & Greeno, 1989, p. 136);
- die „Nature of Competence“ wird von den Autoren in instruktionstheoretische Ansätze eingebettet, die zur Identifizierung kompetenten situationspezifischen Handelns auf kognitionstheoretische Überlegungen zur Systematisierung von Wissensbeständen und Lerneingangsvariablen und auf solche zum Prozess des Lernens zurückgreifen (u. a. Glaser, 1976; Anderson & Krathwohl, 2001).

Insbesondere diese zwei Leistungsmerkmale begründen die Auswahl der „Nature of Competence“ als Referenzrahmen. Mit dessen Hilfe soll im Folgenden die Reichweite von Kompetenzen in konkreten Anforderungssituationen präzisiert werden. Die Autoren betonen, dass eine Aufgliederung von Kompetenz in verschiedene Teilkomponenten nachhaltige Implikationen auf die Interpretation der gezeigten Performanz und die Entwicklung von Kompetenz haben dürfte (Gelman & Greeno, 1989, p. 128). Auch wenn die Ausführungen auf die Entwicklung mathematischer Kompetenz bezogen sind, ergeben sich m. E. doch zahlreiche Generalisierungsannahmen für den Bereich der beruflichen Bildung und damit für einen in der beruflichen Bildung verankerten Kompetenzbegriff.

Conceptual Competence

Conceptual competence wird als regelbasiertes, abstraktes Wissen in einer Domäne (zum Domänenbegriff vgl. Kap. 3.2) charakterisiert, das in einen spezifischen Handlungsplan umgesetzt und in der spezifischen Handlung wirksam wird. Zentrales Element dieser Kompetenzkomponente sind erlernte Regeln, die sich entsprechend kognitivistischer Annahmen in Abhängigkeit vom Vorwissen herausbilden:

„The common point in all these accounts is that the perceived order in a stimulus is often constructed on the basis of available structures [...] In sum, two issues need to be considered in treatments of competence. The first concerns the need to distinguish between competence that is required for the domain and a competent plan of action. [...] Second, [...] some clues as to how to account for self-initiated and self-guided learning of further knowledge within a domain“ (Gelman & Greeno, 1989, p. 131).

In der Konsequenz bedeutet dies, dass zwischen der Kompetenz, gegebene Anforderungssituationen konzeptual zu erfassen, und der Kompetenz, in einer spezifischen Situation adäquat handeln zu können, unterschieden werden muss. Dies entspricht zunächst der Argumentation Andersons, der ebenfalls davon ausgeht,

dass ein aktiver Zugriff auf situationsspezifisches Wissen die Grundlage für eine erfolgreiche Handlung ist (vgl. Kap. 2.2.1). Die Annahmen Greenos und Kollegen gehen jedoch darüber hinaus, indem sie zwischen operationalen und interpretativen kognitiven Strukturen unterscheiden: operationale kognitive Strukturen als Voraussetzung für kognitive Aktivität in einer spezifischen Situation und interpretative kognitive Strukturen als Grundlage dafür, entsprechende Settings bzw. Anforderungssituationen zu erfassen (Gelman & Greeno, 1989, p. 136). Darüber hinaus differenzieren Greeno et al. (1984) zwischen domänenspezifischer und domänenverbundener (domain-linked) Kompetenz, wobei die Generalisierung in der Anwendung das relevante Unterscheidungskriterium ist. Domänenspezifische Kompetenz bezieht sich auf die Bewältigung von Anforderungen in der Domäne, während domänenverbundene Kompetenz die Bewältigung von einzelnen Anforderungen in der Domäne unterstützen kann/soll.

Der von Greeno und Kollegen verwendete Begriff „PICK-UP-SCHEMES“ für die Klasse der domänenverbundenen Kompetenzen in Abgrenzung zu den domänenspezifischen Kompetenzen lässt sich in diesem Zusammenhang wie folgt interpretieren: Pick-up-Schemata charakterisieren verfügbare Repräsentationen und/oder Verhaltensmuster, die stark ausführungsbezogen sind, um die Bewältigung der gegebenen Anforderungssituation zu unterstützen. Für die Konstruktion von Anforderungssituationen, in denen Kompetenz erfasst werden soll, hat diese Differenzierung insbesondere mit Blick auf die Reichweite des Kompetenzbegriffs Konsequenzen. Es ist (1) nach spezifischen kognitiven Strukturen zu fragen, die in einer gegebenen Anforderungssituation als aktives Zugriffswissen verfügbar sein müssen (vgl. Anderson, 1996), und (2) sind unterstützende Verhaltensmuster und Repräsentationen zu ermitteln, die die Bewältigung der Anforderungssituation begünstigen. Von besonderem Interesse ist hier die Lokalisierung des Übergangs von domänenverbundenen allgemeinen zu domänenspezifischen kognitiven Fähigkeiten als Repräsentation einer spezifischen Anforderungssituation. Übertragen auf die Reichweite des Kompetenzbegriffs ist aus diesen theoretischen Annahmen Folgendes abzuleiten (*Merksatz III*):

Merksatz III

Die Reichweite des Kompetenzbegriffs ist im Unterschied zur Kompetenz-Performanz-Ableitung Chomskys nicht auf eine Gegenüberstellung und die damit verbundene Unterscheidung von Leistung und Kompetenz begrenzt, sondern wird um handlungsbasierte Komponenten erweitert, die sich als das Ergebnis des operationalen und interpretativen wissensbasierten Zugriffs in spezifischen Anforderungssituationen beschreiben lassen (vgl. Gelman & Greeno, 1989, p. 172). Dies bedeutet, dass für kompetente Handlungen sowohl domänenspezifische als auch domänenverbundene Repräsentationen von Fakten, Prozeduren, Prinzipien und Theorien verfügbar sein müssen.

Insbesondere die Differenzierung zwischen domänenverbundenen und domänenspezifischen Kompetenzen ist der für den Bereich der Berufs- und Wirtschaftspädagogik entscheidende theoretische Schritt, wenn Übergänge zwischen fachspezifischen berufsbezogenen Fähigkeiten und allgemeinen bereichsübergreifenden kognitiven Dispositionen untersucht werden sollen. Konzeptuale Kompetenz ist in diesem Zusammenhang weniger auf übergreifende kognitive Fähigkeiten und Dispositionen zurückführbar, sondern auf mentale Netzwerke domänenspezifischen Wissens sowie auf domänenspezifische Techniken und Arbeitsroutinen, welche domänenverbunden ausgebildet werden (Gelman & Greeno, 1989, p. 144; hierzu auch Reif, 2008, p. 261).

Procedural und Utilizational Competence

Neben der *conceptual competence* werden zwei weitere Kompetenzkomponenten aufgenommen (Greeno & Gelman, 1984): (1) *Procedural competence* bezieht sich auf die Prozeduren und Fertigkeiten bei der Anwendung des konzeptualen Wissens in konkreten Anforderungssituationen und ist damit in der Tradition von Anderson (1983, 1993) zu interpretieren; (2) *utilizational competence* umfasst die Fähigkeit, Anforderungssituationen regulativ zu erfassen, d. h. Anforderungsmerkmale und individuelle Zielmerkmale einzuschätzen und aufeinander zu beziehen. In dieser Kompetenzkomponente wird folglich die Interaktion mit dem Kontext erfasst oder – umfassender im Sinne situierter Kognition – die Konstruktion der Situation als Bewertungsprozess. In der modifizierten Modellkonzeption von Gelman und Greeno (1989, p. 144) wird dieser Kompetenzbereich als *interpretative competence* bezeichnet, da unter diesen die Kulturtechniken zum Verstehen und Interpretieren konkreter Anforderungssituationen (u. a. sprachliches und mathematisches Verständnis) sowie – als Erweiterung des kognitiven Ansatzes – die Regulationsmechanismen zum Handeln in konkreten Situationen subsumiert werden können.

„These interpretative processes deal with the interactive aspects of the setting. Greeno et al. (1984) discussed children's use of features of the task setting to achieve goals and called the general knowledge for that utilization competence. That discussion was concerned with children's assessments of such domain-relevant props as the arrangement of items in a line or a circle that provide resources for achieving goals that have been activated in the child's understanding of the task“ (Gelman & Greeno, 1989, p. 143).

Forschungsergebnisse belegen, dass die Auseinandersetzung mit einer spezifischen Situation/Domäne umso erfolgreicher ist, je spezialisierter das Wissen über die Domäne ausgeprägt ist (vgl. Chase & Simon, 1973; Chi, 1978; Gobet & Simon, 2000). Hierbei sind die beobachteten Effekte in der Domäne und nicht in persönlichen Dispositionen begründet. Die Experten-Novizen-Forschung belegt u. a., dass Experten nicht generell intelligenter oder befähigter sind als Novizen; Ex-

perten können jedoch die Anforderungssituation besser erfassen sowie leichter kausale Bezüge herstellen (u. a. Chi, Hutchinson & Robin, 1989; Hirschfeld & Gelman, 1994). Der Ansatz von Gelman und Greeno (1989) entspricht in hohem Maße der Auffassung, das Kompetenzkonzept mit Hilfe einer differenzierten Betrachtung der Wissensdimension zu erfassen (vgl. u. a. Achtenhagen, 2005): Experten verfügen im deklarativen Bereich über eine umfangreichere Wissensbasis sowie über einen höheren Anteil richtiger Verknüpfungen von Wissen; im prozeduralen Bereich greifen Experten beim Abbilden der Situation auf weniger Oberbegriffe und abstraktere Merkmale zurück als Novizen, da sie über effiziente, bereichsspezifische Problemlöseschemata verfügen, die den Suchaufwand reduzieren und eine sichere Lösung ohne kognitive Überlastung erlauben; im interpretativen Bereich ist die Konstruktion der Situation durch verfügbare kognitive Strukturen in sehr kurzer Zeit möglich (Gelman & Greeno, 1989; Perkins, 1992; Anderson, 1993; Spada & Wichmann, 1996). Die breitere konzeptuale Struktur wirkt damit nachweisbar auf die interpretative Struktur. Neuere Befunde zeigen zudem, dass konzeptuale und interpretative Strukturen sich nicht nur in Domänen unterscheiden, sondern auch bereits auf der Ebene der einzelnen Aufgaben variieren können (Gelman, 1996; Linhares & Brum, 2007). Auch die Lehr-Lernforschung zeigt deutlich, dass die Lernleistung maßgeblich von der Bewertung der Situation und deren Repräsentation bestimmt ist (u. a. Winther, 2006; Ames, 1992; Pintrich & De Groot, 1990; Pressley, Borkowski & Schneider, 1987). Für die Reichweite des Kompetenzbegriffs ergeben sich damit folgende Schlussfolgerungen: Menschliche Handlungen werden in konkreten Situationen durch den Kontext geleitet und determiniert, wobei die Konstruktion der Situation durch die Stimuli der Umwelt erfolgt. Für die Kompetenz bedeutet dies, dass Handlungsangebote und Handlungseinschränkungen einer spezifischen Situation immer mit in das Verständnis von Kompetenz eingehen müssen.

„A concept of attunements to constraints and affordances provides an alternative, more general form of explanation of activity“ (Greeno, 1998, p. 8).

Die Einbeziehung von Handlungsangeboten und -einschränkungen als Kriterien der Reichweite des Kompetenzbegriffs lässt sich auf drei Problembereiche der Kognitionsforschung zurückführen (Greeno, 1998, pp. 7/8):

- Der Aktionsraum einer Anforderungssituation lässt sich selten exakt definieren; bedeutende Aspekte der Lösung ergeben sich häufig erst im Lösungsprozess selbst (*emergent problem spaces*).
- Die Konstruktion von Wissen und Verstehen ist selten starr auf zuvor festgelegte Repräsentationen zurückzuführen, sondern entwickelt sich dynamisch und in Interaktion mit der Umwelt (*interactive constructing of understanding*).
- Der Grad der Auseinandersetzung mit einer konkreten Anforderungssituation ist durch motivationale und volitionale Vermittlungsleistungen beein-

flusst. Zudem bestimmen die Interaktionen mit anderen deren Einstellungen und Verhaltensmuster – im Sinne einer *community of practice* –, in welchem Maße Handlungsräume genutzt werden (*engagement in activities*).

Diese theoretischen Annahmen lassen sich im Hinblick auf die Reichweite des Kompetenzbegriffes wie folgt zusammenfassen (*Merksatz IV*):

Merksatz IV

Die zu definierende Kompetenz kann (1) nur so umfangreich wie die zugrundeliegende Anforderungssituation bzw. ein Set aus Anforderungen sein und ist damit an die Handlungsspielräume des Kontextes gebunden; sie kann ferner (2) nicht als unveränderlich beschrieben werden, sondern ist kontextsensitiv von aktuellen Konstruktionsleistungen des Individuums abhängig.

Dass die Kompetenz der Lernenden stark von dem zugrundeliegenden Kontext und den in diesem Kontext angeregten kognitiven Stimuli sowie den motivationalen und volitionalen Vermittlungsleistungen abhängig ist, konnte in eigenen Arbeiten sehr deutlich gezeigt werden (u. a. Winther, 2006; 2007; Winther & Achtenhagen, 2008a). Die Befunde des DFG-Projekts „Integrierte Kompetenzförderung in den beruflichen Fächern des Wirtschaftsgymnasiums“ belegen, dass ein ausgeprägtes Verstehen von Begriffen und Konzepten den Grad der Auseinandersetzung mit einer konkreten Anforderungssituation zwar fördert, dass jedoch selbst bei einer ausreichenden Wissensbasis eine erfolgreiche Umsetzung des Wissens in konkrete Handlungen scheitern kann, wenn die dazu erforderlichen motivationalen und volitionalen Vermittlungsleistungen nicht bestehen (vgl. dazu auch Artelt, Demmrich & Baumert, 2001). Im Rahmen des Projekts konnten drei *kontextsensitive states* identifiziert werden, mit deren Hilfe sich Handlungsspielräume des Kontextes präzisieren ließen (Winther & Achtenhagen, 2008a, S. 272):

- Motivationale Regulation (Variationsbreite der Kontexte/intraindividuelle Unterschiede: $F = 1,62$ bis $11,15$; $p < ,05$; $\eta^2 = ,24$ bis $,60$) im Sinne aktualisierter Motivation der kognitiven Motivationsmodelle ist als Reflexionsprozess das Ergebnis kognitiver Bewertungen im Handlungsvollzug. Die kognitiven Bewertungen betreffen das Aufgabenwissen, die Wahrnehmung der Aufgabenvalenz und den Strategieeinsatz zur Aufgabenbewältigung. Die Reflexion über vorhandene Wissensbestände stärkt die Persistenz des Lernverhaltens, forciert die Auswahl und die Adaption von Lernstrategien und setzt Zielerreichungsprozesse in Gang (vgl. Pintrich, 2000; 2004). Individuelle Kompetenzunterschiede ließen sich vor dem Hintergrund dieser Überlegungen auf die Qualität der motivationalen Regulation zurückführen, die das Zusammenspiel von bereichsspezifischem Wissen, wahrgenommener Aufgabenvalenz und zielführendem Strategieeinsatz beschreibt.
- Volitionale Kontrolle (Variationsbreite der Kontexte/intraindividuelle Unterschiede: $F = 4,77$ bis $17,52$; $p < ,01$; $\eta^2 = ,31$ bis $,62$) regt Vermittlungs-

prozesse an, die die Verwirklichung der aktuellen Absicht begünstigen, indem die Steuerung der Ausführung der Absicht kontrolliert wird. Lernende regulieren Anstrengung und Ausdauer entsprechend dem Anforderungsniveau der Situation, um so die Ausführung der Absicht zu gewährleisten. (u. a. Kuhl, 1996).

- Kognitive Kontrolle (Variationsbreite der Kontexte/intraindividuelle Unterschiede: $F = 3,15$ bis $7,77$; $p < ,01$; $\eta^2 = ,42$ bis $,50$) ist im Lösungsprozess zum einen auf der Ebene der Repräsentation der Anforderungssituation und zum anderen hinsichtlich des metamemorialen Wissens über die Anforderungssituation relevant. Die Repräsentationsebene bestimmt das Abstraktionsniveau; die Qualität des metamemorialen Wissens beeinflusst eine realistische Situationseinschätzung (Baumert, Klieme, Neubrand, Prenzel, Schiefele, Schneider et al., 2000). Es lassen sich Aussagen hinsichtlich der subjektiven mentalen Repräsentation der Domäne treffen, indem die Anforderungssituation durch die Nutzung von zentralen, lösungsrelevanten Merkmalen klassifiziert wird (u. a. Chi, Bassok, Lewis, Reimann & Glaser, 1989).

Insgesamt zeigen die Ergebnisse deutlich, dass Kompetenzen nicht stabil sind, sondern sich kontextsensitiv ausbilden. Mit dem Projekt liegen Ergebnisse bezüglich der Selbstbewertungen der Lernenden vor, die verdeutlichen, dass die Kompetenz des Lernenden abhängig ist von der Antizipation des Kontextes und von seinen Fähigkeiten zur kognitiven, volitionalen und motivationalen Regulation im konkreten Handlungskontext ($\beta = ,18$ bis $,34$; $p < ,001$; $\chi^2 < 3,28$; $GFI > ,992$; $RMSEA < ,093$; vgl. Winther, 2006, S. 220). In Konsequenz bedeutet dies, dass Handlungsangebote und Handlungseinschränkungen einer spezifischen Situation tatsächlich in das Verständnis von Kompetenz mit eingehen. Die Reichweite des Kompetenzbegriffs ist damit äußerst begrenzt: Kompetenzen sollten sich auf spezifische Kontexte beziehen, deren Handlungsspielraum möglichst klar definiert werden kann. Die Präzisierung des Handlungsspielraums lässt sich beispielsweise über Schwellenwerte der zur Anforderungsbewältigung notwendigen kognitiven, motivationalen und volitionalen Ressourcen erreichen. Der enge Kontextbezug bedeutet nicht, dass nur domänenspezifische Kompetenzbereiche betrachtet werden, wohl aber, dass mögliche domänenverbundene Kompetenzbereiche zur Anforderungssituation passen müssen und sich folglich als Repräsentationen dieser konkreten Anforderungssituation darstellen lassen.

Nach Hartig (2008, S. 21) stellt der Bezug auf eine Menge hinreichend ähnlicher realer Situationen, in denen bestimmte, ähnliche Anforderungen bewältigt werden müssen, eine Möglichkeit dar, die Willkür der Kontextdefinition und damit der Bestimmung der Reichweite von Kompetenz einzuschränken. Die zentrale Herausforderung ist, gemeinsame ähnliche Anforderungen zu beschreiben. Hierfür ist eine hinreichend genaue Vorstellung darüber zu entwickeln, welche Prozesse beim Bearbeiten und Lösen der Anforderungssituationen ablaufen. Die Struktur von Kompetenz ergibt sich damit aus der Struktur der Anforderungssituation. Vor diesem Hintergrund sind neben der kognitiven Struktur und dem

Bedeutungsumfang von Kompetenzen für eine präzisierende Kompetenzdefinition in die Modellkriterien von Bedeutung, in die der Kompetenzbegriff einzubetten ist.

2.3 Kompetenzmodelle: Zusammenhang von kognitiver Struktur und Reichweite

Die Ausführungen zur Beschreibung der kognitiven Struktur und der Reichweite von Kompetenzen haben gezeigt, dass in konkreten Anforderungssituationen unterschiedliche Teilkomponenten von Kompetenz systematisch zusammenhängen und damit die Struktur von Kompetenz bestimmen. Kompetenzmodelle, in denen diese Zusammenhänge theoretisch aufgeschlüsselt und damit diagnostizierbar gemacht werden sollen, stellen die Grundlage für die Entwicklung valider Messinstrumente dar. Dabei ist die Wahl des theoretischen Modells abhängig von den angenommenen Zusammenhängen und Entwicklungstendenzen der zugrundeliegenden Kompetenz, dem angestrebten Bedeutungsumfang des Kompetenzbegriffs sowie der Zielsetzung der Kompetenzerfassung. Für die verschiedenen Bereiche der Kompetenzdiagnostik wurden folglich auch verschiedene Kompetenzmodelle entwickelt (vgl. u. a. Abs, 2007):

- Kompetenzmodelle, die sich mit der Dimensionalität von Kompetenzen befassen. Diese Modelle gehen von der theoretischen Annahme aus, dass in einer konkreten Anforderungssituation verschiedene Kompetenzen bzw. Kompetenzbündel differenziert erfasst werden können (*Kompetenzstrukturmodell*).
- Kompetenzmodelle, die die Graduierung von Kompetenzen und damit unterschiedliche Stufen der Kompetenzerreichung abbilden, fokussieren auf die Hierarchie von Kompetenzen und sind stark auf die Anforderungssituationen bezogen (*Kompetenzniveaumodell*).
- Kompetenzmodelle, die sich vorrangig auf der Basis kognitionstheoretischer Annahmen mit den mentalen Modellen von Experten und Novizen und deren Unterschieden beschäftigen (*Kompetenzentwicklungsmodell*).

Neben diesen drei zentralen Kompetenzmodellen haben sich in der Literatur weitere Modelle herausgebildet. Ihnen ist gemeinsam, dass sie Verhaltensoptionen und Bewältigungsstrategien in verschiedenen Anforderungssituationen auf ihre Generalisierbarkeit prüfen. Je mehr Handlungsoptionen zur Verfügung stehen bzw. je mehr Anforderungssituationen sich bewältigen lassen, umso stärker wäre in diesem Fall die Kompetenz ausgeprägt.

2.3.1 Struktur, Graduierung und Entwicklung

In der Expertise „Zur Entwicklung nationaler Bildungsstandards“ werden Kompetenzmodelle als Mittler zwischen abstrakten (curricularen) Bildungszielen und konkreten Aufgabenstellungen definiert (Klieme, Avenarius, Blum, Döbrich, Gruber et al., 2003, S. 71), deren Ziel es ist, zu einer theoretischen Beschreibung der Struktur spezifischer Fähigkeiten und der Stufen ihres Erwerbs beizutragen (ebd., S. 65). Im Folgenden sollen unterschiedliche Kompetenzmodelle kurz skiz-

ziert und an Beispielen aufgezeigt werden, welche Konsequenzen sich aus der Modellwahl für die Definition und die Erfassung von Kompetenzen ergeben.

Kompetenzstrukturmodelle

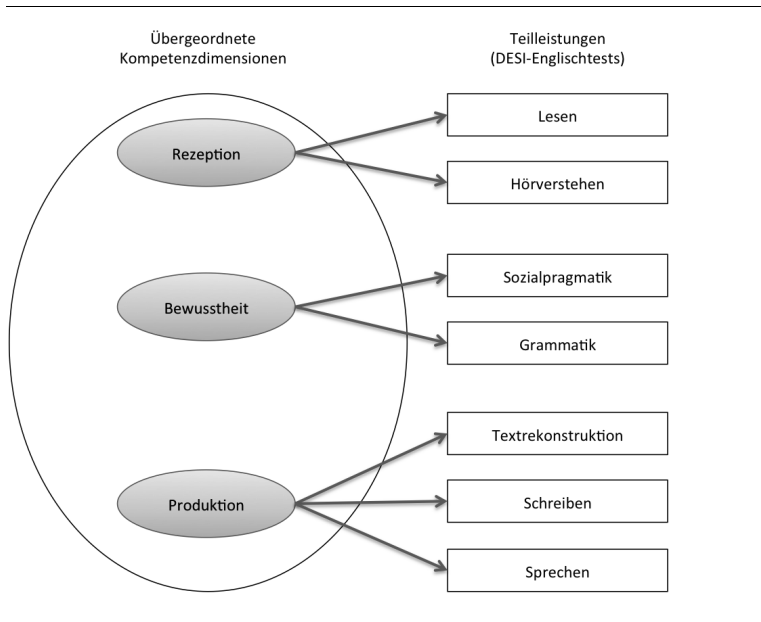
Kompetenzstrukturmodelle befassen sich einerseits mit der Identifizierung von Kompetenzbereichen. Hierunter ist die inhaltlich-systematische Ausgestaltung einer Kompetenzdomäne zu fassen. Im Kern geht es um übergreifende Ideen und Prinzipien einer Fachdisziplin (oder eines Unterrichtsfaches), die aus der jeweiligen Fachwissenschaft sowie der vorherrschenden Fachdidaktik abgeleitet werden (vgl. insbesondere die Arbeiten im Bereich der Mathematikdidaktik: u. a. Neubrand, Klieme, Lüdtke & Neubrand, 2002; van de Rijt, Godfrey, Aubrey, van Luit, Ghesquiere et al., 2003). Andererseits geht es um die Strukturierung dieser Kompetenzbereiche über Teilkompetenzen. Hierfür liegen insbesondere aus dem Bereich der Psychologischen Diagnostik elaborierte Taxonomien (vgl. u. a. Anderson & Krathwohl, 2001) und außer- sowie innerfachliche Modellierungsprozesse vor (Klieme, Neubrand & Lüdtke, 2001), die ihrerseits die Fachdidaktik der entsprechenden Disziplin maßgeblich beeinflusst haben.

Die Grundidee besteht darin, dass sich kontextspezifische Situationen über die kognitiven Strukturen und Prozesse der Anforderungsbewältigung modellieren lassen: Bei Sprachkompetenz kann z. B. gefragt werden, ob es sich um die Bearbeitung von geschriebenen oder gesprochenen Inhalten handelt und in welchem Ausmaß der Erfolg bei der Anforderungsbewältigung vom Vokabelschatz oder von der Beherrschung grammatikalischer Regeln abhängig ist. Die Herausforderung besteht einerseits darin, den Abstraktionsgrad des Modells möglichst gering zu halten, um die Testentwicklung zu erleichtern; andererseits jedoch bedürfen diese Modelle einer gewissen Abstraktion, um mit ihnen Kompetenz in einer spezifischen Domäne auch global beschreiben zu können. So werden in den naturwissenschaftlich oder auch sprachlich orientierten Kompetenzstrukturmodellen die verschiedenen Kompetenzbereiche in der Regel über konkrete kontextspezifische Handlungen differenziert: Argumentieren, Modellieren, Nutzung von mathematischen, physikalischen, chemischen oder biologischen Basiskonzepten bzw. Lesen, Hören, Sprechen in der Mutter- oder Fremdsprache. Ein sehr anschauliches Beispiel eines Kompetenzstrukturmodells ist der DESI-Studie (Deutsch Englisch Schülerleistungen International; Jude & Klieme, 2007; Hartig & Klieme, 2006) entnommen:

Den einzelnen Kompetenzdimensionen Rezeption, Bewusstheit und Produktion sind entsprechende Teilleistungen zugeordnet. In dieser Zusammenstellung ist der Zusammenhang zwischen Struktur der Anforderungssituation und Struktur der Kompetenz sehr deutlich erkennbar. Mit einer Anforderungssituation, die auf das Lesen eines Textes abstellt, kann kein grammatikalisches Verständnis erfasst werden. Oder anders formuliert: Zur Erfassung grammatikalischer Fähigkeiten sind Anforderungssituationen zu konstruieren, die auf den Umgang mit Sprache im Sinne der Bewusstheit von Regeln abstellen. Die Ausgangsfrage ist folglich,

welche Kompetenzen in einem spezifischen Zusammenhang erfasst werden können und in welcher Beziehung diese zueinander stehen.

Abbildung 2.2 Kompetenzstrukturmodell DESI-Studie



Die verschiedenen Dimensionen lassen sich faktorenanalytisch mittels Interkorrelationen der Messungen bestimmen; werden theoretische Modellierungen vorgenommen, lässt sich die Gültigkeit des Modells beispielsweise über lineare Strukturgleichungsmodelle prüfen. Dass die kognitive Modellierung von Kompetenzen jedoch ein stärker inhaltlich akzentuiertes als empirisches Problem ist, zeigen die PISA-Befunde. Für PISA 2003 berichtet u. a. Wu (2004), dass sich für die erfassten Kompetenzbereiche Mathematik, Naturwissenschaften, Lesen und Problemlösen messfehlerbereinigte Interkorrelationen von 0,77 bis 0,89 zeigen. Die einzelnen Kompetenzbereiche sind demnach empirisch nicht eindeutig zu isolieren, wohl aber inhaltlich konkret zu beschreiben und auf diesem Wege zu differenzieren. Diese Befunde stützen die Argumentation derjenigen, die dafür plädieren, dass für die in PISA erfassten Performanzen eine empirisch einheitliche und bildungsabhängige kognitive Fähigkeit verantwortlich ist (vgl. insbesondere Rindermann, 2006, S. 70). Demnach messen die internationalen Vergleichsstudien domänenübergreifend eine einheitliche Merkmalsdimension. Diese Annahme wurde durch die Gegenüberstellung eines g-Faktormodells, in dem die Intelligenz (g) als eindimensionaler Faktor die Lösung der Anforderungssituationen erklärt, und eines Nested-Faktormodells, das mehrere Prädiktoren für die

Aufklärung interindividueller Performanzunterschiede annimmt, geprüft. Die deskriptiven Fit-Indizes und interferenzstatistische Vergleiche sprechen für das Nested-Faktormodell und damit für ein mehrdimensionales Kompetenzmodell (vgl. detailliert Baumert, Brunner, Lüdtke & Trautwein, 2007; Brunner, 2005).

Kompetenzniveaumodelle

Kompetenzniveaumodelle befassen sich mit einer kriteriumsorientierten Interpretation der Kompetenzmesswerte, also mit der konkreten inhaltlichen Beschreibung empirisch erfasster Kompetenzen.

„The concept of criterion referenced measurement was conceived to encourage the development of procedures whereby assessments of proficiency could be referenced to stages along progressions of increasing competence“ (Glaser, 1981, p. 935).

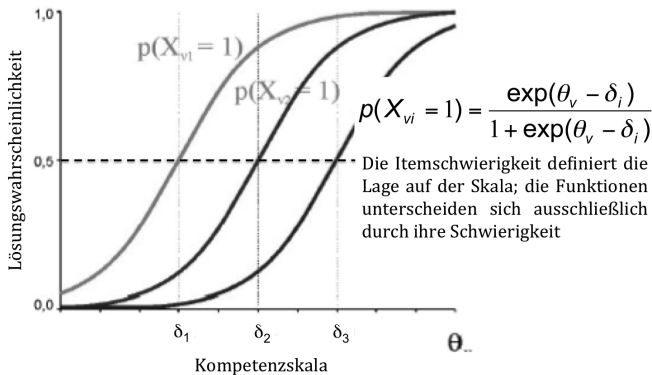
Zentral ist hier die Frage, über welchen Grad an Kompetenz ein Individuum verfügt. Wird diese Frage bezogen auf die Anforderungssituationen umformuliert, geben Kompetenzniveaumodelle darüber Auskunft, welche spezifischen Anforderungssituationen sich bei welchem Kompetenzniveau gerade noch bewältigen lassen. Diese Graduierung von Kompetenzen erfolgt bislang vorrangig formal (vgl. u. a. OECD, 1995; 2003), indem über probabilistische Testmodelle Kompetenzstufen charakterisiert werden. Diese Verfahren werden im Bereich der Pädagogik als defizitär bewertet, da sie keine Analyse der Leistungsprozesse beinhalten, sondern lediglich die Gesamtschwierigkeit einer spezifischen Anforderungssituation bestimmen (vgl. u. a. Neubrand, Klieme, Lüdtke & Neubrand, 2002).

„Several principles were considered for developing and establishing a useful meaning for ‚being at a level‘, and therefore for determining an approach to locating cut-off points between levels and associating students with them: First, it is important to understand that [...] there are no natural breaking points to mark borderlines between stages along these continua. Dividing each of these continua into levels, though useful for communication about students’ development, is essentially arbitrary“ (Adams & Wu 2002, p. 197).

Empirisch sind Kompetenzniveaus bzw. Kompetenzstufen nichts anderes als formale Abschnitte auf der kontinuierlichen Skala der Kompetenzmesswerte (Beaton & Allen, 1992). Die Interpretation dieser Abschnitte ist folglich nur unter Einbezug der konkreten Anforderungssituation möglich und sinnvoll. Probabilistische Testmodelle, wie beispielsweise das Rasch-Modell, machen sich diese Beziehung zwischen Inhalt der Anforderungssituation und beobachteter Leistung in der Anforderungssituation zu nutze. Die Lösungswahrscheinlichkeit konkreter Anforderungssituation und die Kompetenz einer Person werden auf einer gemeinsamen Skala in Beziehung gesetzt, um die Frage zu beantworten, welche Anforderungssituation diese Person gerade noch bewältigen kann. Hierfür werden

die Schwierigkeiten der einzelnen Aufgaben auf der jeweiligen Kompetenzskala verortet.⁵ Zur Illustration ist in Abbildung 2.3 die Verankerung von drei verschiedenen Anforderungssituationen auf der Kompetenzskala am Beispiel des dichotomen Rasch-Modells dargestellt.

Abbildung 2.3 Skalierung von Kompetenzen im dichotomen Rasch-Modell



Die Schwierigkeit einer Anforderungssituation ist als der Punkt auf der Skala definiert, an dem Personen mit genau entsprechender Fähigkeit die Anforderungssituation mit einer Wahrscheinlichkeit von 50 Prozent lösen können. Die dargestellten Funktionen unterscheiden sich ausschließlich durch ihre Schwierigkeitsparameter. So wie die Schwierigkeit der Anforderungssituationen in Beziehung zur Fähigkeit einer Person gesetzt werden können, lassen sich umgekehrt auch die Leistungswerte einer Person auf eine spezifische Anforderungssituation beziehen. Personen, die eine Anforderungssituation der Schwierigkeit δ_2 mit einer Wahrscheinlichkeit von 50 Prozent lösen können, verfügen über die Fähigkeit θ_2 , die genau der Schwierigkeit der Anforderungssituation entspricht. Personen mit der Fähigkeit θ_2 können die Anforderungssituation 1, die sich durch die Schwierigkeit δ_1 charakterisieren lässt, folglich mit einer höheren Wahrscheinlichkeit ($p_{(v1)} > 0,5$) und die Anforderungssituation 3 mit einer geringeren Wahrscheinlichkeit ($p_{(v3)} < 0,5$) lösen. Entscheidend für die Lösungswahrscheinlichkeit ist die Differenz zwischen der individuellen Fähigkeit θ_v und der Schwierigkeit δ_i der jeweiligen Anforderungssituation i .

Ein solches Modell erlaubt es, Aussagen über die Interaktion zwischen getesteter Person und zu lösender Anforderungssituation zu treffen. Diese Aussagen

⁵ Die folgenden Bemerkungen greifen der detaillierten Darstellung in Kapitel 4.1.2 vor; sie erfolgen hier, um in der Systematik der Kompetenzdifferenzierung zu bleiben.

konzentrieren sich dabei vorrangig auf kognitive Strukturen, mit denen sich Unterschiede zwischen den getesteten Personen beschreiben lassen. Es ist die Frage zu beantworten, welche kognitiven Strukturen als aktives Zugriffswissen zur Bewältigung einer spezifischen Anforderungssituation notwendig sind. Hypothesen über den Lösungsprozess werden folglich in Hypothesen über die Inhalte der Anforderungssituationen und deren Schwierigkeitsparameter übersetzt, um Unterschiede zwischen den getesteten Personen in Form von Niveaus oder Abschnitten auf der kontinuierlichen Kompetenzskala interpretieren zu können (vgl. Popham, 1993, p. 130).

Kompetenzentwicklungsmodelle

Kompetenzentwicklungsmodelle spezifizieren den Verlauf des Kompetenzerwerbs; sie können damit Fragen der Erlern- und Vermittelbarkeit von Kompetenzen beantworten und lassen sich folglich insbesondere als Planungsmodelle für Lehr-Lernprozesse nutzen. Sie sind als empirisch fundierte Modelle zu definieren, die Strukturen und Entwicklungsverläufe von Kompetenzen unter Berücksichtigung spezifischer Inhalte aufführen. Das Ziel von Kompetenzentwicklungsmodellen besteht darin, Kompetenzstufen bzw. -niveaus nicht allein aus der Analyse der Anforderungssituation abzuleiten (vgl. Kompetenzniveau Modelle), sondern die Stufen als entwicklungspsychologisch begründete qualitative Unterschiede zu interpretieren (vgl. IPN, 2003, S. 18).

Vorbild hierfür sind in erster Linie die Arbeiten Piagets zum Modell der kognitiven Entwicklung (u. a. 1959). Zentral ist die Annahme, dass in jeder Entwicklungsstufe die vorhergehenden einfacheren Stufen differenziert erfasst und neu integriert werden. Dies impliziert, dass die Vorgänge beim Lernenden sehr detailliert zu erfassen sind. So sind Art und Umfang der kognitiven Repräsentation, Lernschwierigkeiten bzw. -hindernisse, Lerngelegenheiten oder auch vorurteilliche Vorstellungen Faktoren, die die qualitativen Unterschiede der Denk- und Problemlösemodi hinsichtlich gleicher Probleme in den verschiedenen Alters- bzw. Entwicklungsstufen begründen können.

Ein Beispiel eines Kompetenzmodells, das sich an den Arbeiten Piagets orientiert, befasst sich mit Entwicklungsstudien zur mathematischen Grundbildung (van de Rijt, van Luit & Pennings, 1999). Für acht theoretisch abgeleitete und fachbezogene Aspekte mathematischen Wissens – z. B. „concept of comparison“, „seriation“, „structured counting“ – wurden Items konstruiert und in eine unidimensionale Skala transformiert, die mathematische Kompetenz differenziert abzubilden vermag (vgl. auch van de Rijt, Godfrey, Aubrey, van Luit, Ghesquiere et al., 2003). Weitere Beispiele finden sich in Anlehnung an die bei PISA beschriebene Scientific Literacy für den Bereich des Experimentierens (vgl. IPN, 2003).

Die Modelle machen meines Erachtens Entwicklungsstufen der Kompetenz nicht plausibel. Sie zeigen, welche Herausforderungen und Anforderungen an die Stufenbildung zu stellen sind, und können illustrieren, aus welchen Gründen der Forschungsstand zu empirisch abgesicherten Kompetenzentwicklungsmodellen in der Literatur defizitär ist. Die Modellierung von Kompetenzerwerb und dessen

Entwicklung erfordert ein integriertes Zusammenspiel von empirischer Forschung, Fachdidaktik, allgemeiner Pädagogik, pädagogischen Psychologie und Entwicklungspsychologie (dazu auch IPN, 2003).

„Developmental learning is the discipline which integrates psychology, education, neurology, human physiology, computer, and information sciences, embracing learning as an evolutionary process where beliefs (knowledge structures) evolve through experience“ (Hashway, 1998, p. 203).

Insbesondere die empirische Fundierung von Kompetenzentwicklungsmodellen stößt dann an Grenzen, wenn es um mehr geht, als die Herausbildung und Entwicklung von kognitiven Strukturen im Experten-Novizen-Vergleich zu beschreiben. Kompetenzentwicklungsmodelle müssen davon ausgehen, dass Kompetenzerwerb und -verlauf von vielseitigen Interaktionen des Lernenden mit der Umwelt geprägt sind. Bislang ist dies, in den Strukturgleichungsanalysen und Distanzmodellen, die vorrangig als empirische Methoden eingesetzt werden, nicht befriedigend umgesetzt (Hashway, 1998).

2.3.2 Modellierung von Kompetenzen in den internationalen Vergleichsstudien

Die internationalen Vergleichsstudien der OECD wie IALS (International Adult Literacy Survey), PISA (Programme for International Student Assessment) oder PIAAC (Programme for the International Assessment of Adult Competencies) konzentrieren sich auf „literacy“- sowie „numeracy“-Ansätze. Dies bedeutet, dass text-, bild- und/oder zahlssprachliche Informationen in Alltagsbezügen zu verarbeiten und zu nutzen sind, um eigene Ziele erreichen und das eigene Wissen und Handlungspotential ausbauen zu können (vgl. OECD, 1995; Deutsches PISA-Konsortium, 2001).

„Key features associated with PISA include: [...] its innovative ‚literacy‘ concept, which is concerned with the capacity of students to apply knowledge and skills in key subject areas and to analyse, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations“ (OECD, 2005, p. 13).

PISA konzentriert sich dabei auf Lesekompetenzen sowie mathematische und naturwissenschaftliche Kompetenzen. In jedem PISA-Zyklus dominiert einer dieser drei Bereiche den Test (ca. 70 Prozent der Testzeit): Lesekompetenz in den Jahren 2000 und 2009, mathematische Kompetenz im Jahr 2003 und naturwissenschaftliche Kompetenz im Jahr 2006. Diese zentralen Kompetenzbereiche sind für alle Erhebungszyklen definiert und werden in ihrer Struktur und Graduierung mittels probabilistischer Verfahren analysiert und geprüft (vgl. Abbildung 2.4).

Abbildung 2.4 Definition der Kompetenzbereiche (OECD, 2005, p. 14)

Mathematical literacy: An individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individuals' life as a constructive, concerned and reflective citizen.

Reading literacy: An individual's capacity to understand, use and reflect on written texts, in order to achieve one's goals, to develop one's knowledge and potential and to participate in society.

Scientific literacy: The capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity.

Problem-solving skills: An individual's capacity to use cognitive processes to confront and resolve real, cross-disciplinary situations where the solution path is not immediately obvious and where the literacy domains or curricular areas that might be applicable are not within a single domain of mathematics, science or reading.

Die Definitionen verdeutlichen anschaulich, dass das Ziel der PISA-Studien darin liegt, Kompetenzen als Ergebnisse von Bildungsprozessen in unterschiedlichen Kompetenzbereichen oder Domänen zu erfassen. Es sollen die anwachsenden Anforderungen im gesellschaftlichen Austausch und die dazu notwendigen Fähigkeiten dargestellt werden (vgl. OECD, 1995, p. 22).

Kompetenzstruktur

Für die Darstellung der Kompetenzstruktur werden für jeden Kompetenzbereich Skalen gewählt, die sich vorrangig an kognitive Taxonomien anlehnen und sich i. d. R. auf die kognitiven Dimensionen ‚Identifizieren‘, ‚Verstehen‘ und ‚Operieren‘ beziehen lassen. Diese Skalen tragen dem Umstand Rechnung, dass Kompetenzen kontextspezifisch sind und sich sowohl über situations- als auch personenspezifische Komponenten charakterisieren lassen. So basiert beispielsweise die Beschreibung der Lesekompetenz der PISA-Studie darauf, Anforderungen beim Lesen situationsspezifisch zu differenzieren (z. B. Textreorganisation vs. Textproduktion), die auf Seiten der Person durch die entsprechenden Fähigkeiten zu bewältigen sind (z. B. Informationen herausfiltern vs. Interpretationen entwickeln) (vgl. Deutsches PISA-Konsortium, 2001, S. 88ff.). Dieses Vorgehen entspricht im weitesten Sinne der Taxonomy of Educational Objectives (Bloom, 1956) und deren Revision durch Anderson und Krathwohl (2001). Ausschlaggebend ist, dass zur Beschreibung von Lernzielen bzw. Kompetenzen neben der über kognitive Dimensionen abgebildeten Verhaltensseite auch immer der Bezug auf Inhalte ge-

hört. In diesen Kontext ist die Diskussion um Modellierungsprozesse einzubinden (vgl. u. a. Blum & Neubrand, 1998). Der Entwicklung von Skalen liegt ein umfangreicher, mehrstufiger Prozess zu Grunde, der für alle gewählten Kompetenzbereiche bzw. Domänen identisch ist: (1) die theoretische, fachdidaktisch-orientierte Identifizierung möglicher Skalen, (2) die Konstruktion entsprechender Items, (3) die Itemanalyse, (4) die Einbindung der Items in einen Feldtest und die (5) sich anschließende Definition der Kompetenzdimensionen für den Haupttest.

Kompetenzniveau

Hinsichtlich der Kompetenzniveaus der OECD-Studien gelten folgende Überlegungen: Zur Darstellung der Kompetenzbereiche ist eine Niveaustufung in den Leistungen notwendig und sinnvoll, um feststellen zu können, welche Anforderungssituationen Individuen bei welcher Ausprägung einer Kompetenz bewältigen.

„Underlying the concept of achievement measurement is the notion of a continuum of knowledge acquisition ranging from no proficiency at all to perfect performance. An individual's achievement level falls at some point on this continuum as indicated by the behaviors he displays during testing [...] The standard against which a student's performance is compared when measured in this manner is the behavior which defines each point along the achievement continuum“ (Glaser, 1963, pp. 519/520).

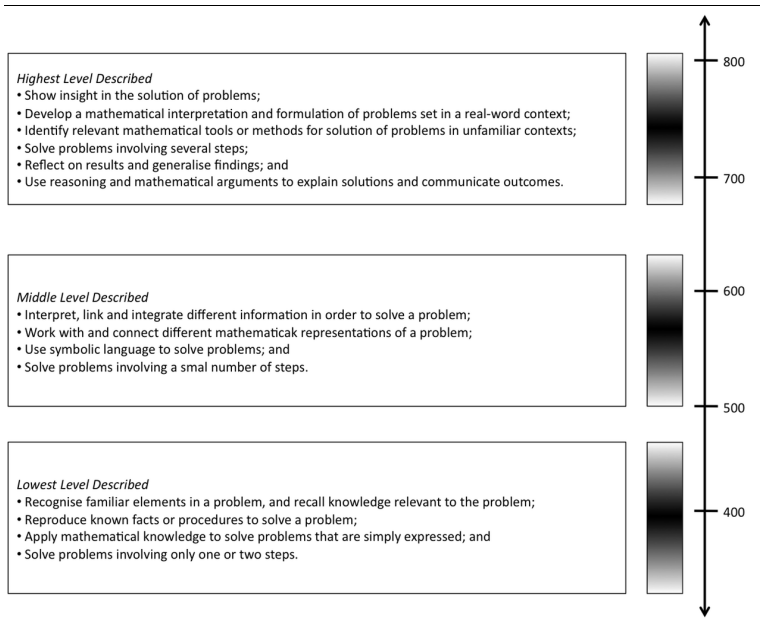
Die OECD-Vergleichsstudien definieren diese Niveaustufen post-hoc auf Grundlage der testmodell-konform übrig gebliebenen Aufgaben auf der Grundlage eines zweikategoriellen Raschmodells, wobei die Festlegung der Kompetenzstufengrenzen durch Unterteilung der Fähigkeitsskala hinsichtlich der Mindestlösungschancen erfolgt (u. a. Klieme & Leutner, 2006, S. 7). Dieses Verfahren, ebenso wie die von Beaton und Allen (1992) vorgeschlagene Festlegung der Kompetenzstufengrenzen durch rekursive Bestimmung von Markieraufgaben, sind bislang mit Blick auf die inhaltliche Ausgestaltung noch nicht hinreichend theoretisch begründet. Um zu einer theoriegeleiteten Modellierung der Niveaus zu gelangen, werden insbesondere fachspezifische Ansätze und Zusammenhänge integriert (vgl. u. a. mathematisches Problemlösen: Klieme & Baumert, 2001). Hierfür sind die Arbeiten aus dem Bereich der mathematischen Kompetenz besonders anschaulich, und meines Erachtens lassen sich aus diesen auch wertvolle Implikationen für den Bereich der Berufs- und Wirtschaftspädagogik ableiten. Die Kernidee der mathematischen Kompetenzmodellierung besteht darin, dass sich die Bewältigung von Anforderungssituationen in charakteristische Modellierungsschritte zerlegen lässt. Diese Modellierungsschritte machen Anforderungssituationen und die entsprechenden Lösungsprozesse miteinander vergleichbar. Diese Modellierungsschritte sind:

- die Übersetzung einer gegebenen Anforderungssituation in ein konkretes mathematisches Problem (das *Mathematisieren* der Situation),

- die algorithmische oder begriffliche Lösung des Problems (innermathematische *Verarbeitung* des Modells),
- die Übersetzung der mathematisch gefundenen Lösung im Hinblick auf die Anforderungssituation (*Interpretation* von Modell und Lösung) und
- die Überprüfung der Richtigkeit und Anwendbarkeit des gewählten mathematischen Modells (*Validierung* des Modells).

Die Modellierungsschritte sind für jede Anforderungssituation identisch, aber unterschiedlich komplex. Die Komplexität des mathematischen Modellierungsprozesses ist dabei von mehreren Faktoren abhängig. So spielen beispielsweise die Anzahl der Lösungsmöglichkeiten oder der Umfang der zu verarbeitenden lösungsrelevanten Variablen eine Rolle; die Art der Repräsentation sowie das Niveau der Reflexion sind ebenfalls von Bedeutung. Diese Faktoren, die die Schwierigkeit einer gegebenen Anforderungssituation beeinflussen, ermöglichen es, die empirisch erzeugten Kompetenzstufen inhaltlich zu interpretieren. Abbildung 2.5 zeigt exemplarisch die Stufenbildung für den Bereich der mathematischen Kompetenz auf.

Abbildung 2.5 *The Mathematical Literacy Scale* (Adams & Wu, 2004, p. 209)



Die Konstruktion dieser Kompetenzstufen ist als Hilfsmittel zu interpretieren, um die Ergebnisse der PISA-Studie aussagekräftig zu illustrieren. Es lässt sich mit ihrer Hilfe einerseits zeigen, wie Populationen auf den verschiedenen Stufen ver-

teilt sind; andererseits können über die Verteilung auf den Stufen auch Aussagen über Grundfähigkeiten in den Sub-Populationen getroffen werden.

2.4 Kompetenzdefinitionen und Kompetenzmodelle im Bereich der beruflichen Bildung

Die in der beruflichen Bildung geführte Kompetenzdiskussion ist im Gegensatz zu der dargestellten kognitiven Perspektive auf das Konzept der beruflichen Handlungskompetenz ausgerichtet. In diesem Zusammenhang ist die Frage zu stellen, in welchem Umfang dieses Konzept anschlussfähig an aktuelle Entwicklungen der Bildungsforschung ist und welche Optionen sich aus diesem Konzept für ein allgemein akzeptiertes Kompetenzmodell in der beruflichen Bildung ableiten lassen. Die Fokussierung auf das Konzept der beruflichen Handlungskompetenz wird durch die ordnungspolitische Vorgabe in Lehrplänen und Ausbildungsordnungen verstärkt (vgl. Stiller, 1998). Ein wesentliches Desiderat der berufs- und wirtschaftspädagogischen Forschung ist die empirische Fundierung ihrer Kompetenzmodellierungen. Es mangelt nicht an inhaltlich ausgearbeiteten Konzepten, die ihren Ursprung in weiten Teilen in der Anthropologie Heinrich Roths (1971) haben, jedoch aber an einer Umsetzung dieser Konzepte in operationalisierbare Modelle. Roths Ansatz ist der einer ganzheitlichen Handlungsfähigkeit, die sich über Sach-, Sozial- und personale Kompetenz ausdrücken lässt. Sacheinsichtige, sozialeinsichtige und moraleinsichtige Handlungsfähigkeit entwickelt sich in Abhängigkeit von der Lernumwelt, den Lern-, Sozialisations- und Erziehungsprozessen. Diese Typologie von Kompetenzen bzw. Kompetenzgefügen wird in der Literatur im deutlichen Gegensatz zur kognitiv geprägten Kompetenzdiskussion der internationalen Schulleistungsstudien gesehen. Das zentrale Argument ist, dass das Kriterium des Domänenbezugs oder der Fachlichkeit, das dem kognitiven Kompetenzverständnis zugrunde liegt, nicht erfüllt sein kann (vgl. die Beiträge in Erpenbeck & von Rosenstiel, 2003). So wird in der Expertise zur Entwicklung nationaler Bildungsstandards der Kompetenzbegriff, wie er insbesondere aus den OECD-Vergleichsstudien bekannt ist, strikt von dem in der Berufs- und Wirtschaftspädagogik verwendeten Konzept der Handlungskompetenz abgegrenzt (Klieme, Avenarius, Blum, Döbrich, Gruber et al., 2003, S. 22). Diese Divergenz zeigt sich jedoch meines Erachtens nicht in den Systematisierungen von Kompetenz innerhalb der Berufs- und Wirtschaftspädagogik. Die hier diskutierten Modelle enthalten ebenso die Ebenen der internalen Bedingungen, der aktuellen Vollzüge und der Umgebungsbedingungen, wie sie für die Formulierung von Kompetenzmodellen im allgemeinbildenden Bereich konzeptionell herangezogen werden (vgl. dazu u. a. Böhner & Straka, 2005).

Exemplarisch soll dies an ausgewählten Kompetenzmodellen, wie sie in der Berufs- und Wirtschaftspädagogik zu finden sind, diskutiert werden. Der Fokus liegt hierbei nicht auf der Vollständigkeit möglicher Modelle, sondern auf der Veranschaulichung der Gemeinsamkeiten und den daraus ableitbaren Empfehlungen zur Kompetenzmodellierung in beruflichen Handlungsfeldern.

2.4.1 Beispiele für Kompetenzmodelle der Berufs- und Wirtschaftspädagogik

Die Begrifflichkeit der Handlungskompetenz, aufgeschlüsselt über die Dimensionen Sach-/Methoden-, Sozial- und Selbstkompetenz, ist vorrangig durch Lothar Reetz (u. a. 1989a; 1989b; 1990; 1999) in der berufs- und wirtschaftspädagogischen Forschung verortet. Reetz arbeitet theoretisch analysierend heraus, wie sich das Konzept der Schlüsselqualifikationen in das Konzept einer Persönlichkeits-theorie einbinden lässt, mit dem Ziel, Persönlichkeitsentwicklung in das Zentrum von Lehr-Lern- und Arbeitsprozessen zu stellen.

„Diese genuin pädagogische Zielsetzung hat ihre eigene Berechtigung und darf nicht instrumentalisiert werden. Das ist deswegen besonders hervorzuheben, weil inzwischen auch die Betriebe immer deutlicher erkennen, daß die umfassende Persönlichkeitsentwicklung eine wichtige Vorbedingung für das Erbringen ausgezeichneter Arbeitsleistung darstellt, so daß man in dieser Perspektive von einer ‚Koinzidenz ökonomischer und pädagogischer Vernunft‘ sprechen kann“ (Deutsche Forschungsgemeinschaft, 1990, S. VII).

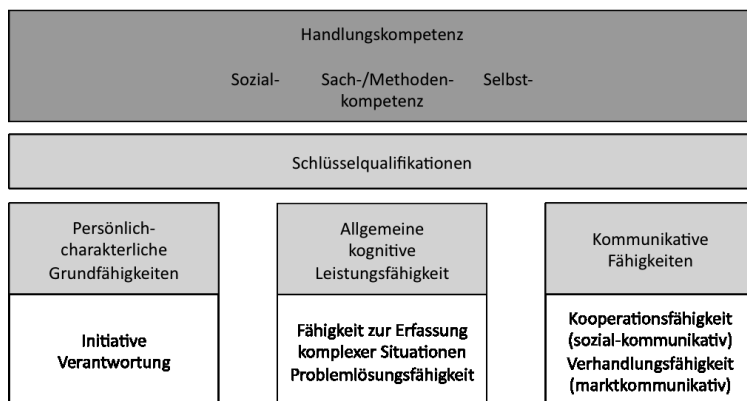
In den Debatten um das Kompetenzverständnis der Berufs- und Wirtschaftspädagogik wird übersehen oder vergessen, dass Reetz Handlungskompetenz ganz im Sinne kognitivistischer Tradition definiert hat (1999, S. 35): Handlungskompetenz wird als die berufliche Fähigkeit bezeichnet, die es Individuen erlaubt, in spezifischen Anforderungssituationen zu handeln. Der Erwerb von Kompetenzen ist – im Sinne Piagets – das Resultat von Entwicklungs- und Lernprozessen (vgl. hierzu Bader & Müller, 2002).

Modelle der Handlungskompetenz

Reetz selbst versteht sein Modell der beruflichen Handlungskompetenz als „Einstieg in die Operationalisierung und Aufklärung über die wechselseitige Bedingtheit“ von Kompetenzen und Schlüsselqualifikationen (Reetz, 1999, S. 37; vgl. Abbildung 2.6). Das Modell bietet eine Komplexitätshierarchie und stellt horizontale, auf die Unterscheidung von Kompetenzdimensionen abzielende Relationen dar. Handlungskompetenz ist folglich die Oberkategorie, die sich als Integrationsergebnis von Sach-/Methoden-, Sozial- und Selbstkompetenz in spezifischen Anforderungssituationen definieren lässt. Sachkompetenz wird als Fähigkeit eines Individuums zu sacheinsichtigem Handeln in konkreten Anforderungssituationen unter Rückgriff auf situativ erlernte Strategien und Heuristiken (Methodenkompetenz) definiert. Sozialkompetenz bezieht sich auf sozialeinsichtiges und kommunikatives Handeln, Selbstkompetenz entsprechend auf moraleinsichtiges Handeln. Die starke Nähe zu Roth stellt dieses Modell in eine entwicklungspsychologische Perspektive, die jedoch auf Grund des zentralen Beziehungsgefüges von Kompetenz und Qualifikation nur graduell angelegt ist. Die Fragen des Kompetenzerwerbs und der Kompetenzentwicklung sind mit der theoretischen Verortung von Kompetenzbeziehungen und deren Bezug auf konkrete Anforderungssi-

tuationen noch nicht beantwortet. Arnold (1997, S. 290) kritisiert in diesem Zusammenhang, der Kompetenzentwicklungsbegriff werde in der berufs- und wirtschaftspädagogischen Literatur weder entwicklungspsychologisch noch im Sinne einer pädagogischen Identitätsbildung verwendet – ein Anspruch, der jedoch von Reetz nicht intendiert ist.

Abbildung 2.6 Modell der Handlungskompetenz (Reetz, 1999, S. 38)

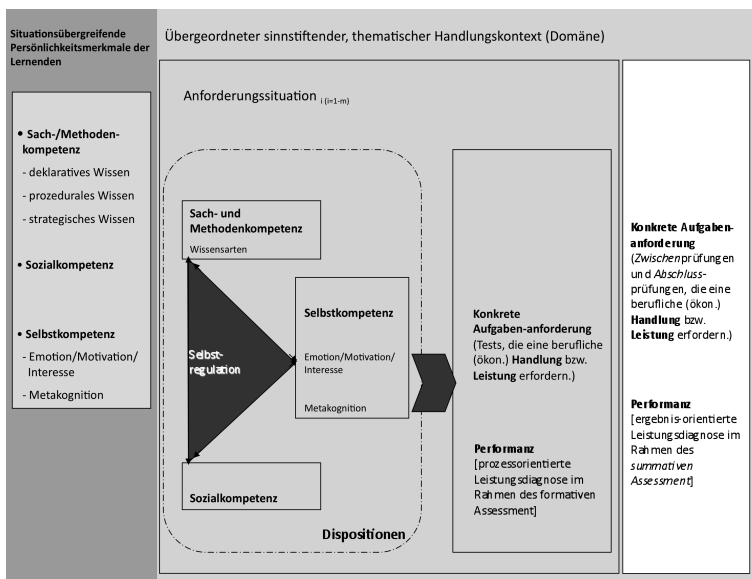


Aktuell wird das Modell von Reetz beispielsweise im Rahmen eines BiBB-Projekts zu Kompetenzstandards in der beruflichen Bildung adaptiert (Hensge, Görmär, Lorig, Molitor & Schreiber, 2008). Die Autoren stellen die Kompetenzdimensionen in konkrete Anforderungssituationen, die sie als Arbeits- und Geschäftsprozesse entsprechend spezifischer Berufsbildpositionen abbilden. Dieses Vorgehen hat zum einen die stärkere Explizierung des Kontextes und zum anderen die Strukturierung der Handlungskompetenz unter Einbeziehung fachlicher, methodischer und sozialer Kompetenzdimensionen zum Ziel. Es zeigen sich (optisch) deutliche Parallelen zu dem Kompetenzraster, das für ein international vergleichendes Large-Scale Assessment in der beruflichen Bildung (VET-LSA) ausgearbeitet wurde. In der Expertise wird vorgeschlagen, in Anlehnung an die Systematisierung von Selbst-, Sach-, Methoden- und Sozialkompetenz interne Bedingungen berufsübergreifender und berufsfachlicher Kompetenzen zu erheben. Hierfür werden die Teilkomponenten der beruflichen Handlungskompetenz mittels individueller kognitiver Dispositionen und nicht anhand von überdauernden Handlungsparametern erfasst (Baethge, Achtenhagen, Arends, Babic, Baethge-Kinsky & Weber, 2006, S. 52ff.).

Handlungskompetenz als kontextspezifische kognitive Leistungsdisposition

In dem Raster zur Expertise eines international vergleichenden VET-LSA werden das Ineinandergreifen und das Wechselspiel verschiedener Kompetenzbereiche und der erforderlichen bzw. zu berücksichtigenden individuellen Fähigkeiten systemisch eindeutig an die Anforderungsbewältigung in spezifischen beruflichen Situationen gebunden. Auch zeigt sich, dass zwischen beruflicher Handlungsfähigkeit und kognitiver Kompetenzdefinition weniger Differenzen bestehen, als dies angenommen wird. Berufliche Handlungskompetenz wird hier als aktives operationales und funktionales Wissen unter Hinzunahme psychologischer Vermittlungsleistungen wie Motivation, Einstellungen und Überzeugungen definiert. Die Beziehungsgefüge des Kompetenzrasters gehen zurück auf das DFG-Projekt „Integrierte Kompetenzförderung in den beruflichen Fächern des Wirtschaftsgymnasiums“ (Achtenhagen & Winther, 2006; Winther, 2006).

Abbildung 2.7 Modell der Handlungskompetenz
(Achtenhagen & Winther, 2006, S. 349)



Kompetenzen werden hier im Sinne der Schulleistungsstudien als kognitive Leistungsdispositionen verstanden, die sich auf einen übergeordneten sinnstiftenden, thematischen Handlungskontext (Domäne) beziehen. Dabei wird von einem generativen Verhältnis zwischen Kompetenz und Aufgabenbearbeitung gesprochen (Performanz). Dies bedeutet, dass sich die Dispositionen in der konkre-

ten Aufgabenbearbeitung (auf der performativen Ebene) zeigen und umgekehrt von dieser auf die Kompetenz zurückgeschlossen werden kann. Die Dispositionen generieren somit Verhalten (vgl. dazu auch Weinert, 2001a).

Die Kompetenzdimensionen Sach-/Methoden-, Sozial- und Selbstkompetenz sind in diesem Modell als operationalisierte kognitive Repräsentationen des spezifischen Handlungskontextes zu interpretieren. Dabei sind hier neben dem aktiven Wissenszugriff (deklaratives Wissen und prozedurale Fähigkeiten) insbesondere die Bewusstheit der Handlungsoptionen in der konkreten Anforderungssituation von Bedeutung: im Sinne sacheinsichtiger, sozialeinsichtiger und moral-einsichtiger Verhaltensspielräume. Die Ergebnisse der Kompetenztests wurden in diesem Projekt mittels probabilistischer Verfahren re-analysiert, um Aussagen zur Modellbildung sowie der Graduierung von Kompetenzen ableiten zu können (im Detail Kapitel 5.1; vgl. Winther & Achtenhagen, 2008b; 2009; 2010). Vor dem Hintergrund dieser Befunde könnte als Schlüssel für eine angemessene Modellierung fachlicher Kompetenz eine systematische Variation der Komplexität der Anforderungssituationen dienen.

Brand, Hofmeister und Tramm (2005) betonen in diesem Zusammenhang, dass nicht die Konzeption, sondern die Akzentuierung der Modellierung im berufsbildenden Bereich eine andere ist: Während die Herausforderung der Kompetenz erfassung in Bezug auf eingeführte Schulfächer darin besteht, traditionelle Lerninhalte in Aufgaben und Problemstellungen einzubetten, über die sie empirisch erfasst werden können, besteht die spezifische Herausforderung im Bereich der Berufsbildung darin, berufliche Handlungs- und Orientierungsleistungen abzubilden, die aus ihrem Begründungszusammenhang heraus bereits situiert sind. Oder anders formuliert: Für traditionelle Schulfächer weist der Entwicklungsweg vom Inhalt über die damit zu verknüpfende kognitive Leistung zur situierten Aufgabe; für den Bereich der Berufs- und Wirtschaftspädagogik gälte eine gegenteilige Argumentation, wenn der Entwicklungsweg von der Situation über die in dieser erbrachten kognitiven Leistung zum Inhalt führt (Brand, Hofmeister & Tramm, 2005, S. 6). Dieses Verständnis ist in einen Klassifikationsrahmen für die Itementwicklung im Rahmen der Hamburger „Untersuchung der Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen“ eingeflossen (ULME III; vgl. Lehmann & Seeber, 2007). Dieser Klassifikationsrahmen stellt ein Analyseinstrument zur Überprüfung des Anspruchsniveaus von Testaufgaben dar. Die Testaufgaben werden in Anlehnung an die von Anderson und Krathwohl (2001) vorgelegte Taxonomierung einerseits nach der Art des Wissens und andererseits nach dem angesprochenen kognitiven Anforderungsniveau differenziert. Hinsichtlich der Wissensarten wird zwischen Faktenwissen, Konzeptwissen und prozeduralem Wissen unterschieden, bezüglich der kognitiven Leistungsdimensionen zwischen Reproduktion, Anwendung und Reflexion (Hofmeister, 2005, S. 5). Formal weist dieser Klassifikationsrahmen ähnliche Strukturen auf wie jene Kompetenzcluster, die im Rahmen von PISA 2003 z. B. für die Mathematik entwickelt wurden (Lehmann & Seeber, 2007, S. 27; Seeber, 2005; 2008). Die Einschätzung der Anforderungen der Testaufgaben

erfolgte a priori durch Fachdidaktiker und damit unabhängig von den Ergebnissen der Leistungstests. Hier geben die Autoren zu bedenken, dass bei der Unterscheidung der kognitiven Leistungsdimensionen implizit Annahmen über die erforderlichen vorgelagerten Lernprozesse getroffen werden, die als subjektive Einschätzungen die Einordnung beeinflussen können. Vor diesem Hintergrund weist die Konzeption des ULME-Projekts meines Erachtens drei zentrale Einschränkungen auf, wenn berufliche Kompetenz erfasst werden soll:

- Es werden kaum Aufgaben formuliert, die einem höheren kognitiven Prozess zuzurechnen wären. So erreicht z. B. keine Aufgabe das Niveau der Reflexion (Brand, Hofmeister & Tramm, 2005, S. 17). Der Komplexität der Konstrukts „berufliche Kompetenz“ kann meines Erachtens so nicht entsprochen werden. Zudem wird ein für berufliche Handlungen maßgebliches Entscheidungsmodell nicht implementiert – hier gehen insbesondere die Möglichkeiten einer äußeren Modellierung der Domäne und der dadurch implizierte situierte Charakter beruflicher Anforderungssituationen verloren.
- Es werden keine Kriterien für spezifische Fachkenntnisse festgelegt. Insbesondere im prozeduralen Bereich waren zur Bewältigung der Aufgaben keine domänenspezifischen, sondern domänenübergreifende allgemeine kognitive Fähigkeiten notwendig (Brand, Hofmeister & Tramm, 2005, S. 18).
- Das a priori durch Fachdidaktiker erfasste Anspruchsniveau der Testaufgaben impliziert keine Aussage im Hinblick auf das Schwierigkeitsniveau. Die empirischen Befunde zeigen, dass die Leistungen z. B. im Ausbildungsberuf Industriekaufmann/-frau tendenziell eine bimodale Verteilung mit einem Gipfel im unteren Leistungsbereich und einem flachen Kurvenverlauf zum oberen Leistungsspektrum aufweisen. Gute Leistungen erzielten die Auszubildenden bei Testaufgaben, die gesamtwirtschaftliche und gesellschaftliche Zusammenhänge thematisieren. Leistungsschwächen waren hingegen bei Testaufgaben auffällig – und dies gilt übergreifend für alle untersuchten kaufmännischen Berufe –, die das Quantifizieren ökonomischer Beziehungen zum Inhalt hatten. Selbst leistungsstarke Subgruppen wiesen in diesem Bereich der kaufmännischen Kompetenz ausgeprägte Defizite auf (Lehmann & Seeber, 2007, S. 140).

In Anlehnung an ULME III zeigt eine Studie zur Entwicklung beruflicher Kompetenzen von Bankkaufleuten, welche Effekte personale, schulische und betriebliche Bedingungen auf den Kompetenzerwerb haben. Zur Erfassung beruflicher Kompetenzen wurden ein Fachtest für Bankkaufleute und Items des Wirtschaftskundlichen Bildungstests (u. a. Beck, 2001) eingesetzt. Die gewonnenen Ergebnisse wurden unter Verwendung probabilistischer Testmodelle reanalysiert und überprüft (u. a. Straka, Fehring & Rosendahl, 2008; Rosendahl, Fehring & Straka, 2008).

Die Entwicklung eines Kompetenzmodells für die elektrotechnische Ausbildung, wie es von Nickolaus und Kollegen (Nickolaus, Gschwendtner & Geißel,

2008) vorgestellt wurde, zeigt exemplarisch auf, wie es gelingen kann, Kompetenzmodellierungen und Forschungen im Bereich der Kompetenzentwicklung aufeinander zu beziehen und parallel voranzutreiben. Am Beispiel der Ausbildung im Kfz-Gewerbe wird veranschaulicht, welche Prädiktoren sich für den fachspezifischen Kompetenzaufbau identifizieren lassen und wie eine Modellierung des Fachwissens aussehen kann. Auch bei diesem Projekt handelt es sich um Re-Analysen mittels probabilistischer Verfahren, die insbesondere auf die Kompetenzgraduierung abstellen.

Diese exemplarischen Entwicklungen aus dem kaufmännischen und gewerblichen Bereich zeigen auf, dass sich hier weitere Forschungsdesiderate anschließen und dass die Berufs- und Wirtschaftspädagogik die Herausforderung psychometrischer Kompetenzmodellierungen angenommen hat.

Anders als in den hier vorgestellten Projekten erfolgt innerhalb der Berufs- und Wirtschaftspädagogik bei der Formulierung der Handlungskompetenz in anderen Modellen bislang keine Konkretisierung der aktuellen Vorgänge. Zwar werden in den Ordnungsmitteln (Ausbildungsordnungen und Rahmenrichtlinien) Wissensarten mit Handlungsarten verknüpft; eine Klassifikation der Handlungsarten und Inhaltsbereiche für konkrete Situationen liegt jedoch nicht vor. Beispielsweise haben Breuer und Höhn (1999) in einem Wirtschaftsmodellversuch zu Abschlussprüfungen für den Versicherungskaufmann/die Versicherungskauffrau ein Qualifikationsförderungssystem entwickelt, das handlungsleitende Begriffe unter Rückgriff auf bereichsspezifische Fakten und Konzepte herausstellt. Rauner, Grollmann und Martens (2007) nutzen Expertenratings, um Lernfortschritte von Auszubildenden in der Automobilbranche zu bewerten. In diesen zwei exemplarischen Projekten werden Anforderungen vor dem Hintergrund einer konkreten Domäne konstruiert und der Ausbildungssystematik angepasst. Weitere Hinweise zu einer differenzierteren Systematik finden sich u. a. bei Faulstich (1997); hier werden die Kompetenzbereiche Fach-, Methoden-, Sozial- und Reflexionskompetenz mit sensomotorischen, emotional-affektiven sowie kognitiven Aspekten in einer Matrix verknüpft. Diese Projekte sind als Fortschritt zu bewerten, da unterschiedliche Perspektiven für den Kompetenzbegriff zusammengeführt werden; für praktische Erhebungsverfahren erscheinen diese Systematiken jedoch nur wenig hilfreich (vgl. insb. Gillen & Kaufhold, 2005).

2.4.2 Kompetenzstrukturmodell für die kaufmännische Bildung

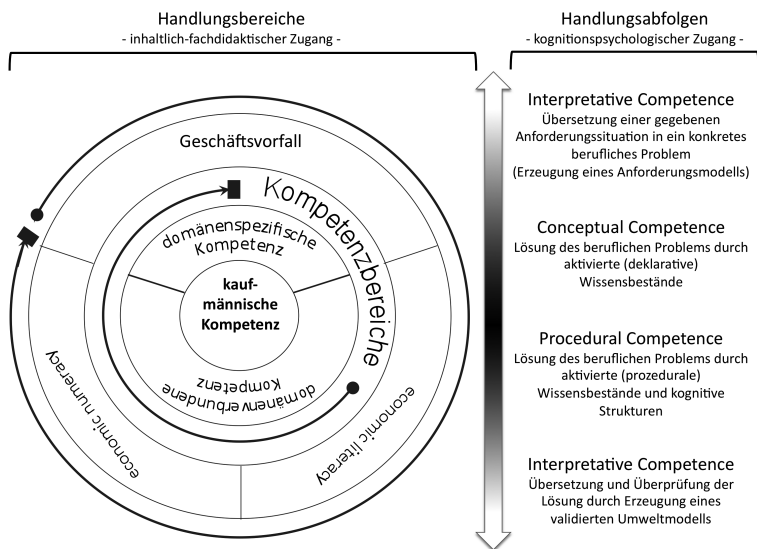
Inhalte und Struktur eines Kompetenzmodells für die kaufmännische Bildung bauen auf der Kompetenzdefinition der PISA-Studien auf (u. a. OECD, 2003). Hieran angelehnt wird für die Analyse von kompetenzgestützter Bildungsmobilität kaufmännische Kompetenz wie folgt definiert:

„Kaufmännische Kompetenz ist die Fähigkeit, auf Grundlage eines systemischen Verstehens betrieblicher Teilprozesse und deren Rekonstruktion aus realen Unternehmensdaten in berufsrealen Situationen unternehmerische Entscheidungen treffen und diese validieren zu können,

um damit das eigene Wissens- und Handlungspotential vor dem Hintergrund der Entwicklung individueller beruflicher Regulationsfähigkeit auszubauen“ (Winther & Achtenhagen, 2008c, S. 100).

Diese Definition unterstreicht die Bedeutung der kompetenten Erstellung angemessener Handlungspläne im Rahmen beruflicher Aktivität. Ziel der Modellierung ist es, die handlungsleitende Funktion von Wissensprinzipien einschließlich ihrer einschränkenden Bedingungen herauszustellen (vgl. Gelman & Greeno, 1989, p. 172). Hierbei wird für die Modellierung kaufmännischer Kompetenz zwischen Handlungsbereichen und Handlungsabfolgen unterschieden. Für die Modellierung der Handlungsbereiche wird einer inhaltlich-fachdidaktischen Logik und für die Modellierung der Handlungsabfolgen einer kognitionspsychologischen Logik gefolgt (vgl. Abb. 2.8)

Abbildung 2.8 Kompetenzstrukturmodell für die kaufmännische Bildung



Hinsichtlich der *Handlungsbereiche* wird vorgeschlagen, Kompetenzbereiche zu erfassen, die explizit auf unterschiedlichen fachlichen Zugriffen auf die Domäne basieren, um tatsächlich einen sinnhaften beruflichen Handlungskontext bestimmen zu können. Die bislang erzielten Ergebnisse bestätigen eine zwischen domänenverbundenen und domänenspezifischen Kompetenzen differenzierende Erfassung: einerseits zur Abbildung kaufmännischer Grundfertigkeiten die do-

mänenverbundenen Bereiche „economic literacy“ sowie „economic numeracy“ zur Erschließung ökonomischen Wissens sowie andererseits zur Bearbeitung komplexer ökonomischer Zusammenhänge domänenspezifische Geschäftsvorfälle auf Basis festgelegter Arbeits- und Geschäftsprozesse. Die Hinzunahme der domänenverbundenen Kompetenzbereiche „economic literacy“ sowie „economic numeracy“ soll die für das (wirtschafts-)berufliche Handeln notwendigen kaufmännischen Grundfertigkeiten abbilden und testen und beschreibt die in den Curricula angesprochenen Interdependenzen dieser Bereiche. Das Konzept der „economic literacy“ stellt auf die kulturelle Teilhabe und Entwicklungszustände in wirtschaftsbezogenen Kontexten ab. In Anlehnung an den International Adult Literacy Survey (IALS; OECD, 1995) sowie an die vom National Council on Economic Education herausgegebenen Standards und den Test ökonomischer Grundbildung (als deutsche Adaption vgl. Beck, 2001) werden text- und bildsprachliche Kenntnisse sowie das Verständnis von quantitativen Werten und Verhältnissen in wirtschaftlichen Alltagskontexten erfasst. Der Bereich der „economic numeracy“ erfasst grundlegende mathematische Kenntnisse und Fertigkeiten vor dem Hintergrund konkreter Unternehmensprozesse aus zwei Perspektiven: (1) Zum einen sollen Rechenoperationen mit unternehmensinternen Bezügen aufgestellt und gelöst werden können, (2) zum anderen sind unternehmensinterne Realwerte hinsichtlich ihres Zustandekommens und ihrer Aussagekraft zu interpretieren. In den Ergebnissen wird bemerkenswert deutlich, dass domänenspezifische Leistungen in den zu bearbeitenden Geschäftsvorfällen von beiden domänenverbundenen Leistungsbereichen abhängig sind.

Im vorliegenden kaufmännischen Kompetenzmodell werden domänenspezifische Kompetenzbereiche über so genannte Geschäftsvorfälle eingeführt. Unter *Geschäftsvorfall* wird in diesem Zusammenhang eine über konkrete Arbeitsprozesse definierbare Anforderungssituation verstanden, die sich in unternehmensspezifischen Geschäftsprozessen verorten lässt. Ein Geschäftsvorfall stellt damit eine inhaltlich-systematisch ausgestaltete arbeitsplatzspezifische Situation dar, die sich detailliert im Hinblick auf das Anforderungsniveau, die Handlungsspielräume sowie die intendierten Zielsetzungen beschreiben lässt. Für die Definitionen von Geschäftsvorfällen ist eine Orientierung an den übergreifenden Ideen und Prinzipien einer Fachdisziplin bzw. Domäne zwingend erforderlich. Geschäftsvorfälle sind in realen Arbeitsumgebungen natürlich beobachtbare Settings; in (schulisch) simulierten Arbeitsumgebungen werden Geschäftsvorfälle als authentische Umgebungen in die Lehr-Lernprozesse implementiert. Der Geschäftsvorfall stellt damit in verschiedenen Arbeits- und Lernumgebungen die der Kompetenzmessung zugrunde liegende situierte Aufgabe bzw. Problemstellung dar.

Die domänenverbundenen und domänenspezifischen Kompetenzbereiche sind über Modellierungsprozesse zerlegbar und werden damit einer detaillierten Erfassung zugänglich. Die verschiedenen Kompetenzbereiche werden unter Berücksichtigung des entwicklungspsychologischen Kompetenzkonzepts von Gree-no und Kollegen (1989) sowie im Hinblick auf den vollständigen Prozess des Mo-

dellierens als Möglichkeiten, Handlungsabfolgen zu systematisieren, in Teilkompetenzen zerlegt: in conceptual, procedural und interpretative competence.

In diesen Teilkompetenzen werden die in der kognitiven Tradition vorzufindenden Unterscheidungen zwischen deklarativem Wissen, prozeduraler Fähigkeit und der Fähigkeit, Anforderungssituationen konzeptual und handlungsleitend zu erfassen, aufgegriffen (konditionales/strategisches Wissen). Die Fähigkeit, spezifische Anforderungen aktiv zu modellieren und kognitiv zu bewerten, ist folglich als eigenständige Kompetenz zu interpretieren.

Die Kompetenzstruktur ist in Anlehnung an berufsreale Anforderungssituationen prozessual: Die gegebene Anforderungssituation ist zunächst in ein konkretes berufliches Problem zu übersetzen. Hierfür ist die Anforderungssituation kognitiv in ihrer Reichweite und im Hinblick auf ihr Lösungspotential zu erfassen (interpretative competence). Dieser Modellierungsschritt begründet die nachfolgenden Bearbeitungsschritte sowie die letztlich gezeigte Performanz, da über das erzeugte Modell der Anforderungssituation die Stimuli und Handlungsspielräume der Anforderungsbewältigung definiert werden. Werden folglich zentrale Elemente der Anforderungssituation nicht erkannt oder fehlerinterpretiert, ist der Proband nicht mehr in der Lage, aus dem konstruierten Modell der Anforderungssituation eine entsprechend gute Lösung zu entwickeln. Dies gilt zunächst einmal modelltheoretisch unabhängig von der Verfügbarkeit deklarativen Wissens und prozeduraler Fähigkeiten. Es ist jedoch unter Rückgriff auf die Ergebnisse der Lehr-Lern-Forschung als evident anzusehen, dass kognitive Repräsentationen und Bewertungen von der Art und dem Umfang deklarativer und prozeduraler Wissensbestände abhängig sind und in diesem Zusammenhang insbesondere ein aktiver Wissenszugriff die Modellbildung der Anforderungssituation stützt (vgl. dazu Kapitel 2.2.2).

Das über das Modell der Anforderungssituation definierte konkrete berufliche Problem wird in einem zweiten Schritt durch aktivierte deklarative (conceptual competence) und prozedurale (procedural competence) Wissensbestände und kognitive Strukturen bearbeitet. Der Erfolg der Bearbeitung ist dabei abhängig von den zur Verfügung stehenden kognitiven Ressourcen (base-level; vgl. Anderson, 1993) und den kognitiven Vermittlungsleistungen, die die Zugriffs- und Verarbeitungsprozesse stützen (contextual-priming sowie Theorien über state-Komponenten in Leistungsprozessen; vgl. Winther, 2006).

Die gefundene Lösung ist in dem sich anschließenden Prozessschritt im Hinblick auf die gegebene Anforderungssituation zu übersetzen. Dieser Schritt dient folglich dazu, sowohl die Lösung als auch das konstruierte Modell der Anforderungssituation zu interpretieren und im Hinblick auf die Ausgangssituation zu validieren (interpretative competence). Hierfür sind erneut konzeptuale Verarbeitungsmodi im Hinblick auf die gegebene Anforderungssituation einzusetzen. Hinzukommen Reflexionen des Lösungsprozesses, mit denen die Anwendbarkeit der Lösung bezogen auf die Anforderungssituation bewertet werden.

Die einzelnen Prozessschritte lassen sich für jede Anforderungssituation als *Handlungsabfolge* identifizieren; sie weisen jedoch in Abhängigkeit zur gegebene-

nen Anforderungssituation unterschiedliche Komplexitätsgrade auf, die die Schwierigkeiten einer Anforderungssituation und damit die Lösungswahrscheinlichkeiten nachhaltig beeinflussen. Die Fähigkeit einer Person in den verschiedenen Kompetenzbereichen lässt sich als Funktion der Schwierigkeitsindizes einer Anforderungssituation darstellen. Je höher der Schwierigkeitsgrad einer Anforderungssituation, desto ausgeprägter muss die Kompetenz des Probanden sein, um die zugrundeliegende Anforderungssituation bewältigen zu können.

Für kaufmännische Leistungsinhalte liegen Beschreibungen dieses Prozesses des Modellierens in Anlehnung an Vorschläge im Bereich der Didaktik der Mathematik (Blum, Neubrand, Ehmke, Senkbeil, Jordan et al., 2004) u. a. für das wirtschaftsinstrumentelle Rechnungswesen (Preiß, 2005) und das bankwirtschaftliche Finanzgeschäft (Winther & Achtenhagen, 2007) vor. Grundidee beider Konzeptionen ist die Abbildung kaufmännisch konzeptueller Kompetenz anhand quantitativer Unternehmensdaten. In allen Bereichen der kaufmännischen Bildung müssen auf Basis realer Belege (Versandlisten, Kontoauszüge, Rechnungen, Registereintragungen etc.) Modellierungen der realen Situation vorgenommen werden. Diese Modellierungen finden ihren Niederschlag beispielsweise über Bestandsveränderungen in den Geschäftsdaten einer Unternehmung und geben Auskunft über unternehmensinterne Prozesse. Die Geschäftsdaten sind der quantitative Beleg für die Initiierung unternehmerischen Entscheidens und Handelns sowie Maß für die Validierung unternehmerischer Aktivitäten.

Ausgangspunkt der Modellierung ist ein in betrieblichen Arbeitskontexten realer und in schulischen Lehr-Lernumgebungen authentisch simulierter Arbeitsprozess mit seinen charakteristischen Merkmalen. Für den Bankenbereich können beispielsweise Kundenorientierung, rechtliche Restriktionen oder Geschäftsfelderpolitik als Merkmale gewählt werden. Diese Merkmale stellen die Außenkriterien für die Modellierung der realen Situation dar (äußere Modellierung). Die Arbeitsprozesse lassen sich in ihre kognitiven Strukturen und Regeln zerlegen. Im kaufmännischen Bereich können in diesem Zusammenhang eine Vielzahl domänenverbundener und domänenspezifischer Kenntnisse und Methoden identifiziert werden, die sich auf unternehmensinterne Geschäftsprozesse beziehen lassen. Diese kontextualisierten Strukturen lassen sich als Innenkriterien der Modellierung definieren (innere Modellierung). Wird exemplarisch aus dem Bankenbereich das Geschäftsfeld der Kontoeröffnung gewählt, können folgende domänenspezifische Kenntnisse als Innenkriterien der Modellierung angenommen werden: (1) deklarative Wissensbestände, die sich auf spezifische Kenntnisse des Produktbereichs Kontokorrent beziehen (Arten, Rechtsgrundlagen, Sonderbedingungen), (2) prozedurale Wissensbestände im Sinne einer funktionalen Systemkontrolle des Produktbereichs Kontokorrent (Verträge auswählen und ausfüllen, Geschäftsabschluss tätigen) sowie (3) die Fähigkeit, Situationen bewerten und in diesen zielbezogen handeln zu können. Im Fall der Kontoeröffnung fallen hierunter z. B. die ökonomische Interpretation eines Kontovertrages aus Bank- und Kundenperspektive.

Diese Strukturierung hinsichtlich einzelner didaktisch und lerntheoretisch ausformulierter Kriterien der Anforderungssituation bringt folgende Vorteile mit sich: Es lassen sich (1) die für die Domäne notwendigen inhaltlichen Systematisierungen durch die Unterscheidung von domänenspezifischen und domänenverbundenen Anforderungsbereichen vornehmen, (2) die kognitiven Prozesse hinsichtlich einer differenzierten Kompetenzstruktur modellieren und (3) die Graduierung der Kompetenzen im Sinne einer Niveaustufung über Approximationen fassen, die auf der Verfügbarkeit von Handlungs- und Verarbeitungsmodellen und dem Zugriff hierauf basieren.

Nachfolgend werden die Entwicklungsschritte der Kompetenzmodellierung sowie Fragen der inhaltlichen Ausgestaltung des Modells diskutiert. Das Kompetenzmodell dient als Hilfsinstrument zur Testentwicklung. Das Ziel ist es, Messinstrumente zu entwickeln, die Kompetenzen im Bereich der kaufmännischen Bildung empirisch erfassbar machen, und mit deren Hilfe auf der Basis eines theoretisch formulierten und empirisch validierten Modells handlungsrelevante Rückmeldungen gegeben werden können.

3 Constructing Measures: Instrumentenentwicklung und deren Gestaltungskriterien

3.1 Assessment-Modelle

Die Herausforderungen, die an Erhebungen im Bereich des Lernens und Leistens gestellt werden, nehmen mit der Entwicklung der technischen und statistischen Methoden kontinuierlich zu (Mislevy, 1994; Beiträge in Wilson, 2004a): Sie sollen das Lernen begleiten, die Entwicklung des Lernenden beschreiben und Prognosen über Lernverläufe stellen. Die Assessmentsituationen werden dabei zunehmend komplexer – dies gilt sowohl im Hinblick auf den Vernetzungsgrad der beobachtbaren Performanz als auch auf den Abstraktionsgrad der zugrundeliegenden Aufgabenstellung. Hierfür sind Abstimmungsprozesse notwendig, die klar definieren, was und wie gemessen werden soll und unter welchen Zielsetzungen eine Messung zu sehen ist.

„One cannot simply construct “good tasks” in isolation, however, and hope that someone down the line will figure out “how to score them.” One must design a complex assessment from the very start around the inferences one wants to make, the observations one needs to ground them, the situations that will evoke those observations, and the chain of reasoning that connects them” (Messick, 1994, zit. n. Mislevy, Almond & Lukas, 2003, p. 1).

Im Folgenden werden exemplarisch drei Beispiele aus der Literatur skizziert, die aufzeigen, wie sich Instrumente zur Messung von Kompetenzen im Sinne eines „Measuring“-Prozess entwickeln lassen und welche (fach)didaktischen Konsequenzen daraus resultieren können: (1) Das Modell der „Assessment-Triangle“ des National Research Council (Pellegrino, Chudowsky & Glaser, 2001), (2) Annahmen eines „Evidence-centered Assessment Designs“ (u. a. Mislevy, 2007; Mislevy & Riconscente, 2006) sowie (3) die „Four Building Blocks“ der Instrumentenentwicklung vor dem Hintergrund probabilistischer Testmodelle (u. a. Wilson, 2008; 2005). Diesen Assessmentmodellen ist gemeinsam, dass die Instrumentenentwicklung auf elaborierten Theorien basiert und nicht allein auf die Konstruktion von Items ausgerichtet ist. Wilson formuliert für die gewählten Beispiele exemplarisch:

„[...] 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, 2005, pp. 3/4).

Mit Hilfe dieser Beispiele wird veranschaulicht, welche Gestaltungskriterien für die Entwicklung von Testinstrumenten zur Erfassung beruflicher Kompetenz

theoretisch auszuarbeiten sind, um Bezugspunkte und Beurteilungsmaßstäbe sowie die praktische Relevanz des Assessments begründen und in ein Messmodell umsetzen zu können. Die zentrale Leitlinie zur Formulierung relevanter Gestaltungskriterien ist der dem Kompetenzkonstrukt innewohnende Kontextbezug. Kompetenzen als „realized abilities“ (Connel, Cheridan & Gardner, 2003, p. 142) beziehen sich demnach immer auf den Umgang und die Bewältigung von spezifischen Anforderungssituationen. Diese Situationen werden aus Domänen bzw. fachbezogenen Leistungsbereichen abgeleitet und bilden den Ausgangspunkt der Instrumentenentwicklung.

Entsprechend der „Standards for educational and psychological testing“ (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999) besteht die Entwicklung eines Tests aus vier Phasen:

„(1) delineation of the purpose(s) of the test and the scope of the domain (content and skills) to be measured; (2) development and evaluation of the test specifications; (3) development, field testing, evaluation, and selection of items and scoring guides and procedures, and (4) assembly and evaluation of the test for operational use“ (ebd., p. 37).

Die im Folgenden dargestellten Assessment Modelle orientieren sich an diesen Prozessstufen; sie wählen jedoch unterschiedliche Abstraktionsgrade im Hinblick auf die Interpretation und die Ausgestaltung der einzelnen Entwicklungsschritte.

3.1.1 Assessment Triangle

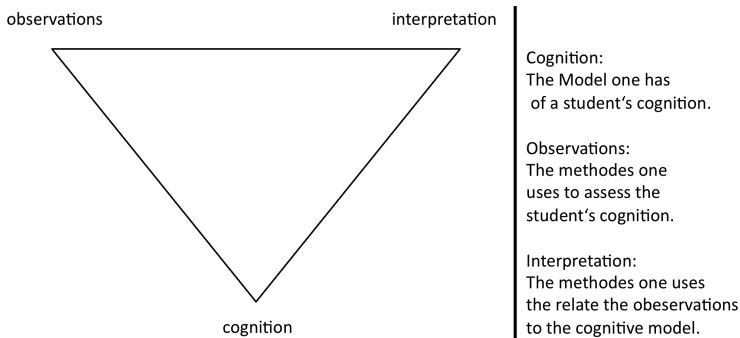
Die Entwicklung eines Instruments zur Erfassung kaufmännischer Kompetenz orientiert sich an der Idee, dass Assessments simultan im Hinblick auf die zu erfassenden individuellen Dispositionen, die zugrundeliegenden Anforderungssituationen und das gewählte Messmodell zu entwickeln sind. Der NRC Report „Knowing What Students Know“ (Pellegrino, Chudowsky & Glaser, 2001) stellt vor diesem Hintergrund ein abstraktes Modell der Assessmententwicklung vor, das gleichsam als Anforderungskatalog an moderne Assessments zu interpretieren ist: Assessments sollten aus drei interagierenden Aspekten bestehen (vgl. Abbildung 3.1).

- Der *kognitive Aspekt* beschreibt aus Perspektive des Testdesigners ein kognitives Modell des Lernens und Leistens, das dem Assessment zugrunde gelegt werden soll; aus Perspektive des Lehrens handelt es sich beim kognitiven Aspekt um Unterrichtsinhalte sowie um Fragen der fachdidaktischen Aufbereitung und methodischen Vermittlung in Lehr-Lernprozessen.
- Mit dem *Aspekt der Beobachtung* werden die Methoden beschrieben, mit denen die Kognitionen der Lernenden gemessen werden können. Hierbei handelt es sich um spezifische Anforderungssituationen oder Aufgaben, die vom Testdesigner unter Bezugnahme auf das angenommene kognitive Modell zu entwickeln sind. Mit Blick in einen Klassenraum oder in eine ent-

sprechende betriebliche Arbeitsumgebung können Beobachtungsmodelle beispielsweise Lernaufgaben, Unterrichtsgespräche, Arbeitsprozesse oder betriebliche Unterweisungen sein.

- Der *Aspekt der Interpretation* befasst sich damit, wie die Ergebnisse der Beobachtung auf das gewählte kognitive Modell zu beziehen sind. Oder anders formuliert, welche Aussagen unter Rückgriff auf die beobachteten Ergebnisse eines Tests (Testscore, Note, Bericht) im Hinblick auf das Leistungsvermögen der Lernenden getroffen werden können.

Abbildung 3.1 Die KWSK Assessment Triangle
(Pellegrino, Chudowsky & Glaser, 2001, p. 44)



Wilson (2004b, pp. 2ff.) weist im Zusammenhang mit der in der Assessment Triangle angesprochenen Kohärenz der Entwicklung, des Einsatzes und der Verwertung von Assessments auf drei zentrale Herausforderungen hin, die den Nutzen von Assessments in Lern- und Arbeitskontexten nachhaltig beeinflussen (vgl. term of *coherence*, Pellegrino, Chudowsky & Glaser, 2001). Er geht im Sinne Black und Wiliams (u. a. 2004) davon aus, dass Assessments – und hier auch solche, die Systemvergleiche anstreben – in erster Linie der Förderung und Stützung von Lern- und Arbeitsprozessen zu dienen haben (vgl. hierzu auch Beiträge in Moss, Pullin, Gee, Haertel & Young, 2008). Unter dieser Bedingung ist eine erste Herausforderung, für jede Art von Assessment Inhalts- und Konstruktvalidität sicherzustellen. Nur unter der Validitätsvoraussetzung lassen sich die Ergebnisse eines Assessments direkt für den Unterricht und/oder für die betrieblichen Arbeitsabläufe nutzen und in die Ausbildung von Lehrpersonen integrieren. Eine zweite Herausforderung stellt die Praktikabilität bzw. Nutzerfreundlichkeit von Assessments dar. Sollen die Ergebnisse eines Assessments direkt für den Unterricht oder betriebliche Abläufe nutzbar sein, müssen sowohl die Auswertungsverfahren als auch die Interpretationen der Ergebnisse transparent und einfach ver-

ständig sein. Den Nutzern von Assessments ist zu vermitteln, wie Ergebnisse erzeugt werden und welche Aussagekraft diese Ergebnisse besitzen (vgl. term of *comprehensiveness*, Pellegrino, Chudowsky & Glaser, 2001). Dies stellt eine Voraussetzung dafür da, dass die Ergebnisse verschiedener Assessments – von Klassenarbeiten bis hin zu internationalen Vergleichsstudien – Eingang in die aktuelle Bildungslandschaft finden. Eng damit verbunden ist die Frage, wie diese verschiedenen Assessments aufeinander bezogen werden können. Eine dritte Herausforderung ist demnach, wie Ergebnisse von Systemvergleichen für schulische und betriebliche Leistungserfassungen genutzt werden können und vice versa (top down-button up Prinzip; vgl. term of *continuity*, Pellegrino, Chudowsky & Glaser, 2001). Die in der Forschung entwickelten Assessmentmodelle sollen hierfür Hilfsinstrumentarien zur Verfügung stellen. Das generelle Prinzip verschiedener Assessmentmodelle (vgl. u. a. Shavelson, Baxter & Pine, 1991, Gustafsson & Undheim, 1996; White & Frederiksen, 1998; Wilson & Sloane, 2000) ist, dass sie für die Entwicklung von Assessments und für die Interpretation von Ergebnissen strukturelle Muster bzw. Beziehungsgefüge anbieten, mit denen sich Assessments unterschiedlichster inhaltlicher Färbung und Zielsetzung miteinander vergleichen lassen (im Detail Kapitel 3.1.2 und 3.1.3).

Die KWSK Assessment Triangle stellt eine (abstrakte) Struktur im Sinne normativer Zielvorgaben zur Verfügung, anhand derer sich domänenbezogene Assessmentmodelle konstruieren lassen. Ein zentrales Element eines Assessmentmodells ist demnach eine konzeptionelle Theorie darüber, wie Lernprozesse funktionieren, wie Wissen erworben und Kompetenzen aufgebaut werden (*kognitiver Aspekt*). Diese Theorie sollte auch Aussagen darüber treffen können, welche Zusammenhänge zwischen Lehr- und Lernprozessen bestehen, wie sich Lerninhalte vermitteln lassen und welche Variablen das Lernen fördern bzw. behindern können. Im Ergebnis sollte auf Basis dieser Heuristik ein Modell entworfen werden können, das kognitive Prozesse domänenunabhängig definiert und in Beziehung zu den spezifischen Inhalten, Arbeitstechniken und Zielen einer Domäne setzen kann. Mit der Entwicklung nationaler Bildungsstandards werden diese Überlegungen beispielsweise genutzt, um über eine Auflistung von Prozessen und Zielen in unterschiedlichen Fachbereichen zu einer Vergleichbarkeit der kognitiven Anforderungen zu gelangen (u. a. Department of Education and Science, 1987; Australian Education Council, 1992; KMK, 2003).

Assessmentmodelle sollten darüber hinaus den Zusammenhang zwischen kognitiven Prozessen und den Möglichkeiten der Testung dieser Prozesse verdeutlichen (*Aspekt der Beobachtung*). Hierzu sind Arten von Assessments (Classroom Assessment, Large-Scale-Assessment), Präsentationsformate (paper-pencil-Test, computerbasiertes Testen) und Auswertungsverfahren (Messmodelle und Testtheorie) mit den Zielen des Assessments und mit den zentralen fachlichen Inhalten abzustimmen.

Diese Analysen sind insbesondere vor dem Hintergrund komplexer Assessments, mit denen differenzierte Wissens- und Handlungspotentiale erfasst werden sollen, von Bedeutung. Insbesondere die Entwicklung von Testformaten und

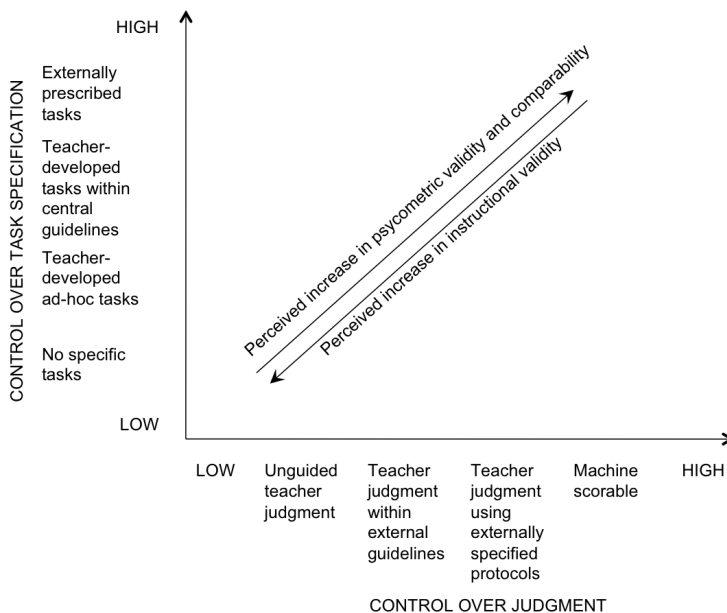
Testitems steht dabei in dem Konflikt zwischen psychometrischer Validität und Vergleichbarkeit und der instruktionalen Validität (Abbildung 3.2). In der Literatur wird dieser Konflikt in der Regel über den Vergleich von standardisierten Multiple-Choice-Tests und alternativen Assessmentformen (vgl. u. a. performance assessment: Shavelson, Ruiz-Primo & Wiley, 1999; portfolio assessment: Wolf, 1989; computer simulation: Baxter, 1999) dargestellt.

In Abstimmung mit den spezifischen Inhalten, Arbeitstechniken und Zielen einer Domäne ist zu prüfen, welche Testformate geeignet sind und wie sich ein Assessment aus verschiedenen Testformaten möglichst optimal zusammensetzt.

„The use of diverse assessment modes, then, is an attempt to use information from a variety of different forms of assessment [...] so that we can capitalize on the opportunities that some assessments might provide for enhancing validity as well as the opportunities that others might provide for increasing reliability“ (Wilson, 2004b, p. 11).

Letztlich ist in ein Assessmentmodell eine Vorstellung darüber zu implementieren, wie sich die über Testformate gesammelten Daten nutzen lassen (*Aspekt der Interpretation*).

Abbildung 3.2 Testformate und Validität (Wilson, 2004b, p. 10)



Dies beinhaltet Modelle und Vorschriften der Datenaufbereitung (Scoring-Guidelines, Rater-Manualen), Annahmen über die Zusammenfassung von Daten im Rahmen der statistischen Analyse und die Richtlinien zur Aufbereitung der Ergebnisse sowie deren Verbreitung. Die aktuelle Literatur verweist hier auf die Vorteile von Item-Response-Modellen, um die Beziehungen zwischen beobachteten Variablen und kriteriumorientiertem Output messtheoretisch zu begründen (vgl. Beispiele in Pellegrino, Chudowsky & Glaser, 2001; im Detail Kapitel 4.2).

In den zwei nachfolgend beschriebenen Assessmentmodellen sind die Annahmen der KWSK Assessment Triangle verankert. In beiden Modellen werden ausgehend von einer kognitiven Theorie Entwicklungsschritte der Assessmentkonstruktion veranschaulicht.

3.1.2 *Evidence-centered Assessment Design (ECD)*

In der Argumentation eines evidence-centered Assessment Designs ist die Assessmententwicklung eng gekoppelt an die Testvalidation. Die Prozessstufen werden in diesem Zusammenhang als (mathematisch-statistische) Beweisführung interpretiert.

„Evidence-centered design views an assessment as an evidentiary argument: An argument from what we observe students say, do, or make in a few particular circumstances, to inferences about what they know, can do, or have accomplished more generally“ (Mislevy & Haertel, 2006, p. 7).

Evidence-centered Designs (ECD) setzen sich aus fünf Stufen bzw. Schichten („layer“) zusammen: (1) die Analyse der Domäne, (2) die Modellierung der Domäne, (3) eine konzeptionelle Theorie der Assessmentkonstruktion, (4) die Konstruktion des Assessments im Sinne einer Pilotierung und (5) der Einsatz des Assessments als Produkt (vgl. u. a. Mislevy & Riconscente, 2005).

Die Zerlegung eines Assessments in einzelne Schichten soll sowohl die Entwicklung neuer Assessments unterstützen als auch eine Analyse und Bewertung bereits bestehender Assessments kriterienorientiert leiten. Hierbei wird von den Autoren betont, dass ECD-Modelle für jede Domäne und für jedes denkbare Assessmentformat anwendbar sind. Dies ist möglich, da die Entwicklung bzw. Bewertung eines Assessments an verschiedenen Repräsentationsformen innerhalb der Beweiskette orientiert ist („knowledge representation“), die für jede Schicht charakteristisch beschrieben werden können. ECD-Modelle bieten vor diesem Hintergrund einen technischen auf die Beweisführung ausgelegten Rahmen, um unabhängig von einer konkreten Domäne, die Arten und Erscheinungsformen der Wissensrepräsentation strukturieren zu können (vgl. Mislevy, 2008). Wissensrepräsentationen werden als Output eines jeden Analyseschrittes interpretiert; sie stellen das Beweismaterial eines geschlossenen Assessmentprozesses dar und sind damit die Produkte des Assessments. Für einen geschlossenen Assessmentprozess müssen alle Schichten in sich stimmig konstruiert sein und in ihrer Inter-

ferenz die unterschiedlichen konzeptionellen, strukturellen und ausführenden Stufen der Instrumentenentwicklung berücksichtigen.

Abbildung 3.3 ECD-Schichten (in Anlehnung an Mislevy & Riconscente, 2005, p. 6)

ECD-Schicht	Interpretation	Merkmale	Repräsentationsformen
Analyse der Domäne („domain analysis“)	Informations-sammlung der Domänenspezifika, einschließlich Informationen darüber, wie die Domänenmerkmale erlernt und vermittelt werden	Konzepte, Terminologien, Methoden und spezifische Repräsentationsformen	Formeln, Symbole, MindMaps, Lehrpläne, technische Modelle
Modellierung der Domäne („domain modeling“)	Narrative Struktur der Domänen-spezifika (zentrale Ideen, typische Arbeits- und Lernprozesse)	Wissen, Fertigkeiten, Fähigkeiten, mögliche Arbeitsprodukte oder Beobachtungsergebnisse	Modell der Domäne (Beziehungen, Argumentationen, Bezugspunkte)
Theorie der Assessmentkonstruktion („conceptual assessment framework“)	Modellformulierung zur Konstruktion von Assessments (Blue Print)	(latente) Variablen, mögliche Messmodelle sowie Spezifikationen des Testaufbaus	Mathematische und / oder graphische Repräsentation des Messmodells
Assessmentkonstruktion („assessment implementation“)	Erstellung des Assessments (Itemkonstruktion)	Aufgabenmaterial (einschließlich aller notwendigen Informationen zur Aufgabenbewältigung)	Itembeschreibungen, Präsentationsregeln, Kodierungen, Scoringverfahren
Einsatz des Assessments („assessment delivery“)	Operativer Assessment-durchlauf und Reporting	Aufgaben, Arbeitsprodukte und das entsprechende Antwortverhalten, Testscores	Numerische und graphische Zusammenfassung des Assessments, Testgüte, Reports auf unterschiedlichen Analyseebenen

Abbildung 3.3 gibt einen Überblick über die ECD-Schichten einschließlich einiger Beispiele für Repräsentationsformen als Outputs der Beweiskette (Mislevy & Riconscente, 2005, p. 6). Die Beziehungen zwischen den einzelnen Schichten werden über die Outputs der Schichten inhaltlich fixiert, d. h., dass über die Art und den Umfang der Repräsentation einer Schicht auf den Output der folgenden Schichten Einfluss genommen wird. Die getroffenen Annahmen und gewählten Entscheidungen bei der Analyse der Domäne wirken folglich auf die Wahl des Messmodells und auf die Verfahren der Testanalyse und der Testinterpretation.

Analyse der Domäne (domain analysis)

Die Analyse der Domäne dient der Filterung relevanter Informationen aus einer spezifischen Domäne, die in Beziehung zu dem angestrebten Assessment gesetzt

werden können. Die Art der zu filternden Informationen ist dabei abhängig von den Zielsetzungen des Assessments. Typische Informationen sind Inhalte, Konzepte, Theorien, Terminologien, Strategien, Arbeitsmethoden und Repräsentationsformen einer Domäne. Charakteristisch ist, dass sich diese Informationen in domänenspezifischen Anforderungssituation (in Lern- oder Arbeitskontexten) beobachten lassen, ihre Interpretation jedoch von den Vorstellungen, wie Wissen erworben und vermittelt wird, und in diesem Zusammenhang auch von der Vorstellung, wie sich Kompetenzen definieren und entwickeln, abhängig ist.

„As the first stage in assessment design, Domain Analysis leads us to understand the knowledge people use in a domain, the representational forms, characteristics of good work, and features of situations that evoke the use of valued knowledge, procedures, and strategies“ (Mislevy & Haertel, 2006, p. 7).

Greeno, Collins und Resnick (1997; vgl. auch Greeno, 1998) betonen, dass unterschiedliche *(lern-)psychologisch determinierte Domänenzugänge* zu verschiedenen Domänensichtweisen führen (müssen). Unter Bezugnahme auf mathematische Inhalte unterscheiden die Autoren zwischen behavioralen, kognitiven und soziokulturellen Zugängen: Während aus behavioraler Perspektive Aspekte des Lösungsvorgangs verschiedener mathematischer Probleme im Sinne eines Stimulus-Verhaltens-Musters von zentraler Bedeutung sind, wird unter der kognitiven Perspektive auf Informationsverarbeitungsprozesse fokussiert und damit auf die Fragen, wie Wissen erworben und wie auf Wissensrepräsentationen zugegriffen wird. Aus soziokultureller Perspektive ist entscheidend, wie und mit welchem Umfang mathematische Kenntnisse und Fertigkeiten zu einer aktiven gesellschaftlichen Teilnahme in unterschiedlichen Gemeinschaften beitragen können. Testsituationen werden konstruiert, um mit ihnen hinreichend genaue Informationen erfassen zu können – dies geschieht vor dem Hintergrund des gewählten Domänenzugangs. Ein Testentwickler entscheidet folglich aus (lern-)psychologischer Perspektive über die Art der Aufgabenstellung und folglich auch über die Inhalte der Evaluation und des Reportings. Verschiedene Assessments innerhalb derselben Domäne sind unabhängig von dem eigentlichen Assessmentziel zwingend die Folge.

Neben (lern-)psychologisch differenzierenden Sichtweisen führen unterschiedliche *fachlich determinierte Domänenzugänge* zu wechselnden Vorstellungen über die Domäne und folglich zu verschiedenen Assessments. So geht Achtenhagen (2004) davon aus, dass in beruflichen Situationen unterschiedliche fachliche Zugänge integriert sind. Sloane und Dilger (2005) beschreiben beispielsweise unter Rückgriff auf Getsch und Preiß (2003a; 2003b) die Integration von mathematischem Regelwissen und ökonomischem Begründungswissen im Rahmen einer Deckungsbeitragsrechnung. Von Mathematiklehrern wird eine Deckungsbeitragsrechnung als praktischer Anwendungsfall für den Mathematikunterricht gesehen; für Ökonomielehrer ist die mathematische Lösung ein Hilfs-

mittel, dem eine inhaltliche, ökonomisch sinnvolle Interpretation zu folgen hat. In einem weiteren Beispiel wird auf betriebswirtschaftliche Teilgebiete verwiesen, deren Zugänge selbst in der zugrundeliegenden Bezugswissenschaft kontrovers diskutiert werden. Marketing z. B. kann folglich mathematisch-empirisch oder sprachlich profiliert sein, was eine unterschiedliche Fachinterpretation zur Folge hat (Sloane & Dilger, 2005, S. 16).

Die zentrale Frage in diesem Zusammenhang ist, welches die Leitideen einer Domäne sind. Um folglich die innere Struktur einer Domäne zu erfassen, sollte im Sinne einer geschlossenen Beweisführung auf beobachtbare Außenkriterien zurückgegriffen werden. Beispiele hierfür finden sich im Rahmen der „structured outcomes of learning“-Taxonomie (Biggs & Collis, 1982) oder aktuell im Zuge der Entwicklung nationaler Bildungsstandards (u. a. KMK, 2003).

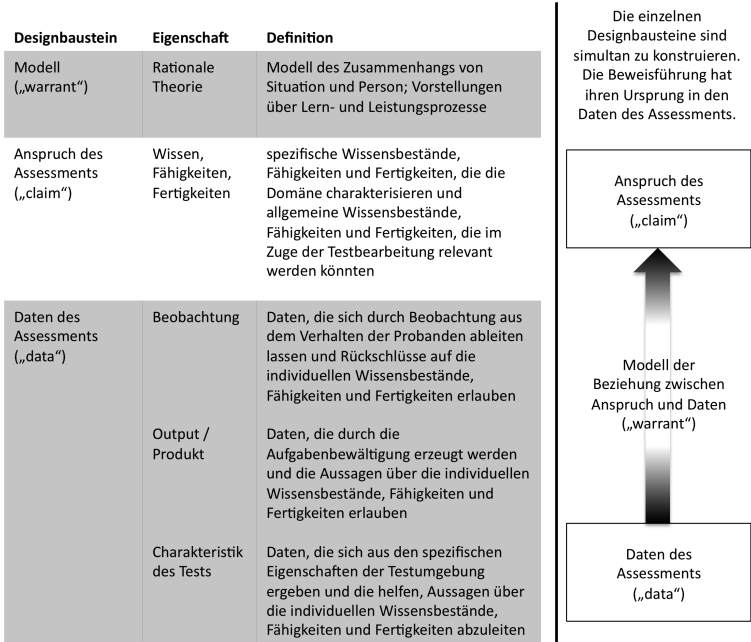
Modellierung der Domäne (domain modeling)

Die Modellierung der Domäne dient zunächst einmal der Strukturierung der in der Domänenanalyse gesammelten Informationen. ECD-Modelle bieten hierfür standardisierte Schemata („design pattern“) an, mit denen Inhalte verschiedener Domänen analysiert und für ein Assessment aufbereitet werden können. Die Grundlage dieser strukturellen Ordnung bilden in Anlehnung an Toulmin (1958) drei Designbausteine, die sich eindeutig auf ein konkretes Assessment beziehen lassen und für dieses Assessment simultan zu konstruieren sind (vgl. Abbildung 3.4):

- Baustein 1: Der Anspruch des Assessments („claim“) ist vor dem Hintergrund möglicher Aufgabenformate und der durch diese Stimuli erzeugten Verhaltensmuster zu formulieren. Der Anspruch bezieht sich insbesondere auf die Frage, wie Wissen, Fähigkeiten und Fertigkeiten einer spezifischen Domäne (Informationen der Analyse der Domäne) zu strukturieren sind. Hierbei sind sowohl die als domänenspezifisch identifizierten Wissensbestände, Fähigkeiten und Fertigkeiten zu strukturieren, als auch jene, die für die Bewältigung der Testsituationen eine zentrale Rolle spielen können. Hierunter fallen insbesondere lernstrategische Ressourcen sowie Heuristiken der Problemlösung.
- Baustein 2: Die Aussagen über die Strukturierung der kognitiven Ressourcen basieren auf Daten („data“). Diese Daten sind Outputs des Assessments, die sich einerseits auf das Antwortverhalten oder das Aktionsverhalten von Probanden während des Assessments beziehen und andererseits Informationen aus den einzelnen Testumgebungen und Aufgabenstellungen enthalten können. Der Anspruch des Assessments ist damit abhängig von der Datenqualität. Die zu erfassenden Daten müssen folglich so beschaffen sein, dass der angestrebte Assessmentanspruch auch erfüllt werden kann. Die Daten können ihren Ursprung in manifesten oder latenten Variablen haben, mit deren Hilfe (generalisierbare) Aussagen über individuelle Wissensbestände, Fähigkeiten und Fertigkeiten getroffen werden sollen. Als Datenquellen stehen die direkte Beobachtung, die Aufgabenlösungen als Output

bzw. Produkt des Assessments und die Charakteristik der Aufgaben einschließlich derer Eigenschaften zur Verfügung (vgl. insbesondere Mislevy, 2007).

Abbildung 3.4 Schemata zur Modellierung der Domäne (vgl. Mislevy & Haertel, 2006)



- Baustein 3: Die Beziehung zwischen den Daten des Assessments und seinem Anspruch basiert auf theoretischen Modellvorstellungen („warrant“) des Zusammenhangs von (Test-)Situation und (Test-)Leistung. Die Problematik besteht darin, dass unterschiedliche Daten produziert werden und damit unterschiedliche Modellrestriktionen in die Beweisführung einfließen.

Im Zentrum der Modellierung der Domäne stehen Vorstellungen über spezifische Wissensbestände, Fähigkeiten und Fertigkeiten sowie über charakteristische Handlungsweisen in domänentypischen Anforderungssituationen:

„Centered around some aspect of KSAs [knowledge, skills, and attributes], a design pattern is meant to offer a variety of approaches that can be used to get evidence about that knowledge or skill, organized in such a

way as to lead toward the more technical work of designing particular tasks“ (Mislevy & Riconscente, 2005, p. 12).

Für einen geschlossenen Assessmentprozess ist zentral, wie sich diese Vorstellungen in Aufgabenstellungen/Anforderungssituationen übersetzen lassen, mit denen schließlich Daten erzeugt werden können, die sich stimmig unter Rückgriff auf ein theoretisch rationales Modell von Lernen und Leisten auf den Anspruch des Assessments beziehen lassen. Dieser Zusammenhang zwischen erzeugten Daten und angestrebter Aussagekraft des Assessments wird in den ECD-Modellen durch eine konzeptionelle Theorie der Assessmentkonstruktion begründet. Im Gegensatz zur Modellierung der Domäne, die bewusst weit gefasst ist und ohne technische Details auskommt, werden im Rahmen der Assessmentkonstruktion konkrete technische Spezifikationen – Messmodelle, Repräsentationsformate – zum Gegenstand der Theorie.

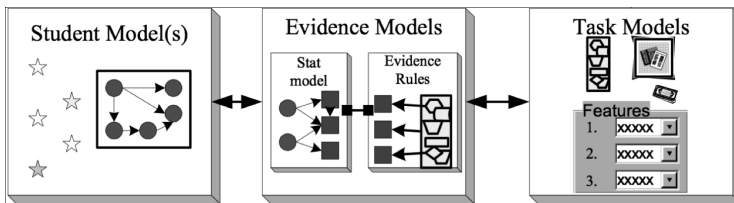
Theorie der Assessmentkonstruktion (conceptual assessment framework)

Die Sammlungen und Strukturierungen aus der Analyse und Modellierung der Domäne werden in Teilmodelle und Prozesse eingebettet, mit deren Hilfe Operationalisierungen von domänentypischen Konstrukten und deren konkrete Umsetzung in Assessments erarbeitet werden können. Hierzu werden alle notwendigen statistischen, operationalen und heuristischen Prozesse aufeinander abgestimmt: Welche statistischen Modelle sollen genutzt werden? Wie werden die Anforderungssituationen des Assessments aussehen? Wie werden die Ergebnisse des Assessments kodiert und gescort? Welche Aussagen lassen sich aus den Testergebnissen ableiten? Fragen wie diese charakterisieren ein Assessment. Um diese Fragen in einen theoretisch begründeten Zusammenhang zu stellen, werden drei Modelle entwickelt, die die zentralen Komponenten des Assessments auf inhaltlicher, methodischer, didaktischer und empirischer Ebene synchronisieren (vgl. Abbildung 3.5): Student Model, Task Model und Evidence Model.

Die über den Anspruch des Assessments strukturierten domänentypischen Wissensbestände, Fähigkeiten und Fertigkeiten werden mit Hilfe des „*Student Model*“ operationalisiert. Hierzu ist zu klären, was genau über das Assessment gemessen werden soll und wie die Beziehung zwischen dem „analytischen“ Testergebnis (θ) und der „realen“ Fähigkeit des Probanden ($\theta \pm |x|$) zu modellieren ist. Die Komplexität des Modells ist dabei abhängig von (1) der Anzahl der Fähigkeitsparameter, die über das Assessment abgebildet werden sollen, und von (2) der Anzahl der intermediierenden Umwelt- und Personendeterminanten. Es sei zur Illustration ein einfacher Fall angenommen: Das Assessment soll Auskunft über eine spezifische Fähigkeit geben – die Fähigkeit eines Lernenden, grammatikalische Gesetzmäßigkeiten zu erkennen, eine quadratische Gleichungen zu lösen oder im Bereich der kaufmännischen Bildung eine Lieferantenauswahl mathematisch begründet durchzuführen („claim“; Mislevy, Almond & Lukas, 2003). Das Antwortverhalten in der Testsituation gibt Auskunft über die Fähigkeit eines Lernenden, diese konkrete Testsituation (grammatikalische Gesetzmäßigkeit,

quadratische Gleichung, quantitative Lieferantenauswahl) tendenziell zu bewältigen oder in dieser zu scheitern (θ). Das Testergebnis kann jedoch durch verschiedene Personen- und Umweltdeterminanten beeinflusst sein ($\theta \pm |x|$). Fehlinterpretationen der Aufgabenstellung, Ablenkungen durch Mitschüler sowie motivationale Defizite sind möglich. Je mehr Sicherheit über diese Intermediate besteht, umso zuverlässiger kann die generelle Fähigkeit, beispielsweise eine Lieferantenauswahl durchzuführen, geschätzt werden. Diese Schätzungen basieren auf Wahrscheinlichkeitsannahmen und sind in den verschiedenen Testtheorien über das Bayes-Theorem verortet (z. B. Messfehlermodell der Klassischen Testtheorie; vgl. u. a. Mislevy, 2007). Mit der Entscheidung für eine Testtheorie wird im „Student Model“ auch der mathematisch definierte Zusammenhang zwischen der Fähigkeit einer Person, eine Anforderungssituation zu bewältigen (θ_i), und dem Schwierigkeitsniveau dieser Anforderungssituation (δ_j) festgelegt (im Detail Kapitel 4.2).

Abbildung 3.5 Modelle der Itemkonstruktion (Mislevy & Haertel, 2006, p. 13)



Die Spezifika der Anforderungssituation werden über das „Task Model“ operationalisiert. Hier werden Entscheidungen über das Testformat, die Teststimuli sowie die Repräsentationsform des Tests getroffen.

„Task models describe the presentation material that is presented to the examinee and the work products generated in response. They also contain task model variables that describe features of tasks as well as how those features are related to the presentation material and work products“ (Mislevy, Almond & Lukas, 2003, p. 10).

Insbesondere die Forschungen zum formativen Assessment haben integrierte, in das Lernen und Arbeiten eingebundene Formen des Assessments weiterentwickelt („embedded assessment“, Black, Harrison, Lee, Marshall & Wiliam, 2003; Black & Wiliam, 2004; „in-basket assessment“, Shavelson, Young, Ayala, Furtak, Ruiz-Primo, Tomita & Yin, 2008; Shavelson, Baxter & Pine, 1992). Die zentrale Idee ist, Assessments als „natürlichen“ Bestandteil in Lern- und Arbeitsprozesse zu implementieren, um den Einfluss externer Störfaktoren möglichst gering zu halten. Die Assessmentformate sollten vor diesem Hintergrund domänentypisch

schen Lern- und Arbeitsweisen ähneln. Für Assessments im Rahmen der beruflichen Bildung in Deutschland bedeutet dies zum einen die Abstimmung der Assessmentinhalte zwischen den verschiedenen Lernorten sowie zum anderen eine Konzentration der Assessmentformate auf handlungsbasierte Anforderungssituationen, da der Aktionsraum der Lernenden bedingt durch den Lern- und Arbeitsort Betrieb weit umfassender ist, als man ihn in schulischen Umgebungen vorfinden kann – dies schließt Interaktionen mit Kollegen und Kunden sowie die Abstimmung von Arbeitsprozessen über Abteilungen hinweg mit ein. Vergleichbares gilt für die Assessmentstimuli: Die Arbeitsmaterialien der Assessments sollten möglichst authentische, an domänentypischen Lern- und Arbeitsprozessen orientierte Anreize geben. Für die kaufmännische berufliche Bildung bieten sich in diesem Zusammenhang technologie-basierte Assessments an, die typische Unternehmenssoftware, authentische Arbeits- und Geschäftsprozesse sowie branchenabhängige Markt- und Umwelteinflüsse simulieren (vgl. Winther & Achtenhagen, 2009a).

„Student Model“ und „Task Model“ werden mit Hilfe des „Evidence Model“ aufeinander bezogen. Es enthält Hinweise darauf, wie die in den Anforderungssituationen gezeigte Performance der Probanden gemessen werden soll. Hierfür sind zwei Modellkomponenten vorgesehen: eine Evaluationskomponente sowie die Entscheidung für ein spezifisches Messmodell. Die Evaluation bezieht sich auf die Identifizierung von Fähigkeitsindizes auf Basis der in den Anforderungssituationen beobachtbaren Variablen durch Kodierungs- und Scoringvorschriften. Die Evaluationskomponente gibt damit vor, wie automatische Scoringverfahren zu gestalten oder wie Manuale und Scoringguides aufzubereiten und in entsprechenden Schulungen zu vermitteln wären. Insbesondere bei komplexen Assessments spielen Evaluationsregeln und -prozeduren eine entscheidende Rolle, da der Output der Anforderungssituation unter Rückgriff auf verschiedene Fähigkeitsdimensionen und Zielstellungen interpretiert werden kann und sich damit aus einer Testsituation verschiedene Daten erzeugen lassen. Um die über Evaluationsprozesse erzeugten Daten im Hinblick auf den angestrebten Assessmentanspruch zu identifizieren und zu quantifizieren, ist ein Messmodell zu hinterlegen. Hierbei handelt es sich in der Regel um psychometrische Modelle, mit denen die generierten Daten zusammenzufassen und zu beschreiben sind.

Psychometrische Modelle, unabhängig davon, ob es sich um Modelle der klassischen Testtheorie oder der Item Response Theorie handelt, dienen dazu, die in Anforderungssituationen beobachtbaren Variablen in interpretierbare Outputs zu übersetzen, indem mathematische Regeln darüber aufgestellt werden, wie sich die Wahrscheinlichkeiten, eine konkrete Anforderungssituation zu bewältigen oder in dieser zu scheitern ($p(X_i)$), auf die Fähigkeit des Lernenden (θ_i) beziehen lassen oder vice versa.

Konstruktion des Assessments (assessment implementation)

Die Konstruktion des Assessments implementiert alle zuvor durchlaufenen Schichten des Assessmentprozesses. Hier geht es darum, die gesammelten Infor-

mationen und theoretischen Modellvorstellungen in ein anwendbares Assessmentprodukt umzusetzen. Dazu gehören die Entwicklung und Zusammenstellung der Items einschließlich aller hierfür benötigten Materialien, die Überprüfung des Messmodells, die Vorbereitung von Scoringverfahren sowie die Erprobung von Durchführungstechniken. Die theoretischen Vorarbeiten dienen dazu, Items inhaltlich differenziert zu akzentuieren sowie eine Variation des Anspruchsniveaus zu erreichen. Hierfür werden für inhaltlich ähnliche Testbereiche Designhilfen konstruiert, mit deren Unterstützung Itembündel für domänencharakteristische Anforderungssituationen entwickelt werden können.

Bei der Konstruktion der Items ist zwischen psychometrischer und instruktionaler Validität abzuwägen (vgl. Abbildung 3.2). Items mit hoher instruktionaler Validität zeigen eine starke Adaptation im Hinblick auf die domäentypischen Lern- und Arbeitsbedingungen, jedoch eine geringere Reliabilität als standardisierte Testformate.

Einsatz des Assessments (assessment delivery)

Die operative Schicht des Assessmentprozesses basiert auf systematischen Prozeduren, die darüber Auskunft geben, wie das Assessment einzusetzen ist und welche Effekte der Einsatz des Assessments zeigt. Eine detaillierte Planung macht insbesondere vor dem Hintergrund zunehmend komplexerer Assessments, die verschiedene Itemformate aufweisen und verschiedene Qualitäten von Daten erzeugen, Sinn.

„Any assessment requires some processes by which items are actually selected and administered, scores are reported, and feedback is communicated to the appropriate parties. Operational processes may differ substantially from one assessment to another, and even within a given assessment system the processes may evolve over time as needs arise“ (Mislevy & Riconscente, 2004, p. 81).

Ein idealtypischer Assessmentdurchlauf lässt sich über vier Phasen beschreiben (u. a. Mislevy & Riconscente, 2005): In der *Phase der Itemauswahl* („activity selection process“) wird über ein automatisiertes System (Zufall oder Itemplan) oder manuell durch den Tester ein Item aus dem Assessment gewählt, das dem Probanden vorgelegt wird. Jedes Item eines Tests unterliegt zuvor konstruierten Präsentationsregeln, die der Sicherung der Durchführungsobjektivität dienen. Hierzu zählen Itemattribute wie die Testzeit, die Darstellungsform oder eine Aufzählung der zur Verfügung stehenden Hilfsmittel (Taschenrechner, ergänzende Literatur). Diese Informationen sind dem „Task Model“ zu entnehmen. Während der *Präsentationsphase* („presentation process“) können erstmals Informationen darüber gesammelt werden, wie Probanden mit dem Item interagieren, wie die konstruierte Testsituation wahrgenommen wird und in welchem Umfang das im „Student Model“ theoretisch modellierte Antwortverhalten im Praxiseinsatz eintritt. Die Antworten werden in der *Phase der Beweisidentifikation* („evidence iden-

tification process“) entsprechend der Annahmen des „Evidence Model“ evaluiert, um Aussagen über Outcomes des Items oder die Ergebnisse der Anforderungssituation in Folge des Scorings treffen zu können. In der *Phase der Beweisbildung* („evidence accumulation process“) werden schließlich die gescorten Variablen mit Hilfe des gewählten Messmodells zusammengefasst und beschrieben, indem Rückschlüsse auf die Personenfähigkeiten gezogen und die Itemparameter geschätzt werden. Vor dem Hintergrund adaptiven Testens können diese zusammenfassenden Informationen genutzt werden, um in der Phase der Itemauswahl die Anforderungssituationen zu wählen, die dem Leistungspotential des Probanden entspricht (vgl. hierzu im Detail Almond & Mislevy, 1999).

Das evidence-centered Assessment Design (ECD) gibt Testentwicklern Hinweise darauf, welche wechselseitigen Beziehungen zwischen Aspekten der kognitiven Psychologie, den inhaltlichen und technologischen Spielräumen bei der Ausgestaltung von Assessments sowie den empirischen Modellen zur Absicherung eines Assessmentziels gegeben sind. Es bietet strukturierte Schemata für die Testentwicklung an, anhand derer Assessments in verschiedenen Domänen vorbereitet und in ihrer Gestaltung theoriebasiert begründet werden können. Ein besonderes Leistungsmerkmal der ECD-Modelle ist ihre hohe Adaptationsfähigkeit. Sie bieten einerseits die Möglichkeiten, konzeptuelle Forschungsarbeiten zur Testentwicklung mit denen praktischer Testmethoden zu verbinden, andererseits werden den Testentwicklern wertfreie Analysemuster für bestehende und zukünftig zu entwickelnde Assessments an die Hand gegeben.

3.1.3 *Four Building Blocks*

In Übereinstimmung mit den Ergebnissen des National Research Council (NRC; Pellegrino, Chudowsky & Glaser, 2001) sind Assessments für Lehr-Lernprozesse auf Basis (kognitiver) Theorien zu entwickeln, um *begründete* und *konsistente* Erhebungen zu planen, durchzuführen, zu evaluieren und für Lehr-Lernprozesse und bildungsorientierte Entscheidungen nutzbar zu machen: Das Adjektiv „begründet“ bezieht sich in diesem Zusammenhang auf die Formulierung von Theorien über (1) den Gegenstand der Messung (Konstrukt), (2) die Entwicklung von Items bzw. Aufgaben auf Basis des Konstrukts (Design) und (3) die Beziehung zwischen Konstrukt und Testformat bzw. zwischen den latenten und manifesten Variablen (Messmodell); „konsistent“ bedeutet einheitlich im Rahmen der gewählten (kognitiven) Theorie und nachhaltig im Hinblick auf die Interpretation und die Verwertung der Ergebnisse. Die Arbeiten Wilsons (u. a. 2004a; 2004b; 2005; 2008) im Rahmen des Berkeley Evaluation and Assessment (BEAR)-Systems zeigen auf, welche Entwicklungsschritte für eine begründete und konsistente Messung gegangen werden sollten. Diese Entwicklungsschritte sind in den „Four Building Blocks“ konzeptionell erarbeitet. Mit den „Four Building Blocks“ werden für die Instrumentenentwicklung Kriterien zur Verfügung gestellt, mit deren Hilfe einerseits Theorie und Instrument kausal verknüpft werden können; andererseits werden Strukturen und Techniken aufgezeigt, wie auf Basis der Da-

ten aus dem Instrument auf das zu messende theoretische Konstrukt rückgeschlossen werden kann (Wilson, 2005, p. 12).

Abbildung von Kausalität

Die zentrale Ausgangsfrage der Instrumentenentwicklung ist stets die nach dem Ziel und damit nach dem Gegenstand der Messung. Das Ziel der Messung ist objektbezogen und benötigt als Anker einen spezifischen Kontext: Beispiele hierfür sind die Fähigkeit des Textverstehens oder die ökonomische Interpretation betriebswirtschaftlicher Konzepte. Diese Konzepte können allgemeiner formuliert sein (Gewinnkonzept, Kosten- und Ertragskonzept) oder auf spezifische betriebliche Teilbereiche abstellen (Deckungsbeitrag, Fertigungsgemeinkosten). Zentral aber ist jedoch, dass für eine Messung dieser Beispiele eine theoretische Vorstellung über die Zusammenhänge zwischen Aufgabe und Antwortverhalten bestehen muss. Wie muss eine Aufgabe konstruiert sein, um die gesuchte Fähigkeit abbilden zu können und wie lässt sich diese Fähigkeit wissenschaftlich definieren? Antworten auf diese Fragen können mit Hilfe eines kognitiven Modells gefunden werden. Dieses Modell kann ein theoretisches kognitives Konstrukt sein, das beispielsweise den Aufbau von Wissen in einer spezifischen Domäne beschreibt. Vorstellbar sind auch theoretische Konstrukte, die als Bestandteil einer soziologischen oder gesellschaftskritischen Theorie beispielsweise die Stabilität von Einstellungen und Meinungen charakterisieren. Ausschlaggebend ist, dass sich das gewählte Konstrukt (a) über die zu messende Fähigkeit und (b) über den Kontext, in dem die Messung stattfinden soll, definieren lässt.

Im Rahmen der „Four Building Blocks“ werden Konstrukte der Messung im *Block I: Construct Map* systematisch ausdifferenziert. Für die Entwicklung einer Construct Map wird das zu definierende Konstrukt auf einer kontinuierlichen Skala angeordnet: Die Fähigkeit zum Textverstehen ist stark oder schwach ausgeprägt; die ökonomische Interpretation ist fachlich sinnvoll oder falsch. Construct Maps beschreiben demnach spezifische Konstrukte unter Berücksichtigung von verschiedenen qualitativen Leveln. Damit werden zwei Dinge parallel erreicht. Zum einen werden die Grenzen (Extremwerte) des Konstrukts festgelegt und zum anderen wird das Konstrukt technisch in eine eindimensionale latente Variable transformiert. Dies sind Eigenschaften, die sowohl für die Konstruktion von Items (Aufgaben/Anforderungssituationen) zur Messung des Konstrukts als auch für die mathematische Beziehung zwischen Konstrukt und Antwortverhalten in den konstruierten Anforderungssituationen von zentraler Bedeutung sind. Das zentrale verbindende Element zwischen der Construct Map und der Entwicklung von Items ist der Kontext der Messung:

„[...] the task should be a sampling from those [content; Anm. d. Verf.] typically used across the range of contexts where the assessments are taking place“ (Wilson & Draney, 2004, p. 138).

Der *Block II: Item Design* beschreibt die Zusammenstellung von Items auf Basis der Construct Map (causality). Für schulische Assessments geben Wilson und Kollegen zwei Zielsetzungen für die Konstruktion von Items vor: Sie sollen sich (1) am Curriculum orientieren und (2) die instruktionale Praxis reflektieren (u. a. Wilson, 2005, p. 42; Wilson & Draney, 2004, p. 134). Hierfür sind verschiedene Repräsentationsformen wie Itemformate, Materialien und Hilfsmittel in Abhängigkeit zum Kontext zu wählen und aufeinander abzustimmen. Der *Block II: Item Design* hat neben der konstruktiven vor allem eine beschreibende Funktion. Die konstruierten Items sind hierfür auf der eindimensionalen Skala der Construct Map heuristisch zu verorten – es ist einzuschätzen, ob es sich um ein leichtes Item oder um ein schweres handelt, und es ist zu beschreiben, auf welche Komponenten sich die Unterschiede des Schwierigkeitsniveaus zurückführen lassen. Diese Systematisierungen sollten in den Auswertungsrastern/Scoring-Guidelines Berücksichtigung finden. Ein praktisches Beispiel, bei dem kohärente kausale Beziehungen zwischen curricularen Anforderungen, fachspezifischen Ausdifferenzierungen und kognitionspsychologischen Modellen als Rahmen für die Instrumentenentwicklung formuliert wurden, ist das Science Education for Public Understanding Program (SEPUP) des BEAR Center der University of California, Berkeley (u. a. Roberts, Wilson & Draney, 1997; Wilson & Draney, 2004).

Abbildung 3.5 Construct Map und Item Design (Roberts, Wilson & Draney, 1997)

Score Level	Description	Scoring Guide for "Using Evidence to Make Tradeoffs"
Advanced (4)	Goes beyond expected answer (Level 3) in some significant way.	Accomplishes Level 3 AND goes beyond in some significant way, such as suggesting additional evidence beyond that activity
Correct (3)	Correct and complete answer.	Makes a tradeoff using at least two perspectives or options AND provides objective reasons, supported by relevant and accurate evidence, for each option.
Incomplete (2)	Shows understanding but something important is missing.	States at least two perspectives or options AND provides some objective reasons using some relevant evidence BUT reasons are incomplete and/or part of the evidence is missing.
Incorrect (1)	Appropriate terms or concepts but response is incorrect.	States at least on perspective/option BUT only provides subjective reasons and/or uses inaccurate or irrelevant evidence.
Off task (0)	No response, or response is irrelevant.	Response missing, illegible or lacks reasons AND offers no evidence to support decision made.
X		Student had no opportunity to respond.

Für ein Curriculum aus dem Bereich der naturwissenschaftlichen Grundbildung „Issues, Evidence and You“ wurden fünf verschiedene Konstrukte über Con-

structs Maps ausdifferenziert. Diese gelten als charakteristisch für Lernvorgänge in der Domäne und bilden typische Situations- und Handlungsmuster der Domäne ab: „designing and conducting investigations“, „evidence and tradeoffs“, „understanding concepts“, „communicating scientific information“ sowie „group interaction“. Jedes domänenspezifische Situations- und Handlungsmuster (Konstrukt) wurde über „Abschnitte“ auf einer kontinuierlichen Skala definiert (Construct Map). Diese kontinuierliche Skala dient u. a. als Vorlage für Scoring-Guidelines (Item Design). Als Beispiel sind im Folgenden die qualitativen Level und die dazugehörigen Scoring-Guidelines für das Konstrukt „evidence and tradeoffs“ dargestellt (Abbildung 3.6). Mit diesem Konstrukt soll die Fähigkeit der Lernenden, naturwissenschaftliche Phänomene identifizieren und klassifizieren zu können, abgebildet werden.

Für die Entwicklung von Scoring-Guidelines ist die Konstruktdefinition entlang einer kontinuierlichen Skala des möglichen Antwortverhaltens abzubilden. Für das Beispiel „evidence and tradeoff“ variiert die Konstruktdefinition zwischen einem „advanced“ und „incorrect“ Level der Antwortqualitäten. Wilson spricht in diesem Zusammenhang von einer „Item-Response Construct Map“ (Wilson, 2005, p. 26). Für die Entwicklung und die Ausgestaltung von Items ist darüber hinaus bedeutend, wie sich das Konstrukt in Hinblick auf die Fähigkeiten des Lernenden differenzieren lässt. Bei einer Construct Map, die die Fähigkeiten des Lernenden differenziert und dabei Kriterien nennt, mit deren Hilfe unterschiedliche Fähigkeitslevel im Item Design beschrieben werden können, spricht Wilson von „Respondent Construct Map“:

„The two different aspects of the construct – the respondents and their responses – lead to two different sorts of construct maps: (a) a respondent construct map, where the respondents are ordered from greater to less, and qualitatively may be grouped into an ordered series of levels; and (b) an item-response construct map, where the item responses are ordered from greater to less, and qualitatively may also be grouped into an ordered series of levels“ (Wilson, 2005, p. 26).

Diese bereits in der Definition des Konstrukts angelegte Beziehung zwischen Personenfähigkeit und Itemschwierigkeit spielt später bei der Beschreibung und Implikation des für die Messung relevanten psychometrischen Modells eine entscheidende Rolle. Bis zu diesem Punkt ist jedoch entscheidend, dass durch die Formulierung von Construct Maps und den darauf basierenden Itementwicklungen und Scoringverfahren eine kausale Beziehung zwischen Konstrukt und Messinstrument hergestellt werden kann.

Abbildung von Inferenz

Der nächste Schritt ist, auf Basis der erhobenen Daten auf das zu messende Konstrukt zurückzuschließen, um Aussagen beispielsweise über die kognitiven Fähigkeiten von Lernenden in einer spezifischen Domäne treffen zu können. Eine

erste Grundlage hierfür ist die Zusammenstellung von gescorten und kategorisierten Messwerten. Im Rahmen der „Four Building Blocks“ der Instrumentenentwicklung handelt es sich hierbei um den *Block III: Outcome Space*. Der Outcome Space ist definiert

„[...] as a set of categories that are well defined, finite and exhaustive, ordered, context-specific, and research-based“ (Wilson, 2005, p. 62).

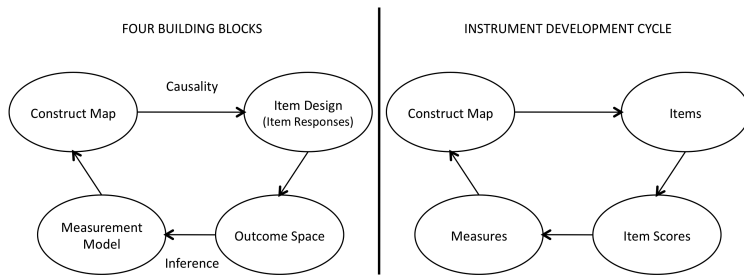
Die Sammlung von Daten im Outcome Space ist eng an den Block II: Item Design gebunden. Die Daten ergeben sich auf Basis des Scoringverfahrens – dieses sollte umfassendes Material darüber enthalten, welches Konstrukt mit dem jeweiligen Item abgebildet wird und es sollte Scoringhilfen anbieten, beispielsweise in Form von Itemvorlagen und deren Antworten. Die Zusammenstellung einer Rohdatenmatrix als zentrales Ziel des Outcome Space ist keinesfalls trivial: Sowohl die Fragen des Itemformats als auch die der Datenqualität haben einen entscheidenden Einfluss.

- Itemformat: Während sich Multiple-Choice-Items einfach scoren lassen („1“ für korrekte Antwort; „0“ für falsche Antwort), sind für offene Antwortformate oder für Antwortformate mit mehreren richtigen Alternativlösungen die Scoringverfahren weit komplexer. In der Assessmentpraxis werden aus diesem Grund häufig erst nach Sichtung der Antworten der Lernenden die Scoring-Guidelines entwickelt.
- Datenqualität: Werden dichotome Daten angestrebt, ist zumindest der Umfang einer richtigen Antwort zu formulieren; bei ordinalen Daten sind hingegen auch qualitative Kategorien im Hinblick auf das unterschiedliche Antwortniveau zu entwickeln.

Die Entscheidung darüber, welche Datenqualität erzeugt wird, hat darüber hinaus weitreichende Konsequenzen für die Wahl des Messmodells (wird wahlweise auch als psychometrisches oder statistisches Modell bezeichnet). Grundlage für das Messmodell ist die Rohdatenmatrix des Outcome Space. Das Ziel des Messmodells ist es, die Daten/Scores des Outcome Space mathematisch begründet auf das Konstrukt zu beziehen. Das Messmodell, das im Rahmen des *Block IV: Measurement Model*, beschrieben wird, hat mehrere Funktionen zu erfüllen: Es soll dabei helfen, (1) die Daten zu verstehen und (2) zu evaluieren, und es soll (3) Kriterien vorgeben, wie die Befunde des Assessments für praktische Ansprüche und Fragen genutzt werden können. Die Wahl des Messmodells ist abhängig von dem Konstrukt, der gewünschten Analyseverfahren und den Präferenzen des Testentwicklers. Im Rahmen der „Four Building Blocks“ wird explizit auf die Vorteile der Nutzung von Modellen der Item Response Theorie hingewiesen. Dazu sei bemerkt, dass die kausale Beziehung zwischen Konstrukt und Messinstrument, so wie sie in den Blöcken I und II: Construct Map und Item Design, theoretisch modelliert wird, auf den zentralen Annahmen der Item Response Theorie basiert. Um letztlich jedoch über das Messinstrument Aussagen zum Messkonstrukt ableiten zu können, sind die Blöcke III und IV: Outcome Space und Measurement Mo-

del in den Assessmentzyklus einzubinden. Sie sind die zentralen Elemente zur Abbildung der Inferenz zwischen theoretischem Konstrukt und messpraktischem Instrument (Abbildung 3.7).

Abbildung 3.7 Four Building Blocks und der Zyklus der Instrumentenentwicklung (Wilson, 2005)



Die „Four Building Blocks“ lassen sich in eine Abbildung überführen, die den Ausgangspunkt und die Richtung der Instrumentenentwicklung vorgibt (Abbildung 3.7). Die Vorstellung, über theoretisch akzentuierte Assessmentmodelle auch die (operative) Instrumentenentwicklung zu strukturieren, ist nicht neu. So formulierten beispielsweise Wright und Stone (1979):

„Four steps must be taken to use a test to measure a person. First, we must work out a clear idea of the variable we intend to make measures on. Second, we must construct items which are believable realizations of this idea and which can elicit signs of it in behavior of the persons we want to measure. Third, we must demonstrate that these items when taken by suitable persons can lead to results that are consistent with our intention. Finally, before we can use any person's score as a basis for their measure, we must determine whether or not their particular pattern of response is, in fact, consistent with our expectations“ (Wright & Stone, 1979, p. 4).

Im Folgenden wird, basierend auf den Ideen der vorgestellten Assessmentmodelle, dargestellt, wie sich vor dem Hintergrund der kaufmännischen Ausbildung Instrumente zur Messung kaufmännischer Kompetenz entwickeln lassen und welche zentralen Gestaltungskriterien die Instrumentenentwicklung leiten.

3.2 Gestaltungskriterium 1: Das konzeptionelle Verständnis der Domäne

In der beruflichen Bildung wird vor dem Hintergrund der Messung von Kompetenz vorrangig diskutiert, auf welche fachlichen und sozialen Inhalte sich Kompetenzen beziehen. Eine weitverbreitete Meinung ist, dass sich Kompetenzen aus der Beschreibung konkreter beruflicher Tätigkeiten ableiten lassen sollten. Dies würde bedeuten, dass sich die Instrumente der Kompetenzmessung an der beruflichen Tätigkeit orientierten und sich auch aus dieser entwickelten. Die Testaufgaben wären – so die theoretische Vorstellung – in diesem Fall einem spezifischen beruflichen Handlungskontext entnommen, der Eigenschaften einer Domäne aufzeigt (vgl. u. a. Wigdor & Green, 1991; Sloane & Dilger, 2005). Achtenhagen (2004, S. 23f.) weist vor dem Hintergrund der deutschen Bedingungen darauf hin, dass die Rahmenrichtlinien und Lehrpläne der beruflichen Bildung als ein Konzept, das konkrete Aufgabenanforderungen für zukünftige, bestimmte Lebenswelten erfasst, interpretiert werden können. Dies hieße, dass die curricularen Vorgaben der Bildungsziele bereits kontextualisiert wären. Der Kontextbezug gilt als eine der zentralen Annahmen der Kompetenzmessung; er ist explizit im Domänenbegriff verankert.

3.2.1 Hypothese von der Domänenspezifität des Lehrens und Lernens

Wird versucht, den Domänenbegriff im Hinblick auf die Erfassung und Entwicklung von Kompetenzen näher zu bestimmen, so sind eingrenzende Vorbemerkungen nötig. Am weitesten geht Beck (2005) mit dem Hinweis darauf, dass die Hypothese von der Domänenspezifität des Lehrens und Lernens dringlich überprüfungsbedürftig sei:

„Über welchem bzw. welchen Kriterien konstituiert sich eine Domäne? Verlaufen ihre Grenzen dort, wo die Regelmäßigkeiten des Lehrens oder Lernens sich ändern? Wo ist das der Fall? Ändern sie sich etwa beim Übergang von der Versicherungs- zur Bankbetriebslehre? Oder beim Übergang vom Rechnungswesen zur Statik für Bauhandwerker? [...] Die Domänehypothese ist in der Sache nichts anderes als eine Einschränkung von Generalisierungsbehauptungen“ (Beck, 2005, S. 551).

In der Tat fällt es schwer, in der einschlägigen Literatur Hinweise und strukturelle Differenzierungen für den Kontrast zwischen domänenspezifischen und allgemein kognitiven Lernprozessen zu finden. Den Arbeiten Funkes zum vorwärts-verkettenden vs. rückwärtsverkettenden Vorgehen im Rahmen der Problembeurteilung (Funke, 2004, S. 288f.) lässt sich beispielsweise entnehmen, dass mit zunehmendem Fachbezug allgemeine Fähigkeiten zur Bewältigung von Anforderungssituationen von spezifischen Fachkenntnissen abgelöst werden. Diese Befunde lassen darauf schließen, dass sich in Anforderungen, die einen konkreten fachlichen Bezug aufweisen, Lernende über ihre zur Bewältigung der Anforderung eingesetzten kognitiven Ressourcen unterscheiden lassen: Lernende mit ausgeprägten fachspezifischen Kenntnissen und Fähigkeiten werden diese auch

zur Anforderungsbewältigung einsetzen, auch wenn allgemeine kognitive Fähigkeiten oder Heuristiken zur Lösung ausreichen würden bzw. die einfachere Lösungsoption darstellten. Übertragen auf die Hypothese der *Domänenspezifität des Lernens* lässt sich folgende Schlussfolgerung ziehen:

Schlussfolgerung I

Unter der Voraussetzung zunehmender Domänenspezifität in fachlich orientierten Anforderungssituationen können mit zunehmender Kompetenz der Lernenden/Auszubildenden Übergänge zwischen allgemeinen kognitiven Fähigkeiten und domänenspezifischen wissensbasierten Entscheidungen erwartet werden (Perspektive des Lernens).

Bei diesem ersten Befund stehen, anders als bei Beck (2005) betont, weniger die Generalisierungen über die Domänen als vielmehr Generalisierungen innerhalb einer spezifischen Domäne im Vordergrund.

Etwas konkreter sind Ergebnisse, die sich aus den Arbeiten zu epistemologischen Überzeugungen des Erwerbs von Wissen ableiten lassen. An dieser Stelle sollen die entsprechenden Forschungsergebnisse nicht im Detail dargestellt werden; jedoch sollen in aller Kürze Ansatzpunkte skizziert werden, mit deren Hilfe sich die Hypothese der Domänenspezifität von Lehren und Lernen prüfen lässt (für Überblicksarbeiten vgl. Cohen & Wartofsky, 1983; Pollock, 1999; Hofer & Pintrich, 2002).

Epistemologische Überzeugungen sind subjektive Konzepte, die sich auf den Erwerb und die Entwicklung von kognitiven Dispositionen beziehen. In der wissenschaftlichen Diskussion weitgehend ungeklärt ist bislang die Frage, ob diese subjektiven Konzepte in Domänen verankert sind oder unabhängig von spezifischen Domänen ausdifferenziert werden. Es gibt eine Vielzahl von Studien, die untersuchen, ob und wie sich epistemologische Überzeugungen von Lernenden in verschiedenen Domänen unterscheiden (vgl. Paulsen & Wells, 1998); die Anzahl der Studien, die der Frage nachgehen, ob eine Person in Abhängigkeit von der fachlichen Domäne über verschiedene epistemologische Überzeugungen verfügt, ist jedoch weit weniger hoch. Insgesamt zeigen sich die Befunde in folgenden Punkten eindeutig:

- Epistemologische Überzeugungen sind in verschiedenen Fachdomänen unterschiedlich ausgeprägt. Die Studien von Schommer (u. a. 1993) zeigen beispielsweise, dass Studierende der Sozialwissenschaften tendenziell über weniger stark ausdifferenzierte epistemologische Überzeugungen verfügen als Studierende der technischen Disziplinen.
- Epistemologische Überzeugungen sind – wie alle subjektiven Konzepte – abhängig von individuellen Variablen wie Geschlecht, Alter, Bildungsstand etc. (u. a. Schommer & Dunnell, 1994).

- Epistemologische Überzeugungen können im Rahmen eines Fachgebietes in Abhängigkeit von der fachlichen Schwerpunktsetzung variieren (z. B. zwischen den mathematischen Teilgebieten der Geometrie und der Algebra; u. a. De Corte, Op't Eynde & Verschaffel, 2002).

Vor dem Hintergrund dieser Befunde wird davon ausgegangen, dass Lernende in Abhängigkeit des Kontextes unterschiedliche epistemologische Überzeugungen vertreten (können) – eine Schlussfolgerung, die weitreichende Konsequenzen im Hinblick auf die Vermittlung von Lerninhalten hat (vgl. Trautwein, Lüdtke & Beyer, 2004). Die individuelle Vorstellung davon, wie Wissen erworben und ausgebaut werden kann, hat Einfluss auf die Fragen der Instruktion: Allgemeine, domänenübergreifende epistemologische Überzeugungen erforderten aus fachdidaktischer Perspektive eine andere Aufbereitung des Lehrstoffs und aus methodischer Perspektive auch eine andere Vermittlungsform, als dies bei domänenspezifischen epistemologischen Überzeugungen der Fall wäre.

Innerhalb der kognitiven Theorien scheint der Domänenbezug des Wissens und der Wissenserwerbsprozesse allerdings eindeutig geklärt. In diesem Zusammenhang ist die Forderung nach domänenspezifischen Vermittlungsprozessen bereits seit den 1970er Jahren verstärkt hörbar. Shulman (1986) beklagt, dass Forschungen im Bereich der Instruktion lange von den falschen Zielen geleitet waren. Anstatt zu fragen, welche Vorstellungen die Lehrenden von den Inhalten ihrer Fachdisziplin haben und wie sie sich eine Vermittlung der Inhalte vor dem Hintergrund von Wissenserwerbsprozessen vorstellen, sei die Forschung auf pädagogisches Verhalten, charakteristische Personenmerkmale von Lehrpersonen oder Lehrmethoden ohne konkreten situativen Bezug fokussiert gewesen:

„The thrust of the cognitive research program in learning is subject matter specific rather than generic. That is, the schemata used to make sense of instruction on photosynthesis in a biology class are completely different from those used to understand the concept of inertia in physics. [...] Most of the cognitive research on teaching has ignored the teacher's cognitive processes in this sense“ (Shulmann, 1986, p. 25).

Stevens und Wineburg (2005) sprechen in diesem Zusammenhang davon, dass die Hypothese der Domänenspezifität des Lernens und Lehrens als allgemeine Norm akzeptiert ist („de facto truth“, Stevens & Wineburg, 2005, p. 125). Das zentrale Forschungsdesiderat ist vor diesem Hintergrund folglich nicht in der Prüfung der Hypothese domänenspezifischer Lehr- und Lernprozesse zu sehen, sondern darauf zu beziehen, wie domänenspezifische und allgemeine, domänenübergreifende instruktionale Konzepte, epistemologische Überzeugungen oder Wissenserwerbsprozesse zusammenwirken und miteinander interagieren. Vor dem Hintergrund der *Domänenspezifität des Lehrens* kann aus diesen Befunden folgende Schlussfolgerung abgeleitet werden:

Schlussfolgerung II

Unter der Annahme der Domänenspezifität ist zu folgern, dass sich fachliche Inhalte in ihren kognitiven Strukturen unterscheiden und folglich vor dem Hintergrund der kognitiven Annahmen zum Wissenserwerb eine unterschiedliche instruktionale Aufbereitung und Vermittlung erfordern (Perspektive des Lehrens).

Zur Definition des Domänenbegriffes sind im Hinblick auf obige Überlegungen die folgenden Aspekte zu betrachten: Zum einen ist zu klären, wie Lernende in unterschiedlichen Domänen agieren, welche typischen Konzepte und Methoden sie anwenden und wie sich Wissenserwerbsprozesse in einer spezifischen Domäne beschreiben lassen. Dem gegenüberzustellen sind zum anderen domänentypische Instruktionsstrategien und fachtheoretische Konzepte der spezifischen Domäne. Unter Rückgriff auf die Ausführungen von Mislevy und Kollegen (u. a. Mislevy, 2008; vgl. Kapitel 3.1.2) kann in diesem Zusammenhang auf die Schicht der Domänenanalyse verwiesen werden.

3.2.2 Domänen- vs. Fachverständnis

In der Expertise „Zur Entwicklung nationaler Bildungsstandards“ werden Domänen als fachspezifische Leistungsbereiche definiert, die sich über einen Bereich von Anforderungssituationen charakterisieren lassen (Klieme, Avenarius, Blum, Döbrich, Gruber et al., 2003).

„Der Bereich von Anforderungssituationen, in denen eine bestimmte Kompetenz zum Tragen kommt, umfasst immer ein mehr oder weniger breites Leistungsspektrum. Die Entwicklung und Förderung von Kompetenzen muss daher eine ausreichende Breite von Lernkontexten, Aufgabenstellungen und Transfersituationen umschließen“ (Klieme, Avenarius, Blum, Döbrich, Gruber et al., 2003, S. 74).

Allerdings wird bei der vorliegenden Definition von fachspezifischen Leistungsbereichen keine Klärung des Fachverständnisses vorgenommen. Der Text enthält jedoch implizit den Verweis auf die Beziehung zwischen Unterrichtsfach und korrespondierender wissenschaftlicher Disziplin mit ihren Überzeugungen, Modellen und Arbeitsweisen (ebd., S. 18). Tenorth (2004) verweist in diesem Zusammenhang auf vier traditionale Fachgefüge, die sich als Domänen anbieten. Hierbei handelt es sich um die Abgrenzung zwischen sprachlich-literaturwissenschaftlichen, mathematisch-naturwissenschaftlichen, historisch-sozialen und ästhetisch-expressiven Profilen.

In der Berufsbildung wird eine derartige Profilbildung kritisch diskutiert und die Übertragbarkeit der Profile auf berufsbezogene Fachbereiche angezweifelt. In

diesem Zusammenhang werden insbesondere die folgenden Schwierigkeiten genannt (vgl. Sloane & Dilger, 2005, S. 16):

- Die Profile sind durch die gymnasiale Oberstufe geprägt und auf den Fachkanon der allgemeinen Bildung bezogen. Dies führt zu einer fehlenden Passung zwischen diesen Profilen und den berufsbezogenen Fächern und Lernfeldern;
- die Profile basieren auf traditionellen fachlichen Zuschnitten. In den berufsbezogenen Fächern und Lernfeldern sind jedoch unterschiedliche fachliche Zugriffe integriert (mathematische, sprachliche, technische), wodurch eine eindeutige Zuordnung zu den Profilen nicht gelingen kann;
- die Profile und die für die Kompetenzmessung zentralen Fragen der output-orientierten Bildungsergebnisse sind bislang nicht aufeinander abgestimmt.

Eine Konsequenz dieser Schwierigkeiten ist, dass für berufliche Domänen fachübergreifende und damit modellvariiierende Domänendefinitionen gefunden werden müssen. Vor dem Hintergrund dieser Überlegung wird im Folgenden das Domänenkonzept so zu umschreiben versucht, dass es für weitere wirtschaftspädagogische Forschungen Anwendung finden kann. Ausgangspunkt sind Aufgaben und Probleme, die in beruflichen Situationen zu bewältigen bzw. zu lösen sind und hierfür auf fachübergreifenden Konzepten basieren (müssen). Im Hinblick darauf soll eine Grenzziehung näherungsweise erfolgen (Achtenhagen & Winther, 2008):

- „Domäne“ ist mehr als nur ein Handlungstyp oder ein Verhalten in eng begrenzten Situationen (z. B. Verbuchen von Beschaffungen).
- „Domäne“ ist mehr als ein eng begrenzter Ausschnitt aus einer Rahmenrichtlinie oder einem Ausbildungsplan (z. B. Sortieren von Beschaffungsvorgängen mit Hilfe einer ABC-Analyse).
- „Domäne“ ist gleich oder weniger als das Handlungs- und Wissensspektrum eines Ausbildungsberufs.
- „Domäne“ ist gleich oder weniger als der Ziel- und Inhaltsbereich eines Unterrichtsfachs bzw. des berufsspezifischen Unterrichts an kaufmännischen Schulen.

Ein weiteres Argument ist in diesem Zusammenhang zentral: Während versucht wird, für die „allgemeine“ Bildung lebensnahe Aufgaben zu entwickeln, stehen für die berufliche Bildung lebensreale Aufgaben und Anforderungen zur Verfügung (vgl. Sloane, 2005, S. 491). Ein Bündel solcher lebensrealer Aufgaben, die sich auf die berufliche Handlungs- und Orientierungsfähigkeit in einem Berufsfeld beziehen (vgl. Brand, Hofmeister & Tramm, 2005; Lehmann, Seeber & Hunger, 2006; Lehmann & Seeber, 2007), könnte als Domäne umschrieben werden (z. B. Controlling unter einem wirtschaftsinstrumentellen Verständnis, Dienstleistungsmarketing). Diese Vorstellungen einer grenzweisen Annäherung an eine berufsbezogene Domänendefinition finden sich vergleichbar auch in den theoretischen Annahmen zu Testkonstruktionen der allgemeinen Bildung; hier wird als

Ausgangskriterium ein fachbezogenes Curriculum gewählt (vgl. u. a. Wilson & Draney, 2004, p. 134):

- Die Domäne ist dem Curriculum nicht äquivalent, sondern im Sinne einer Zusammenfassung des Curriculums komprimiert.
- Die Domäne ist weniger eine inhaltliche Zusammenfassung des Curriculums, sondern vielmehr eine Zusammenstellung der intendierten Effekte des Curriculums.
- Die Domäne ist weniger als die Zielvereinbarung über die Kenntnisse und Fähigkeiten eines Ausbildungsprogrammes, sie ist jedoch beschreibbar über Lernaufgaben, mit denen die Zielvereinbarungen erfüllt werden sollen.

Diese Eingrenzungen veranschaulichen, dass die Fragen der Domänenendefinition insbesondere die Meso- und Makroebene der beruflichen und schulischen Bildung ansprechen. Es geht um die Passung von Lernkultur und Instruktion, die Identifikation der subjektiven Unterrichts- bzw. Unterweisungstheorien und Menschenbilder der Lehrperson mit der Schul- und Unternehmenskultur oder aber um Bezüge zwischen Curricula und Lernorganisation (Seifried, Sembill, Nikolaus & Schelten 2005). Sie zeigen jedoch auch, dass für eine Domänendefinition ein externes Validitätskriterium notwendig ist. Dies ist das Maß für die Grenzen und die inhaltliche Struktur einer Domäne. Für kognitive Tests wie die Kompetenzmessung wäre ein praktikables Verfahren, die Handlungen und Situationen zu analysieren, in denen die Lernenden/Auszubildenden ihre Kenntnisse und Fertigkeiten unter Beweis stellen; ein Verfahren, das sich sowohl für schulische Settings anwenden als auch auf berufliche und damit arbeitsplatzorientierte Settings übertragen ließe (u. a. Frederiksen & White, 2004).

Die Analyse der Settings (mit Settings werden im Folgenden sowohl schulisch als auch betrieblich akzentuierte Lern- und Arbeitsanforderungen umschrieben) kann generell auf zwei Fragen fokussieren. (1) Die Frage nach den personalen Kenntnissen und Einstellungen des Lernenden/Auszubildenden beschäftigt sich mit generellen Dispositionen (traits), die in der Lern- und Arbeitsanforderung benötigt werden (person-oriented). (2) Die Frage, welche konkreten Handlungen oder Arbeiten in einem Setting getan werden müssen, ist hingegen auf die Lern- und Arbeitsanforderung selbst konzentriert (task-oriented) (vgl. Wigdor & Green, 1991, p. 6). Während über die erste Frage die Zielsetzungen von Kompetenzmessung definiert werden können, wird über die zweite Frage die Modellierung von Domänen in kaufmännischen schulischen und betrieblichen Lern- und Arbeitsumgebungen zugänglich.

3.2.3 Domänenmodell für die kaufmännische Bildung

Im Folgenden wird auf der Basis zweier kaufmännischer Ausbildungsgänge ein Domänenmodell für die kaufmännische Berufsbildung vorgestellt: Das Modell bezieht sich zum einen auf die *vollzeitschulische Ausbildung am Fachgymnasium Wirtschaft* und zum anderen auf die *Ausbildung zum Industriekaufmann/zur Industriekauffrau im Dualen System*. Das Domänenmodell ist in Assessmentprozesse,

wie in Kapitel 3.1 skizziert, eingebunden und auch innerhalb dieser entwickelt wurden. Ausgangspunkt der Entwicklungen waren die Ergebnisse des DFG-Projekts „Integrierte Kompetenzförderung in den beruflichen Fächern des Wirtschaftsgymnasiums“ (vgl. Kapitel 2.4.1; Achtenhagen & Winther, 2006; Winther, 2006). Im Rahmen dieses Projekts wurden Kompetenzen als variable Leistungsspektren verstanden, die sich auf einen übergeordneten sinnstiftenden, thematischen Handlungskontext (Domäne) beziehen lassen. Wie in den obigen theoretischen Ausführungen aufgezeigt, ist die Definition eines einzigen, allgemein gültigen Domänenbegriffes für die kaufmännische Ausbildung nicht erstrebenswert. Ein derartiges Vorgehen würde den kognitiven Theorien widersprechen, die Vielfaltigkeit der beruflichen Bildung unberücksichtigt lassen und den Blick auf die Leistungspotentiale der Lernenden/Auszubildenden durch die Fokussierung auf Einzelaspekte des Ausbildungssystems verstellen. Dennoch ist es unter einer wissenschaftlich-pragmatischen Perspektive sinnvoll, Modelle zu entwickeln, die den Anspruch der Generalisierbarkeit verfolgen. Vor dem Hintergrund dieser Prämisse basiert die Formulierung des Domänenmodells auf folgenden Annahmen:

- Als externes Validitätskriterium werden für vollzeitschulische Bildungsprogramme die Curricula und für Ausbildungsprogramme im Rahmen des Dualen Systems neben den Curricula zusätzlich die betrieblichen Ausbildungs- und Arbeitsplatzanforderungen zu Grunde gelegt.
- Die Analyse des Validitätskriteriums erfolgt bezogen auf die Lern- und Arbeitsanforderungen (task-oriented).
- Unterschiedliche fachliche Zugriffe auf die Lern- und Arbeitsanforderungen stellen kein Ausschlusskriterium, sondern eine normative Eigenschaft dar.
- Aus lernpsychologischer Perspektive erfolgt der Domänenzugang kognitiv, d. h., dass auf Informationsverarbeitungsprozesse fokussiert wird und damit auf die Fragen, wie Wissen erworben und wie auf Wissensrepräsentationen zugegriffen wird.
- Die Hypothese der Domänenspezifität von Lernen und Lehren hat uneingeschränkte Gültigkeit.

Diese Annahmen stellen sicher, dass unterschiedliche Ausbildungsgänge im Domänenmodell über die gleiche Ausgangsbasis verfügen und miteinander vergleichbar sind. Mit Blick auf die Forderung nach begründeten und konsistenten Assessmentprozessen werden hier Festsetzungen getroffen, die auf die Testentwicklung und die empirischen Verfahren der Testanalyse und -evaluation ausstrahlen.

Formale Grundlagen und Kriterien der Inhaltskomprimierung

Für die vollzeitschulische Ausbildung am niedersächsischen Fachgymnasium Wirtschaft und für die Ausbildung zum Industriekaufmann/zur Industriekauffrau im Dualen System stehen vergleichende Analysen der Curricula zur Verfügung (vgl. John, 2004). Es zeigt sich, dass die fachbezogenen Inhalte beider Ausbildungsgänge einen hohen Überschneidungsbereich aufweisen (ca. 85 %). Dies ist dadurch zu begründen, dass die niedersächsischen Rahmenrichtlinien des Fach-

gymnasiums Wirtschaft (Niedersächsisches Kultusministerium, 2001) an den Rahmenlehrplänen der Industriekaufleuteausbildung (KMK, 2002) orientiert sind und beide Curricula die Industriebetriebslehre als korrespondierende Bezugswissenschaft implementieren. Für die Ausbildung im Dualen System und damit für den Vergleich der verschiedenen Lernorte gilt formal, dass die Ausbildungsordnung (einschließlich des Ausbildungsrahmenplans) und der Rahmenlehrplan durch das „Gemeinsame Ergebnisprotokoll“ (KMK, 1972) aufeinander abgestimmt sind. Die curricularen Inhalte der Ausbildungsgänge sowie das Ausbildungsberufsbild (vgl. § 4 der Verordnung über die Berufsausbildung zum Industriekaufmann/zur Industriekauffrau, BGBl, 2002) lassen sich inhaltlich so komprimieren, dass sich domänencharakteristische Lern- und Arbeitsanforderungen identifizieren lassen.

Zusätzlich zu den formalen Grundlagen wurden zwei betriebswirtschaftliche Theoriekonzepte als Referenzkataloge für die Inhaltskomprimierung hinterlegt: das St. Galler Managementmodell (Dubs, Euler, Rüegg-Stürm & Wyss, 2004) und die betriebswirtschaftliche Entscheidungstheorie nach Laux (2007).

Im St. Galler Managementmodell werden über betriebliche Managementprozesse, Geschäftsprozesse und Unterstützungsprozesse Orientierungshinweise für eine prozessorientierte Betriebswirtschaftslehre gegeben. Die einzelnen Unternehmensprozesse werden durch drei Integrationsebenen – Strategie, Struktur und Kultur – miteinander in Beziehung gesetzt. In der Entscheidungstheorie nach Laux werden Elemente unternehmerischen Entscheidungshandeln identifiziert, die sich generalisiert auf vielseitige unternehmerische Handlungssituationen anwenden lassen (Laux, 2007, S. 8ff.):

- Die Problemformulierung,
- die Präzisierung des Zielsystems,
- die Erforschung von Alternativen (mit ihren Restriktionen für mögliche Alternativen, der Suche nach und der Prognose der Ergebnisse der Alternativen),
- die Auswahl der Alternativen sowie
- die Entscheidung in der Realisationsphase.

Beide betriebswirtschaftlichen Konzepte sind Beispiele dafür, dass unternehmerisches Entscheidungshandeln ein zentraler Bezugspunkt in modernen Unternehmens- und Managementkonzeptionen, die auf einer prozess- und systemorientierten Betriebswirtschaftslehre basieren, ist. Die gewählten Beispiele der Betriebswirtschaftstheorie folgen dabei durchaus dem in den Richtlinien und Lehrplänen als Zielvorstellung verankerten Konzept der Handlungskompetenz, die sich analytisch in zwei zentrale Teilleistungen untergliedern lässt: (1) Die Fähigkeit zur Orientierung, d. h. zur angemessenen Modellierung von Handlungssituationen. Dies erfordert die kognitive Bewertung der Ausgangssituation, die Modellierung des Zielzustandes sowie das Abwägen alternativer Handlungswege. (2) Die operative Kompetenz, d. h. die Fähigkeit, einen Ist-Zustand in einen Soll-Zustand zu transformieren (vgl. Reetz & Tramm, 2000). Wichtig ist es zu verdeutlichen, dass es sich bei der Lösung eines gegebenen betrieblichen Problems um

eine Abfolge zahlreicher Einzelentscheidungen handelt. Die curricularen Vorgaben zum Ausbildungsberuf Industriekaufmann/-frau sowie die Rahmenrichtlinien für das niedersächsische Fachgymnasium Wirtschaft zeichnen sich im Hinblick auf unternehmerisches Handeln bereits durch eine angemessene fachdidaktische Inhaltsauswahl und -sequenzierung aus, indem sie die Ziele und Inhalte unter der Entscheidungsperspektive strukturieren (Winther, 2006, S. 88): (1) Eine solche fachdidaktische Inhaltsauswahl lässt sich (a) durch eine Reduzierung der wissenschaftlichen Theorie (vgl. Fachwissenschaft-Fachdidaktik-Diskurs: u. a. Kaiser & Kaminski, 2003), (b) durch Abbildung der Ideenkonstrukte zur Beschreibung und Bewältigung relevanter Problemfelder der wirtschaftlichen Praxis (vgl. die neuere Curriculum- vs. Fachdidaktikdiskussion: u. a. Achtenhagen, Tramm, Preiß, Seemann-Weymar, John & Schunck, 1992; Preiß, 1999a; 1999b; 2005; Reetz & Tramm, 2000) und (c) durch Expertiseaufbau im (sinnhaft reduzierten) Inhaltsbereich als Transferbedingung bestimmen (vgl. u. a. Achtenhagen, 2003). (2) Die Inhaltssequenzierung stellt auf die Wissenskonstruktion im Sinne der Domänencharakterisierung ab. Damit steht die Bewältigung von Lern- und Arbeitsanforderungen, deren Wert und deren inhaltliche Verankerung den Lernenden/Auszubildenden bewusst werden sollen, bei diesem Diskussionsschwerpunkt im Vordergrund. Über sinnhafte Inhaltssequenzierung wird eine abstrakte Problemrepräsentation angestrebt, da hierdurch zahlreiche Relationen zwischen Intention, Intension und Extension der Konzepte aufgebaut und so grundlegende, allgemeine Prinzipien der Disziplin aufgedeckt und herausgestellt werden können. In diesem Zusammenhang ist die Frage nach den Bildungsinhalten verändert zu stellen: Was sind die Zugänge zu einem beruflichen Handlungsbereich? Welches sind die zentralen domänenbeschreibenden Denkfiguren und Begriffe für das Verständnis eines beruflichen Handlungsbereichs? Der Managementliteratur lassen sich vor dem Hintergrund dieser Fragestellungen unterschiedliche kognitive Anforderungen entnehmen, die das Spektrum der Leistungsanforderungen in der Domäne beschreiben (vgl. Kapitel 3.3.2).

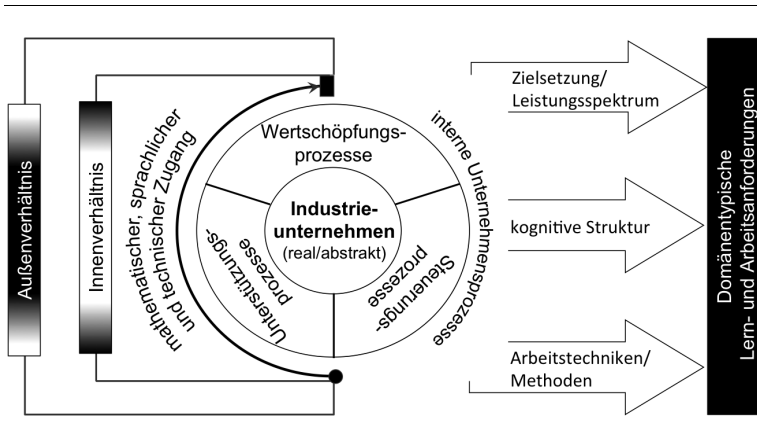
Unternehmensprozesse zur Identifizierung von Lern- und Arbeitsanforderungen

Inhaltsauswahl und Inhaltssequenzierung bereiten die Komprimierung der Curricula bzw. die Beschreibung von Arbeitsplatzanforderungen vor. Die als domämentypisch identifizierten Inhalte lassen sich über folgende Kategorien in ein Spektrum verschiedener Lern- und Arbeitsanforderungen überführen (vgl. Abbildung 3.8): Die Kategorie I: Zielsetzung/Leistungsspektrum, verweist auf die Lern- und Handlungsziele des schulischen Unterrichts bzw. der betrieblichen Unterweisung. Zentral ist, dass die Zielsetzungen eine operationalisierte Form aufweisen müssen, um einer Messung zugänglich zu sein. Hierfür ist es hilfreich, das jeweilige Leistungsspektrum über eine kontinuierliche Einschätzung einerseits des Leistungsniveaus der Lernenden und andererseits des Anforderungsniveaus der Lern- und Arbeitsanforderung zu bestimmen.

In den Rahmenrichtlinien des Unterrichtsfachs Betriebswirtschaft mit Rechnungswesen/Controlling heißt es beispielsweise (Niedersächsisches Kultusmini-

sterium, 2001, S. 31): Die Schülerinnen und Schüler wenden geeignete Programme der Auftragsabwicklung an. Das Leistungsspektrum lässt sich zunächst einmal im Sinne der persönlichen Zielerreichung formulieren. Dies kann beispielsweise dichotom erfolgen, indem geprüft wird, ob die Anwendung erfolgreich („1“) oder nicht erfolgreich („0“) war. Hierzu ist die Zielsetzung jedoch zu operationalisieren. Mit Bezug auf die Frage der Identifizierung domänentypischer Lern- und Arbeitsanforderungen kommt der Kategorie I: Zielsetzung/Leistungsspektrum, noch eine weitere Bedeutung zu. Das operationalisierte Lern- und Handlungsziel kann über Anforderungssituationen mit unterschiedlichem Schwierigkeitsgrad erreicht werden. Das Leistungsspektrum beschreibt sich folglich auch über das Anforderungsniveau der Lern- und Arbeitssituation. Unter Rückgriff auf das Beispiel der Software für die Auftragsabwicklung kann die Spannweite beispielsweise von der Eingabe von Bestelldaten (geringes Anforderungsniveau) bis zur Steuerung von Produktionsprozessen (hohes Anforderungsniveau) reichen.

Abbildung 3.8 Unternehmensprozesse zur Identifikation von Lern- und Arbeitsanforderungen (vorläufiges Domänenmodell)



Die Kategorie II: kognitive Struktur, befasst sich mit der mentalen Repräsentation von Anforderungssituationen. Bezogen auf die Analyse von schulischen und betrieblichen Anforderungssituationen handelt es sich hierbei um die Fragen, über welches Wissen ein Lernender/Auszubildender im Hinblick auf die Anforderungssituation verfügen muss und welche Regeln des Wissenszugriffs greifen, um die Anforderungssituation bewältigen zu können (vgl. u. a. Lesgold, Lajoie, Logan & Eggen, 1990). Unter Rückgriff auf die Arbeiten Greenos (1998; Gelman & Greeno, 1989; vgl. auch „strands of proficiency“: Kilpatrick, Swafford & Findell, 2001, p. 117) wird für eine Analyse zwischen drei verschiedene Strukturen unterschieden (vgl. dazu Kapitel 2.2.2 sowie 2.4.2): deklaratives Wissen, prozedurale Fähigkeit und die Fähigkeit, Anforderungssituationen konzeptual und handlungslei-

tend zu erfassen (konditionales/interpretatives Wissen). Deklaratives Wissen ist einerseits darüber definiert, wie umfangreich das Wissen über die Inhalte der Anforderungssituation ist, und andererseits über die Erfahrung, über die die Lernenden/Auszubildenden zum Lösen der Anforderungssituation verfügen. Prozedurales Wissen beinhaltet die mentale Repräsentation der Ziele und der Prozeduren zur Bewältigung der Anforderungssituation und erfasst darüber hinaus auch die Regeln der Selektion und Anwendung der Prozeduren für die Leistungserbringung. Interpretatives Wissen erfasst im Überschneidungsbereich von deklarativem und prozeduralem Wissen die subjektive Wahrnehmung und Bewertung der Anforderungssituation.

Das operationalisierte Lern- und Handlungsziel und die Anforderungssituation: Anwendung geeigneter Programme der Auftragsabwicklung, zielt für eine Messung auf eine prozedurale kognitive Struktur im Sinne der Verfügbarkeit technischer Fähigkeiten. D. h., dass die kontextualen Wissensbestände sowie die Fähigkeiten der konzeptualen Anforderungserfassung bei dieser Formulierung der Zielsetzung unberücksichtigt bleiben, selbst wenn sie zur Bewältigung der Anforderungssituation im Hintergrund von Bedeutung sind. Diese Sichtweise schließt explizit mit ein, dass Kompetenz in einer Domäne keine eindimensionale kognitive Disposition ist und es folglich im Hinblick auf Kompetenzerwerb und -entwicklung nicht genügen wird, nur auf eine spezifische kognitive Struktur zu fokussieren (hierzu auch Kilpatrick, Swafford & Findell, 2001, p. 116).

Die Kategorie III: Arbeitstechniken/Methoden, beschreibt die operativen Verfahren der Anforderungsbewältigung. Hierzu zählen Materialien, theoretische Modelle und Denkstrukturen, Arbeitsweisen und -methoden sowie Arbeitsmittel, die als typisch für eine spezifische Anforderungssituation anzusehen sind. Das obige Beispiel aus den Rahmenrichtlinien verweist auf den in Unternehmen alltäglichen Einsatz von ERP-Software.

Zur Beschreibung eines für die Domäne typischen Spektrums von Lern- und Arbeitskontexten, Aufgabenstellungen und Transfersituationen wurden die Lerninhalte der Rahmenrichtlinien und des Rahmenlehrplans sowie ausgewählte Tarifbeispiele des Entgelt-Rahmenabkommens (ERA; vgl. Datenbank IG Metall) unter der Prämisse system- und prozessorientierter Unternehmensstrukturen miteinander verglichen – dies war notwendig, da insbesondere die Berufscharakteristiken und die darauf basierenden Tarifbeispiele nach wie vor eher funktionsorientiert formuliert sind.

Exkurs: System- und Prozessorientierung in der Betriebswirtschaftslehre

In den Curricula kaufmännischer Ausbildungsgänge wird gefordert, die Lerninhalts- und Zielstrukturen möglichst komplex anzulegen, um träges Wissen zu vermeiden und ein Verständnis der Dynamik ökonomischer Realität zu entwickeln. Vor diesem Hintergrund werden statisch-linearisierte Sichtweisen betrieblicher Realität aufgegeben und der

systemorientierte Ansatz der Betriebswirtschaft verfolgt. Systemorientierung bedeutet, die Zusammenhänge zwischen einzelnen Themen, Konzepten und Instrumenten herauszustellen, um die einzelnen Wissensэлеmente zu einem sinnhaften Ganzen zu integrieren, das spezifische, dynamische Beziehungen abzubilden vermag.

Die Systemperspektive der Betriebswirtschaft ist zu ergänzen um die Prozessorientierung: Alle Wertschöpfungsaktivitäten einer Unternehmung und die dazu notwendigen organisatorischen und administrativen Arbeiten werden in Prozessen erbracht, die sich durch eine bestimmte sachliche und zeitliche Logik beim Vollzug spezifischer Aufgaben charakterisieren lassen. Funktionsübergreifende Geschäftsprozesse fragen folglich nach der Reihenfolge, dem zeitlichen Verlauf, den Prioritäten von Teilaufgaben sowie nach dem Einsatz von Menschen, Sachgütern und der Verwendung von Informationen (betriebliche Ablauforganisation) (Exkurs entnommen aus Winther, 2006, S. 151f.).

Über folgende Prozesse lassen sich Arbeitsvorgänge in realen Industrieunternehmen respektive Lern- und Handlungsprozesse in authentisch simulierten Lernumwelten strukturieren: Wertschöpfungsprozesse, Steuerungsprozesse und Unterstützungsprozesse. Wertschöpfungsprozesse erfassen den quantifizierbaren Prozess der Leistungserstellung sowie deren Vermarktung, über Steuerungsprozesse werden Entscheidungshilfen für das Management vorbereitet und Unterstützungsprozesse greifen Fragen der Unternehmensführung und Organisation auf (vgl. Tabelle 3.9). Die internen Unternehmensprozesse lassen sich sowohl in einem betrieblichen Innenverhältnis (z. B. Abstimmung zwischen Abteilungen) als auch bezogen auf ihr marktorientiertes Außenbild (z. B. Beziehungen zu Kunden, Lieferanten, Kreditgebern) definieren.

Ein weiteres Beschreibungskriterium ist die Art des Zugangs innerhalb der Prozesse. Kaufmännische Situationen zeichnen sich dadurch aus, dass zur Bewältigung von Lern- und Arbeitsanforderungen mathematische, sprachliche und zum Teil technische Kenntnisse und Fertigkeiten unterstützend und parallel zum Zugriff auf ökonomisch modellierte Wissensbestände in die Lösungsprozesse eingebunden werden. Dieses allgemein kognitive Wissen hat in beruflichen Anforderungssituationen tendenziell eine prozedurale Struktur. Im Rahmen des Kompetenzstrukturmodells für die kaufmännische Bildung wird hierunter *domänenverbundene Kompetenz* verstanden (vgl. Kapitel 2.4.2). In Abbildung 3.9 sind den internen Unternehmensprozessen operationalisierte Lern- und Handlungsziele aus den Curricula sowie Tarifbeispiele und die zugehörigen Berufe entsprechend dem Entgelt-Rahmenabkommen gegenübergestellt. Jeder ausgewählte Beruf erfordert eine dreijährige Ausbildung zum Industriekaufmann/zur Industriekauffrau bzw. eine vergleichbare Qualifikation. Die Gegenüberstellung macht deutlich, dass sich die verschiedenen formalen Vorgaben zum einen inhaltlich auf die gewählten in-

ternen Unternehmensprozesse übertragen lassen und dass zum anderen starke inhaltliche Parallelen zwischen den Curricula und den Arbeitsplatzbeschreibungen bestehen.

Abbildung 3.9 Unternehmensprozesse und zugeordnete Lern- und Arbeitsanforderungen

interne Unternehmensprozesse (Inhaltskomprimierung)	Niedersächsische Rahmenrichtlinien Fachgymnasium Wirtschaft	Rahmenlehrplan für den Ausbildungsberuf Industriekaufmann/Industriekauffrau	Tarifbeispiele und ausgewählte Berufe entsprechend Entgelt-Rahmenabkommen
Wertschöpfungsprozesse <ul style="list-style-type: none"> Erfassen den Prozess der Leistungserstellung sowie deren Vermarktung vor dem Hintergrund des Zielsystems des Unternehmens. Die Wertschöpfung ist messbar über Beschaffungs-, Produktions- und Vertriebsprozesse. 	Lerngebiet 2: Auftragsabwicklung mit Beschaffungs-, Produktions- und Vertriebslogistik (Planung von Beschaffungsvorgängen und Aufgaben und Ziele der Materialwirtschaft erkennen)	Lernfeld 6: Beschaffungsprozesse planen, steuern und kontrollieren (Bedarfsanforderungen bearbeiten, Analyse der Bezugsquellen, Angebotsvergleiche durchführen, Bestellabwicklung kontrollieren)	Einkaufssachbearbeiter/-in (02.01.01.10) Einholen von Angeboten (Bedarfsanalyse, Angebote einholen, auswählen und Bestellung vorbereiten) Durchführung von Bestellungen (Bestellungen überwachen und Störungen klären)
Steuerungsprozesse <ul style="list-style-type: none"> Erfassen Verfahren und Methoden der Unternehmenssteuerung. Steuerungsprozesse basieren auf Betriebsdaten und sind Dienstleistungen im Sinne von Entscheidungshilfen für das Management. 	Lerngebiet 3: Erfassung, Verteilung, Analyse und Zurechnung von Kosten (Kosten- und Leistungsrechnung als Element der Unternehmenssteuerung verstehen, Werte aus der Betriebsdatenerfassung aufbereiten, bewerten und analysieren)	Lernfeld 2: Marktorientierte Geschäftsprozesse erfassen (Rechnungswesen als Mittel zur Erfassung, Steuerung und Überwachung der Wertschöpfung darstellen)	Sachbearbeiter/-in Betriebsbuchhaltung (01.04.01.10) Datenaufbereitung für Abschlüsse (Rückstellungs-, Aufwands- und Ausgleichskonten führen) Vorbereitung der Ergebnisrechnung (Daten für Ist-Ergebnisrechnung zusammenstellen, Ergebnisrechnung kommentieren)
Unterstützungsprozesse <ul style="list-style-type: none"> Erfassen Werte und Strukturen die sich nicht direkt der unternehmerischen Wertschöpfung zuordnen lassen. Hierzu zählen rechtlich und strukturelle Fragen der Führung und Organisation. 	Lerngebiet 6: Konzepte der Unternehmensführung und Organisationsentwicklung (Auseinandersetzung mit Fragen des Personalbedarfs, der Personalauswahl und der Personalbeurteilung)	Lernfeld 7: Personalwirtschaftliche Aufgaben wahrnehmen (Zusammenstellung von Kriterien zur Bewerberauswahl, Bewertung von Arbeitsverträgen unter Berücksichtigung des Arbeits- und Sozialrechts, Tarifrechts und der Betriebsvereinbarungen)	Personalsachbearbeiter/-in (03.01.01.10) Durchführung von Personalmaßnahmen (Bewerberauswahl, Informationen zu betrieblichen und tariflichen Regelungen, Personaldatenerfassung)

Auf Basis dieses Vergleiches lassen sich Items entwickeln, die sich an den Curricula der Ausbildungsgänge sowie an den Beschreibungen von Arbeitsplätzen orientieren. Die inhaltliche Justierung benötigt jedoch zusätzlich eine differen-

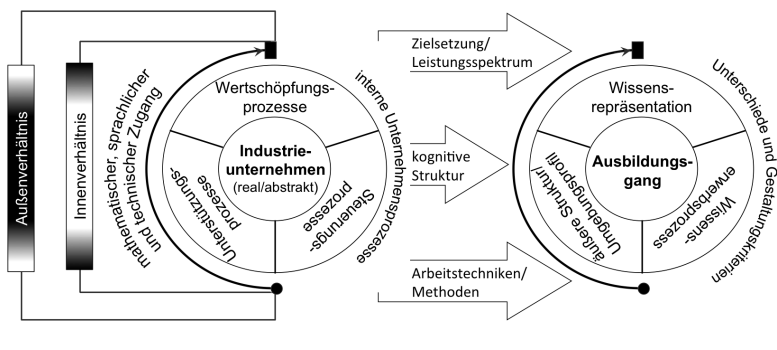
zierte Beschreibung der Unterschiede und Besonderheiten der zu untersuchenden Ausbildungsgänge.

Spezifika der Ausbildungsgänge als Kriterien des Domänenmodells

Eine Analyse und Komprimierung der Ausbildungsinhalte im Hinblick auf die Kategorien Zielsetzung/Leistungsspektrum, kognitive Struktur und Arbeitstechniken/Methoden greift für eine erfolgreiche Itemkonstruktion und Implementierung des Assessments zu kurz. Nicht ohne Grund wird in der Literatur zur Testentwicklung deutlich darauf verwiesen, dass Assessments in Abstimmung mit der Lern- und Arbeitsumgebung und den bestehenden Lern- und Arbeitsbedingungen entwickelt und eingesetzt werden sollen. Dazu zählt auch, dass sich die Lehrenden mit dem Assessment identifizieren und die Notwendigkeit und den Nutzen der Testung vor ihren Schülerinnen und Schülern vertreten können (vgl. Beiträge in Downing & Haladyna, 2004).

Die Spezifika der Ausbildungsgänge variieren nicht nur im Vergleich zwischen den Ausbildungsgängen (vollzeitschulisch vs. dual), sondern auch innerhalb der Ausbildungsgänge selbst (verschiedene didaktisch-methodische Ansätze in den Ausbildungsphasen bzw. Größe und Ausstattung des Ausbildungsbetriebes). Vor diesem Hintergrund ist es notwendig, Kriterien zu entwickeln, die typische Muster und Bedingungen der verschiedenen Ausbildungsgänge und -verläufe für die Konstruktion von Testitems und Testformaten sowie für die Testimplementierung aufbereiten und die den Testentwicklern als Konstruktionshilfen zur Verfügung stehen.

Abbildung 3.10 Domänenmodell für die kaufmännische Bildung



Hierfür wird das vorläufige Domänenmodell (vgl. Abbildung 3.8) um drei Aspekte erweitert, die (a) die Unterschiede zwischen Ausbildungsgängen – hier insbesondere fokussiert auf die verschiedenen Lernorte – theoretisch determinieren können und die (b) Hinweise auf den Gestaltungsspielraum von Assessments in der kaufmännischen Bildung geben. Diese Aspekte sind: (1) Wissensre-

präsentation und (2) Prozess des Wissenserwerbs sowie (3) die äußere Struktur bzw. das Umgebungsprofil der Lern- bzw. Arbeitsanforderung. Anders formuliert beschreiben die Aspekte typische Lernprozesse und deren instruktionale Unterstützung sowie die Beziehung zwischen Lern- und Lehrvorgängen vor dem Hintergrund der Lern- und Arbeitsbedingungen in konkreten Lern- und Arbeitsprozessen. Die ersten beiden Aspekte sind maßgeblich für Fragen der Entwicklung von Testitems; der letztere kann die Entscheidung für Testformate begründet leisten.

Betriebliche und schulische Lernprozesse unterscheiden sich in einem Punkt ganz zentral: Ein betriebliches Setting stellt eine reale Unternehmenssituation dar; in einem schulischem Setting werden (authentische) Modelle einer realen Unternehmenssituation oder Abstraktionen der Realität als Lern- bzw. Arbeitskontext gewählt. Diese Ausgangslage hat entscheidende Bedeutung im Hinblick auf die *Prozesse des Wissenserwerbs* und hier vor allem auf *die Arten der Wissensrepräsentation* (u. a. Mayer, 2004). Zu identifizieren, welche Wissensrepräsentationen vorliegen und wie diese zur Bewältigung von Lern- und Arbeitsanforderungen genutzt werden, ist Teil der Domänenanalyse:

„The analysis of any domain in which learning is to be assessed must include identifying the knowledge representations that people work with and how they use them. Because developing facility with knowledge representations is central to developing expertise in any domain, claims about students' capabilities will be central to assessment: capabilities to choose representations, express information and obtain information from them, transform information from one representation to another, use representations to coordinate actions in situations and interactions with colleagues“ (Mislevy, 2007, p. 282).

Wissensrepräsentationen sind gebunden an die Modelle, die sich Lernende/Auszubildende von der Realität machen, sowie an deren narrative Strukturen. Diese unterscheiden sich zwischen betrieblichen und schulischen Lern- und Arbeitsumgebungen insbesondere durch ihren Authentizitätsgrad und den Anwendungsbezug – Aspekte, die insbesondere in der Diskussion um die Qualität von Instruktion in kaufmännischen Kontexten eine besondere Rolle spielen und auf die *Wissenserwerbsprozesse* in Lern- und Arbeitssituationen ausstrahlen (Tramm, 1992; Achtenhagen, 1998; 2001; 2003).

Einer Untersuchung von Hattie (1999; zit. n. Hays, 2006, p. 75) lässt sich beispielweise entnehmen, dass der Wissenserwerbsprozess maßgeblich von drei Faktoren beeinflusst ist: von der Verstärkung bzw. Übung des Lerninhalts in praktisch relevanten Anwendungskontexten, dem Vorwissen bzw. der Erfahrung der Lernenden und der Qualität der Instruktion. Instruktion ist hier in einer weiten Fassung als Unterstützungsprozess des Wissenserwerbs zu verstehen, der neben den Vermittlungsprozessen auch die Qualität der Lern-/Arbeitsmaterialien umfasst. Von den drei relevanten Faktoren sind folglich zwei konkret auf den

Lern- und Arbeitskontext bezogen: Übungs- und Festigungsprozesse als Elemente des Lernens sowie instruktionale Qualität aus Perspektive des Lehrens. Im Hinblick auf das Wechselspiel von Lernen und Lehren ist es sinnvoll, Wissenserwerbsprozesse in einzelne Phasen zu zerlegen. In Anlehnung an Fitts und Posner (1967) findet Wissenserwerb in drei Phasen – der kognitiven, der inhaltsorientierten und der autonomen Phase – statt. Die kognitive Phase umfasst verschiedene Informationsbearbeitungsprozesse, die die Integration neuer Informationen in alte Wissensstrukturen ermöglichen. Informationsverarbeitungsprozesse werden durch das Generieren visueller Vorstellungen gestützt. Betriebliche und schulische Settings haben vor diesem Hintergrund unterschiedliche Voraussetzungen: Realität mit ihrem Komplexitätsgrad einerseits und (didaktisch reduzierte) Modelle der Realität mit variierendem Authentizitätsgrad andererseits. Während der inhaltsorientierten Phase wird das Erlernte ausgebaut und eingeübt. Die autonome Phase zeichnet sich dadurch aus, dass das Erlernte als aktivierbare habituelle Wissensrepräsentation zur Verfügung steht. Die Phasen unterscheiden sich untereinander durch die Anzahl der Fehler und Fehlerquellen sowie durch die Schwankungsbreite im Leistungsbereich. Während in der kognitiven Phase das Erlernen neuer Inhalte mit Fehlern verbunden und das Leistungsspektrum sehr variabel ist, zeigt die autonome Phase das insgesamt erworbene Leistungsvermögen im Hinblick auf einen spezifischen Lerninhalt.

Für die Itementwicklung ist folglich zu dokumentieren, (1) auf welche Phase des Wissenserwerbsprozesses sich das Assessment bezieht und (2) in welchem Lernkontext und unter welchen instruktionalen Bedingungen das Wissen erworben wurde: In schulischen Lernumgebungen dominieren rezeptive und aktive Lern- und Vermittlungsprozesse auf Basis mehr oder weniger authentischer Materialien und die Phase der Automatisierung von Lerninhalten ist zeitlich i. d. R. extrem reduziert. Im Gegensatz dazu sind die Formen des Lernens in betrieblichen Umgebungen vor allem nachahmend und entdeckend mit dem Ziel des konkludenten und entscheidungsbasierten Handelns angelegt. Zur Herausbildung von beruflichen Routinen werden die inhaltsorientierte und die autonome Phase des Wissenserwerbs deutlich stärker betont, als dies in schulischen Umgebungen der Fall ist/sein kann.

In Abbildung 3.11 werden die Spezifika der Wissensrepräsentation und des Wissenserwerbs in einer Gegenüberstellung der vollzeitschulischen Ausbildung am Fachgymnasium Wirtschaft und der dualen Ausbildung zum Industriekaufmann/zur Industriekauffrau im Hinblick auf ihre Konsequenzen für die Gestaltung von Testitems beschrieben. Für eine bessere Veranschaulichung wird angenommen, dass sich das Assessment für den dualen Ausbildungsgang ausschließlich auf die Leistungsfähigkeiten der Auszubildenden am Lernort Betrieb konzentriert.


Für die Gestaltung von Testitems lässt sich im Hinblick auf die Spezifika der Wissensrepräsentation und des Wissenserwerbs zusammenfassend Folgendes feststellen: Eine optimale Anpassung der Items an domänenspezifische Lern- und Arbeitsanforderungen erfolgt im Hinblick auf die Kategorien Zielsetzung, kogniti-

ve Struktur und Arbeitstechniken, mit deren Hilfe sich (1) Lern- und Anforderungssituationen in der Domäne beschreiben lassen (vgl. Ausführungen zum vorläufigen Domänenmodell) und die (2) als Konstruktionsvorschriften für Testitems im Rahmen des Domänenmodells interpretiert werden können.

Abbildung 3.11 *Spezifika der Wissensrepräsentation und des Wissenserwerbs in verschiedenen Ausbildungsgängen*

Wissens- repräsentation / Wissenserwerb	Duale Ausbildung (Lernort Betrieb)	Vollzeitschulische Ausbildung	Erläuterndes Beispiel (Betrieb ≈ Schule)
Modell und narrative Struktur der Wissensrepräsentation	Abbildung eines Modells betrieblicher Realität auf der Basis von Beobachtung	Abbildung eines Modells didaktisch reduzierter betrieblicher Realität auf Basis betriebswirtschaftlicher Theoriebildung	Abstimmungsprozesse zwischen Produktion und Vertrieb im Ausbildungsbetrieb ≈ Darstellung der Konsequenzen, die Veränderungen im Produktionsablauf für den Vertrieb haben
Wissenserwerb in der kognitiven und inhaltsorientierten Phase	nachahmend bis entdeckend	rezeptiv bis aktiv	Materialdisposition durch Eingabe in das betriebliche ERP-System ≈ Darstellung der Verfahren der Materialdisposition
Wissenserwerb in der autonomen Phase	Ausbildung beruflicher Routinen durch Wiederholung und Einübung im täglichen Arbeitsprozess	Ausbildung des sicheren Umgangs mit Verfahren und Modellen (teilweise ausgelagert in außerschulische Lernprozesse)	tägliche Buchung der Vorsteuer als Teil des Arbeitsprozesses ≈ Einübung und ökonomische Interpretation der Vorsteuerbuchung

Rückschlüsse auf die Konstruktion von Testitems

Zielsetzung	Abbildung von Handlungsfähigkeiten in Arbeitsprozessen im Hinblick auf das Prinzip der Fehlerminimierung	Abbildung von Wissensbeständen im Hinblick auf das Prinzip der Maximierung des ökonomischen Verständnisses	
kognitive Struktur	eher prozedural statt deklarativ	eher deklarativ statt prozedural	
Arbeitstechniken	Handlungsformen des Arbeitsprozesses	betriebswirtschaftliche Modellbildung und deren Anwendung	

Items für ein vorrangig betriebliches Assessment sollten die Abbildung von Handlungsfähigkeiten in spezifischen Arbeitsprozessen bzw. Anforderungssituationen zum Ziel haben. Das Leistungsspektrum ist hierfür im Sinne des *Prinzips der Minimierung* von Fehlern zu formulieren. Dieses Vorgehen entspricht den Ansprüchen der betrieblichen Praxis. Zur Erreichung dieser Zielsetzung werden

Items benötigt, die eher prozedurale Fähigkeiten als deklarative Wissensbestände testen. In der Literatur wird in diesem Zusammenhang auch von „procedural fluency“ oder „procedural competence“ (vgl. u. a. Kilpatrick, Swafford & Findell, 2001; Hays, 2006; Winther & Achtenhagen, 2009b) gesprochen. Hierunter wird verstanden, dass domänenbezogene Handlungen und Prozeduren flexibel, fehlerfrei, effizient und angemessen im Hinblick auf die Anforderungssituation selektiert und ausgeführt werden können. Die Handlungsformen des Arbeitsprozesses sind folglich auch die für das Assessment relevanten Arbeitstechniken.

Items für ein Assessment, das vorrangig in vollzeitschulischen kaufmännischen Ausbildungsgängen eingesetzt werden soll, haben zum Ziel, Wissensbestände im Hinblick auf das *Prinzip der Maximierung* des ökonomischen Verständnisses zu testen. Diese Kategorie ist in der Literatur als „conceptual understanding“ oder „conceptual competence“ (u. a. Greeno, 1998; Kilpatrick, Swafford & Findell, 2001; Winther & Achtenhagen, 2009b; 2010) vertreten und beschreibt das ganzheitliche und funktionale Verständnis zentraler domänenspezifischer Ideen – und umfasst damit mehr als das isolierte Wissen von Fakten und Methoden. Das Wissen ist als kohärentes Ganzes im Hinblick auf betriebswirtschaftliche Theorien und Konzepte strukturiert. Items sind folglich so zu konstruieren, dass diese deklarative kognitive Struktur auch tatsächlich getestet werden kann. Hierzu werden als Arbeitstechniken die betriebswirtschaftliche bzw. theoretische Modellbildung sowie deren Anwendungen hinterlegt.

Wissensrepräsentation und Wissenserwerbsprozess sind eng an die *äußere Struktur bzw. das Umgebungsprofil* des Lern- und Arbeitskontextes gebunden. Während sich die Lern- und Arbeitsanforderungen in kaufmännischen Ausbildungsgängen inhaltlich auf interne Unternehmensprozesse komprimieren lassen, zeigen sich im Hinblick auf die instruktionalen Voraussetzungen und Handlungsformen große Varianzen. In betrieblichen Arbeitsumgebungen ist der Arbeitsprozess des Auszubildenden Bestandteil eines realen, komplexen Geschäftsprozesses. Die Handlungen des Auszubildenden werden durch reale betriebliche Vorgänge ausgelöst und die Arbeitsmaterialien sind speziell auf den konkreten Arbeitsprozess abgestimmt. In schulischen Lernumgebungen bietet es sich an, von *simulierten* (authentischen) Anforderungssituationen zu sprechen. Der Authentizitätsgrad der Simulation ist dabei abhängig von der instruktionalen Aufbereitung des Lerninhalts: An kaufmännischen Schulen lassen sich Fallbeispiele, Fallstudien sowie computerbasierte Modellunternehmen finden, für deren Bearbeitung auf (authentisch) konstruierte Materialien zurückgegriffen wird. Letztlich bedeutet dies, dass eine zugrunde liegende situierte Lern- bzw. Arbeitsanforderung, die in schulischen und betrieblichen Umgebungen auf identische Lerninhalte abstellt, vor dem Hintergrund der verschiedenen Lern- und Arbeitsorte unterschiedlich repräsentiert ist und verschiedene Formen der Bearbeitung und folglich unterschiedliche Arbeitsprodukte hervorbringen kann (vgl. Abbildung 3.12).

Die instruktionalen Voraussetzungen, die zur Verfügung stehenden Materialien und letztlich die in spezifischen Anforderungssituationen erzeugten Produkte sollten Einfluss auf die Wahl des Testformats haben. Die Lern- und Arbeitspro-

dukte sind die Daten des Assessments, die als Ergebnis der Bewältigung von realen Arbeitsprozessen in betrieblichen Kontexten und von simulierten Anforderungen im Rahmen schulischer Ausbildungsgänge erzeugt werden und anhand derer sich Aussagen über die zugrunde liegenden Kenntnisse, Fähigkeiten und Fertigkeiten ableiten lassen.

Abbildung 3.12 Spezifika der äußeren Struktur in verschiedenen Ausbildungsgängen

äußere Struktur Umgebungsprofil	Duale Ausbildung (Lernort Betrieb)	Vollzeitschulische Ausbildung	Erläuterndes Beispiel (Betrieb ≈ Schule)
instruktionale Voraussetzungen	realer Arbeits- prozess als Teil- element eines komplexen Geschäfts- prozesses	simulierte (authentische) Anforderungssituation	Auftragsabwicklung inklusive möglicher Störprozesse ≈ Teilprozesse der Auftragsabwicklung und Einzeleffekte spezifischer Störungen
Materialien	Realdaten der Wertschöpfungs-, Steuerungs- und Unterstützungs- prozesse (konsistenter Datenkranz)	Simulierte Daten des Wertschöpfungs-, Steuerung- und Unterstützungs- prozesses (nicht konsistenter Datenkranz)	reale Datenbanken, Unternehmenssoftware, stoffliche Listen (multimedial) ≈ simulierte und reduzierte Einzeldaten (i. d. R. unimedial)
Lern- und Arbeitsprodukte	Ergebnis eines realen Arbeits- prozesses	Ergebnis der simulierten (authentischen) Anforderungssituation	Output des betrieblichen ERP- Systems (automatisierte Erstellung betrieblicher Belege) ≈ Einzelerstellung von betrieblichen Belegen

Rückschlüsse auf die Wahl des Testformats

Testformat	standardisierte Beobachtung des Arbeitsprozesses/ Evaluation der realen Arbeits- produkte oder Simulation des Arbeitsprozesses anhand eines umfassenden Unternehmens- modells auf Basis eines konsistenten Datensatzes	Evaluation der Arbeitsprodukte (i. d. R. paper-pencil) oder Simulation einer beruflichen Anforderungssituation anhand eines umfassenden Unternehmens- modells auf Basis eines konsistenten Datensatzes	
------------	---	---	--

Als Testformate bieten sich vor diesem Hintergrund sämtliche Aufgaben an, die in betrieblichen und schulischen Umgebungen anzutreffen sind und aus denen Lern- und Arbeitsprodukte hervorgehen, die evaluiert werden können – beispielsweise die Maschinenstundensatzberechnung kalkuliert mit Papier und Bleistift oder anhand eines entsprechenden Computerprogramms. Wird dieses Prinzip verfolgt, werden für schulisch akzentuierte Assessments vorrangig paper-

pencil-Formate gewählt; für betrieblich orientierte Erhebungen müssten die Zugriffe auf Realdaten des Ausbildungsbetriebes und die Anwendung der Unternehmenssoftware als Testformate angesehen werden. Dies ist möglich, aber insbesondere aus zwei Gründen nicht praktikabel:

- Die Realdaten des Ausbildungsbetriebes sind sensibel und nicht alle relevanten Daten sind einem Auszubildenden ohne Einschränkung zugänglich.
- Die Unternehmenssoftware ist zwischen den einzelnen Ausbildungsbetrieben und zum Teil auch innerhalb eines Ausbildungsbetriebes nicht standardisiert.

Für betriebliche Assessments sei daher empfohlen, Arbeitsprozesse anhand eines umfassenden Unternehmensmodells einschließlich eines hierfür nötigen konsistenten Datensatzes zu simulieren und die (technologie-basierte) Simulation als Testformat zu wählen (hierzu im Detail Kapitel 5.2.2). Die gleiche Empfehlung kann für Assessments in schulischen Umgebungen gegeben werden – hier insbesondere mit dem Ziel, dass die simulierten beruflichen Anforderungssituationen über sämtliche interne Unternehmensprozesse hinweg konsistent im Hinblick auf ihre zeitliche und wertmäßige Struktur konstruiert sind.

Für den Einsatz von betrieblichen Simulationen als Testformat sprechen darüber hinaus noch folgende Aspekte: (1) In schulischen Umgebungen kann man tendenziell von einer Individualleistung und auch von einer individuellen Verantwortung für die gezeigte Leistung ausgehen; dies schließt auch Formen der Gruppenarbeit mit ein. In betrieblichen Umgebungen ist im Gegensatz dazu die individuelle Leistung in betriebliche Ablauf- und Funktionsstrukturen eingebunden. Die individuelle Leistung hat folglich einen kollektiven Effekt – dieser kann durch den Einsatz von betrieblichen Simulationen auch für schulische Lern- und Arbeitsprozesse sichtbar gemacht werden. (2) In den Rahmenlehrplänen sowie in der Ausbildungsberufscharakteristik werden Inhaltsbereiche als zentral für die Ausbildung angesehen, die in betrieblichen Anforderungssituationen aufgrund ihrer Datensensibilität oder ihrer unternehmerischen Folgen von den Auszubildenden entweder gar nicht oder nur in Zusammenarbeit mit einem erfahrenen Mitarbeiter bearbeitet werden dürfen. Ein traditionelles Beispiel hierfür sind Maßnahmen der Personalentwicklung. Es ist folglich im Einzelfall zu prüfen, über welche Legitimationen ein Auszubildender verfügt und wie die Verlaufs- und Berechtigungswege in betrieblichen Realitäten aussehen. Diese Beschränkungen sind insbesondere für Assessments von Bedeutung, die sich auf die direkte Bewertung von beruflichen Ausbildungsleistungen am Arbeitsplatz konzentrieren (workplace-based assessments). Durch den Einsatz betrieblicher Simulationen lässt sich dieses Problem lösen, indem Berechtigungen vergeben werden oder der Auszubildende die Position eines entsprechend berechtigten Mitarbeiters einnimmt.

Das Domänenmodell stellt für Testentwickler einen kognitiven Kriterienpool dar, mit dem sich (1) domänenspezifische Inhalte identifizieren lassen und der (2) Regeln vorgibt, mit denen sich Inhalte in Lern- und Arbeitsanforderungen übersetzen lassen. Die identifizierten Inhalte sowie die Bedingungen der Lern-

und Arbeitsanforderungen werden vor dem Hintergrund kognitiver Bewertungen aufeinander bezogen. Damit wird es möglich, die kognitiven Prozesse, Strategien und Wissensbestände, die der Bewältigung von Anforderungssituationen zugrunde liegen, zu explizieren. Vor diesem Hintergrund können Hinweise gegeben werden, die Einfluss auf die Entwicklung von Testitems und auf die Wahl des Testformats haben.

In einem nachfolgenden Schritt sind die lösungsrelevanten kognitiven Prozesse und Ressourcen auf die spezifischen Eigenschaften der Anforderungssituationen zu beziehen, um die Frage zu beantworten, wie sich Lösungsprozesse bei einer systematischen Variation des Anspruchsniveaus der Anforderungssituationen verändern. Ziel ist es, solche Einflussfaktoren zu identifizieren, die das Anspruchsniveau direkt erhöhen bzw. reduzieren können und mit deren Hilfe sich Veränderungen im Lösungsverhalten in Folge der Variation des Anspruchsniveaus möglichst detailliert beschreiben lassen. Hierfür werden die über das Domänenmodell identifizierten Lern- und Arbeitsanforderungen in Eigenschaftsbündel zerlegt, mit deren Hilfe der Zusammenhang zwischen Anforderungssituation und Antwortverhalten modelliert werden kann (Konstruktrepräsentation; vgl. Embretson, 1998; Glaser, Lesgold & Lajoie, 1987) – ein Zusammenhang, der nach der qualitativen Analyse später in das psychometrische Modell eingeht (vgl. Kapitel 4.2).

3.3 Gestaltungskriterium 2: Die Konstruktrepräsentation

Im Zentrum des Domänenmodells für die kaufmännische Bildung stehen Vorstellungen über domänenspezifische Wissensbestände, Fähigkeiten und Fertigkeiten sowie über charakteristische Handlungsweisen in domämentypischen Anforderungssituationen, die sich aus der intendierten Zielsetzung und der typischen Struktur kaufmännischer Ausbildungsgänge ableiten lassen. Im Folgenden ist die Frage zu beantworten, welche kognitiven Prozesse und Ressourcen zur Bewältigung einer spezifischen Anforderungssituation notwendig sind. Während im Domänenmodell ausgehend von einer inhaltlichen Struktur auf die Ebene der Kognition geschlossen wurde, werden im Rahmen der Konstruktrepräsentation die kognitiven Prozesse und Ressourcen ihrerseits als Ausgangspunkt für die Beschreibung von Lern- und Anforderungssituationen gewählt.

Hypothesen über den Lösungsprozess werden folglich in Hypothesen über die Inhalte der Anforderungssituationen und deren Schwierigkeitsparameter übersetzt, um zu einer detaillierten Beschreibung des Konstrukts zu gelangen. Hierfür ist eine hinreichend genaue Vorstellung darüber zu entwickeln, welche Prozesse beim Bearbeiten und Lösen der Anforderungssituationen ablaufen:

„Construct representation refers to the processes, strategies, and knowledge stores that are involved in item responses. Construct representation is studied by task decomposition methods [...] Construct representation is understood when the various components, metacomponents, strate-

gies, and knowledge stores that are involved in solving psychometric items are explicated“ (Embretson, 1985, p. 196).

Die kognitiven Prozesse und Ressourcen geben Auskunft darüber, wie das Konstrukt repräsentiert ist (vgl. Block I: Construct Map, im Rahmen der „Four Building Blocks“); sie sind die charakteristischen Merkmale, mit deren Hilfe sich Anforderungssituationen so zerlegen lassen, dass Aussagen darüber getroffen werden können, wie sich die zur Anforderungsbewältigung notwendigen kognitiven Prozesse und eingesetzten kognitiven Ressourcen verändern, wenn bestimmte Merkmale der Anforderungssituation variiert werden. Embretson (1998) geht im Umkehrschluss davon aus, dass sich die Schwierigkeits- bzw. Komplexitätsgrade der kognitiven Prozesse über die Eigenschaften/Stimuli der Testitems operationalisieren lassen. Eine systematische Variation der Itemeigenschaften lässt damit Aussagen über das zu messende Konstrukt zu. In Anlehnung an Wilson (2005; 2008) sind hierfür möglichst differenzierte qualitative Abstufungen zu modellieren, um eine exakte Messung zu ermöglichen.

„Cognitive models specify the major processes involved in item solving. The relative difficulty of the various underlying cognitive processes are operationalized by varying items stimulus features [...] Test items must represent a broad range of difficulty to measure persons precisely at the various ability levels“ (Embretson, 1998, p. 383).

Das Ziel der Variation von Schwierigkeits- und Komplexitätsgraden der Testitems ist es, Aussagen über die Lösungswahrscheinlichkeiten zu schätzen und in diesem Zusammenhang unter Rückgriff auf ein kognitives Modell begründen zu können, warum einige Items schwieriger als andere zu lösen sind.

Im Folgenden werden drei verschiedene Kategorien von Itemeigenschaften beschrieben, die sich in den Analysen zur Messung kaufmännischer Kompetenz als signifikante Prädiktoren der Itemschwierigkeiten ergeben haben und die der kognitivistischen Auffassung von Kompetenz entsprechen (vgl. Kapitel 2.1.3). Diese Itemeigenschaften sind: Funktionale Modellierung, Inhaltliche Komplexität und Kognitive Taxonomierung.

3.3.1 Das Prinzip der Funktionalen Modellierung

Anforderungssituationen unterscheiden sich hinsichtlich der Verfügbarkeit und des Grades der funktionalen Modellierung (modeling). Bezogen auf ein kognitives Kompetenzmodell bedeutet funktionale Modellierung (a) die Verfügbarkeit eines Verarbeitungsmodells der Anforderungssituation und die Fähigkeit, für die Anforderungssituation adäquate Handlungs- und Bewältigungsstrategien auf Wissens- bzw. Erfahrungsbasis einzusetzen, und (b) den Zugriff der Lernenden/Auszubildenden auf Modelle der gegebenen Anforderungssituation sowie die Aktivierung von situationsadäquaten Verarbeitungsmodellen.

Funktionale Modellierung bezieht sich damit im Verständnis der angewandten Kognitionspsychologie auf Verstehensprozesse, Handlungsschemata und Bewäl-

tigungsstrategien der Lernenden/Auszubildenden in spezifischen schulischen und betrieblichen Lern- und Arbeitsanforderungen (Glaser, 1976; Shavelson, Baxter & Pine, 1992; Yuan, Steedle, Shavelson, Alonzo & Oppezzo, 2006; Reif, 2008). Hier zeigen sich im Vergleich zwischen leistungsstarken und leistungsschwächeren Lernenden Unterschiede, die sich auf die Fähigkeiten der Lernenden, situative Anforderungssituationen in bearbeitungsfähige Modelle der Problem- bzw. Situationslösung zu übersetzen, zurückführen lassen. Über das Prinzip der funktionalen Modellierung werden folglich die Beziehungen zwischen Wissensrepräsentationen und den Prozessen der Wissensanwendung definiert (vgl. Markman, 1999). Hierbei kann eine systematisch heuristische Struktur beginnend bei unvollständiger bzw. nicht nachweisbarer Modellierungsleistung verlaufend bis zur vollständigen Modellierungsleistung angenommen werden. Diese Struktur besagt nichts anderes, als dass Lernende in Abhängigkeit ihrer Modellierungsleistungen bei der Bewältigung einer spezifischen Lern- und Arbeitsanforderung qualitativ unterschiedliches Lösungsverhalten zeigen. Aus der lerntheoretischen Forschung ist bekannt, dass die Modellierungsleistungen abhängig sind von (1) dem Grad der Vertrautheit und (2) der praktischen Erfahrung, die Lernende im Umgang mit spezifischen Lern- und Arbeitsanforderungen erworben haben. Sind Lernende mit den Anforderungssituationen vertraut, fällt es ihnen leichter, die Probleme/Herausforderungen der Lern- und Arbeitssituation zu erkennen und zu definieren; sie können einfacher alle lösungsrelevanten Aspekte der Anforderungssituationen enkodieren und adäquatere Lösungswege wählen, als dies bei Vorlage unbekannter Lern- und Arbeitsanforderungen der Fall wäre (vgl. u. a. Reif, 2008). Unterschiede im Lösungsverhalten lassen sich unter dieser Annahme auf Unterschiede im Hinblick auf die Dekontextualisierung von Wissensbeständen und Fertigkeiten sowie auf die Verstehenstiefe der kontextualisierten Lern- und Anforderungssituation zurückführen (vgl. hierzu „lack of conditional knowledge“ sowie „lack of conceptual knowledge“; u. a. Byrnes, 2008; Bransford, Brown & Cocking, 2000; Bransford & Stein, 1984).

Testentwickler können diese Kenntnisse der kognitiven Theorie dafür nutzen, das Anspruchsniveau der Anforderungssituationen systematisch zu manipulieren, um Unterschiede im Lösungsverhalten zu provozieren und um die erzeugten Unterschiede dann auch begründet definieren zu können. Für das Prinzip der funktionalen Modellierung beziehen sich die Variationen des Anspruchsniveaus auf Hinweise, die als funktionale Modellierungshilfen in die Testitems eingebunden sind. Diese Modellierungshilfen können einerseits den Zugriff auf Wissensrepräsentationen erleichtern und andererseits die Auswahl von situationsangemessenen Lösungsvarianten und -methoden steuern.

Aus Perspektive der Kognitionstheorie beziehen sich diese Modellierungshilfen auf die Beeinflussung der mentalen Modelle, die Lernende/Auszubildende von einem Lern- bzw. Arbeitskontext haben, und auf die Frage, wie sich Funktionalität innerhalb dieser Modelle beschreiben lässt. Klein und Hoffman (2008, p. 65) merken in diesem Zusammenhang an, dass gegenwärtig kein Verfahren im Rahmen der experimentellen oder kognitiven Psychologie existiert, das umfas-

send aufzuklären vermag, wie Lernende denken, handeln und mit ihrer Umgebung interagieren: Die Modelle über das Lernen sind folglich unvollständig und in ihrer Aussagekraft limitiert, sie können jedoch, fokussiert auf einzelne Bereiche des Lern- bzw. Lösungsprozesses, helfen, diesen zu beschreiben und zu verstehen.

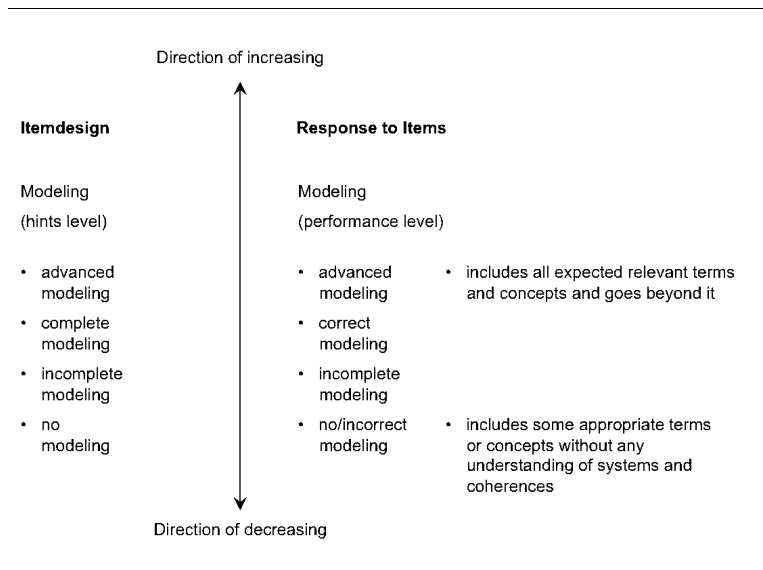
Das Prinzip der funktionalen Modellierung konzentriert sich auf einen solchen Ausschnitt des Lern- bzw. Lösungsprozesses, indem auf die über das Lernen entwickelten mentalen Modelle der Lernenden/Auszubildenden in spezifischen Anforderungssituationen fokussiert wird. Mentale Modelle lassen sich über Beziehungen zwischen Wissensrepräsentationen beschreiben – diese Beziehungen können zeitliche, räumliche, begriffliche, organistorische und/oder systemische Strukturen aufweisen. Über die Ausgestaltung von Testitems lassen sich diese Begründungszusammenhänge stützen bzw. stören.

Als Beispiel sei angenommen, dass Lernende am Fachgymnasium Wirtschaft einen Kundenauftrag abwickeln sollen. Hierzu ist es wichtig zu wissen, dass der Produktionsprozess immer durch die Auftragsabwicklung angestoßen wird, falls die vom Kunden nachgefragte Ware nicht vorrätig bzw. eingeplant am Lager ist. Um die gesamte Auftragsabwicklung zu beherrschen, benötigt der Lernende detaillierte Kenntnisse über unternehmensinterne Beschaffungs-, Produktions- und Vertriebsprozesse – er muss folglich über eine „kognitive Landkarte“ der Inhalte dieser Prozesse und der Wechselwirkungen zwischen den Prozessen verfügen. Auf dieses kognitive Beziehungsgefüge kann der Itementwickler Einfluss nehmen. Er kann beispielsweise Störungen beim Wareneingang oder in der Fertigungssteuerung in das Testitem implementieren, um die Aufgabenstellung komplexer zu machen, oder er kann relevante Eckdaten der Auftragsabwicklung als Ist-Werte vorgeben und damit den Umfang der Modellierungsleistung reduzieren. In beiden Fällen würde er in die zeitlichen und systemischen Strukturen des Modells der Auftragsabwicklung eingreifen: zeitlich im Hinblick auf die Interdependenzen zwischen den Geschäftsprozessen und systemisch unter der Perspektive, dass Daten des Modells als Ankerpunkte für den Lösungsprozess vorliegen und diese Daten den Zugriff auf die relevanten Wissensbestände erleichtern.

Diese bewussten Manipulationen im Rahmen der Itementwicklung einerseits sowie die Modellierungsleistungen der Lernenden/Auszubildenden andererseits, lassen sich auf einer gemeinsamen Skala verorten (vgl. „Construct Map“; Wilson, 2005). Diese Skala definiert sich über Extrempunkte zwischen unvollständiger und weitreichender Modellierung (vgl. Abbildung 3.13). Für den Bereich des Itemdesigns (linke Seite der Abbildung) können Aussagen darüber getroffen werden, in welchem Umfang der Itementwickler in die kognitiven Repräsentationen der Lernenden/Auszubildenden durch eine bewusste Manipulation der Anforderungssituation eingreift. Der Umfang der funktionalen Modellierung ist entscheidend davon abhängig, wie offen bzw. vorstrukturiert eine Anforderungssituation präsentiert wird. Ein herausragende Modellierungsleistung (Stufe 4) wird dann eingefordert, wenn die Anforderungssituation im Sinne betriebswirtschaftlicher Entscheidungsmodelle sämtliche Entscheidungsebenen offen lässt: Pro-

blemformulierung, Präzisierung des Zielsystems, Auswahl und Erprobung von Lösungen sowie die Entscheidung in der Realisationsphase. Der Umfang an eingeforderter Modellierungsleistung reduziert sich in dem Maße, wie das Entscheidungsmodell vorstrukturiert ist. Keine Modellierung ist notwendig, wenn alle Entscheidungsebenen in der Anforderungssituation weitgehend vorgegeben sind, indem beispielsweise auf theoretische Konzepte oder Lösungsmethoden verwiesen wird (Stufe 1). Bei einer umfangreichen Modellierung würden hingegen keine Reduzierungen des theoretischen Modells vorgenommen und dem Lernenden/Auszubildenden keine Modellierungshilfen gegeben (Stufe 3).

Abbildung 3.13 Funktionale Modellierung



Auf Seiten des Antwortverhaltens (rechte Seite der Abbildung) lassen sich Unterschiede im Hinblick auf die Modellierungsleistungen der Lernenden/Auszubildenden beschreiben. Die Frage hier ist, in welchem Umfang die Lernenden/Auszubildenden auf ihre kognitiven Repräsentationen in einer systematisch manipulierten Anforderungssituation zurückgreifen können. In diesem Zusammenhang kann von einer nicht notwendigen inkorrekten Modellierungsleistung gesprochen werden, wenn nur Fragmente bzw. einzelne Konzepte des entsprechenden Modells präsentiert werden können und diese keine zusammenhängende bzw. ökonomisch sinnvolle Struktur aufweisen. Von weitreichender Modellierungsleistung kann hingegen dann ausgegangen werden, wenn der Lernende/Auszubildende die Anforderungssituation umfassend modelliert hat und in seinem Lösungsprozess auf alle relevanten Konzepte und Beziehungen sinnvoll verweisen kann.

Entsprechend der Skala kann nun angenommen werden, dass in einer Anforderungssituation, in der umfangreiche Modellierungshilfen gegeben werden, die Modellierungsleistungen der Lernenden in der Masse positiver ausfallen, als dies in einer Anforderungssituation ohne Modellierungshilfen der Fall wäre. Die Konsequenz ist, dass sich die Lösungswahrscheinlichkeiten der Anforderungssituationen unterscheiden – dies ist gleichsam ein Indiz dafür, dass sich die lösungsrelevanten kognitiven Prozesse und Ressourcen der Lernenden/Auszubildenden in den verschiedenen Anforderungssituationen unterscheiden.

3.3.2 Das Prinzip der Inhaltlichen Komplexität

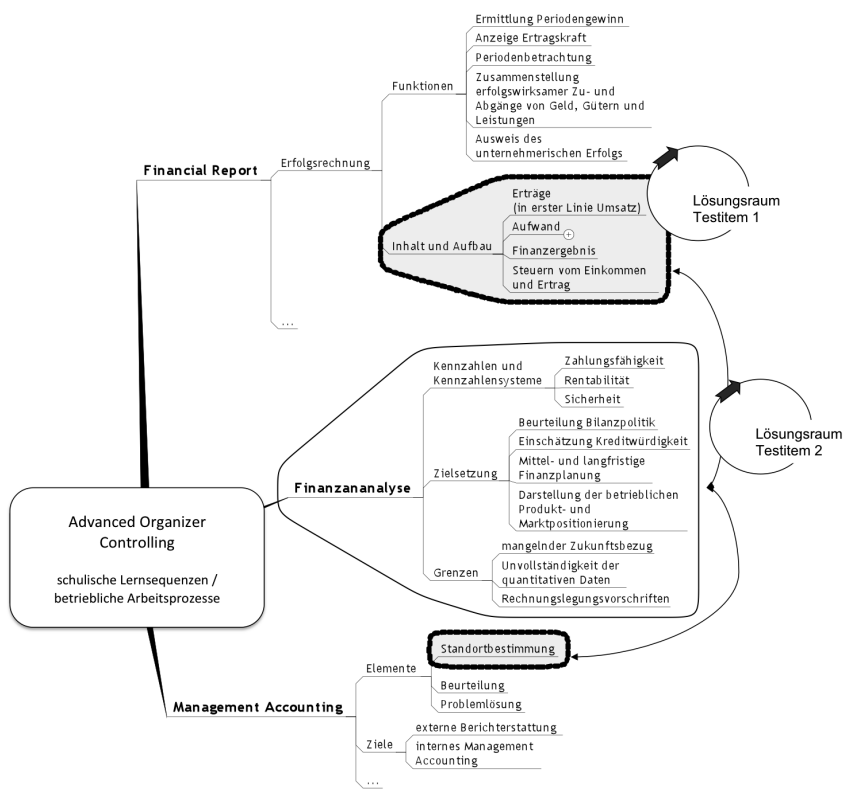
Das Prinzip der funktionalen Modellierung lässt Aussagen über den Schwierigkeits- bzw. Komplexitätsgrad von Anforderungssituationen zu. Mit Hilfe des Prinzips der funktionalen Modellierung können unterschiedlich anspruchsvolle Testitems im Hinblick auf einen spezifischen curricularen bzw. betrieblichen Inhalt konstruiert werden. Die Komplexität der Modellierung von Anforderungssituationen ist dabei jedoch auch von der inhaltlichen Struktur der Lern- und Arbeitsanforderung abhängig: Anforderungssituationen unterscheiden sich folglich auch hinsichtlich ihrer *inhaltlichen Komplexität* (content).

Inhaltliche Komplexität zielt darauf ab, dass Lern- und Arbeitsanforderungen als unterschiedlich anspruchsvoll empfunden werden können. Die Arbeiten zur Mathematical Literacy Scale der PISA-Studien (u. a. Adams & Wu, 2004) verdeutlichen, dass die Wahrnehmung inhaltlicher Komplexität von mehreren Faktoren abhängen kann: Zentral sind in diesem Zusammenhang vor allem (1) die Anzahl der Lösungsmöglichkeiten und (2) der Umfang der zu verarbeitenden lösungsrelevanten Variablen. Beide Faktoren lassen sich in eine logische Beziehung zu dem Umfang des zu testenden Inhaltsbereiches im Rahmen curricularer Lernbereiche/-felder bzw. zu dem Aufgabenprofil in betrieblichen Situationen setzen.

Um curriculare und arbeitsplatzcharakteristische Leistungsanforderungen einheitlich bewerten zu können, wird auf die Arbeiten der korrespondierenden Bezugswissenschaft zurückgegriffen. Insbesondere der Managementliteratur lassen sich unterschiedliche kognitive Anforderungen entnehmen, die das Spektrum der inhaltlichen Leistungsanforderungen in Lern- und Arbeitskontexten beschreiben (Dubs, Euler, Rüegg-Stürm & Wyss, 2004; Laux, 2007). Allen Konzeptionen ist gemeinsam, dass die Leistungsanforderungen über einander bedingende Lernsequenzen bzw. Arbeitsprozesse kumulativ verarbeitet werden. Zur Veranschaulichung werden im Folgenden unterschiedliche Inhaltskategorien des Controlling dargestellt (u. a. Bleicher, 1996; Rüegg-Stürm, 1996). Die Strukturierung des Controlling als Teilgebiet der Betriebswirtschaft orientiert sich zum einen an curricularen Vorgaben und zum anderen an betrieblichen Arbeitsanforderungen. Ausgangspunkte sind die Darstellung, Vermittlung und Aussagekraft finanzieller Größen für unternehmensexterne und -interne Nutzer (Financial Reporting). Die Auseinandersetzung mit den einzelnen Instrumenten der Rechnungslegung und deren Inhalten ist notwendig, um im Rahmen der Finanzanalyse und Unternehmensbewertung die für unternehmerische Entscheidungen not-

wendige Strukturierung und Planung anhand quantitativer Beurteilungen des Unternehmens zu begründen. Diese Bewertungen sind letztlich Entscheidungsgrundlagen im Rahmen des Management Accounting. Das Reporting-System sammelt Informationen auf allen Stufen der Unternehmung (kurzfristige Erfolgsrechnung, Kostenermittlung), beurteilt Chancen und Risiken der laufenden Aktivitäten anhand von Budgets und Abweichungsanalysen und bewertet Alternativstrategien vor dem Hintergrund der finanziellen Stabilität und des erfolgswirtschaftlichen Wertzuwachses eines Unternehmens als Zielgrößen des Controlling.

Abbildung 3.14 Advanced Organizer für den Inhaltsbereich Controlling



In schulischen und betrieblichen Lern- und Arbeitsprozessen wird ein fachdidaktisch begründetes System domänenspezifischer Fakten, Prozeduren und Entscheidungstheorien angestrebt, welches dem kumulativen Charakter von Lern- und Verstehensprozessen gerecht wird. D. h., die Lernenden werden zunächst in die Inhalte des Financial Reportings eingeführt; es folgen die Inhalte der Finanz-

analyse und letztlich werden die Konzepte des Management Accounting vermittelt. Abbildung 3.14 zeigt die verschiedenen Inhaltskategorien des Controlling in Form eines Advanced Organizers für die Planung und Struktur der Instruktion (Abbildung gekürzt). Der Advanced Organizer kennzeichnet jedoch nicht nur den Verlauf schulisch kumulierter Lernprozesse; er gibt auch Hinweise auf betriebliche Strukturen und hier auf die betrieblichen Arbeitsprozesse und deren inhaltliche Verknüpfung mit dem Controlling. Damit sind die Strukturen im Inhaltsbereich Controlling im besonderen Maße geeignet, das Prinzip der inhaltlichen Komplexität zu veranschaulichen. Die Wissens- und Erfahrungsbasis sowie die verfügbaren theoretischen Methoden und Modelle der Lernenden/Auszubildenden in den einzelnen Inhaltsbereichen des Controlling sind sehr gut separierbar und können folglich in konkreten Testitems direkt angesprochen werden. Folgende Testitems sind zur Darstellung des Prinzips der inhaltlichen Komplexität gegeben:

Testitems Inhaltsbereich Controlling

Testitem 1: Der Controller legt Ihnen die aktuelle Erfolgsrechnung des Unternehmens vor und bittet Sie, die fehlenden Werte zu ermitteln, um die Erfolgsrechnung zu vervollständigen.

Testitem 2: Der Controller bittet Sie, folgende Kennzahlen aus den Ergebnissen der Erfolgsrechnung zu berechnen: a) Return on Sales sowie b) EBITDA. Bitte klären Sie in diesem Zusammenhang zusätzlich, ob diese Kennzahlen geeignete Instrumente der unternehmensinternen Steuerung sind. Interpretieren Sie hierfür die Aussagekraft der Kennzahlen und ermitteln Sie gegebenenfalls alternative Indikatoren.

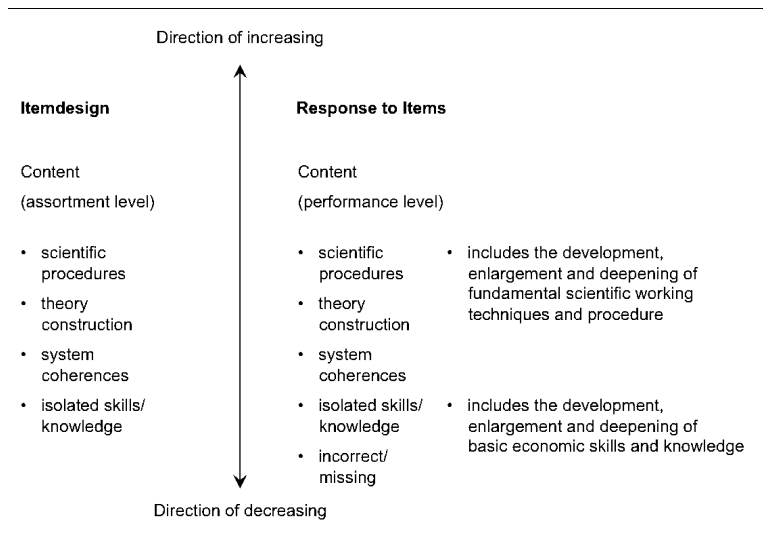
Testitem 1 und 2 unterscheiden sich zentral hinsichtlich ihrer inhaltlichen Komplexität. Testitem 1 verlangt von den Lernenden/Auszubildenden die Verarbeitung eines Repertoires an Fertigkeiten und Algorithmen, das in der gelernten Form direkt und isoliert einsetzbar ist. Die Lernenden/Auszubildenden müssen zur Bewältigung der Anforderungssituationen den Aufbau einer betrieblichen Erfolgsrechnung kennen und in diesem Zusammenhang Werte der Erfolgsrechnung benennen bzw. diese berechnen. Damit erfasst das Testitem 1 einen sehr begrenzten Inhaltsbereich aus der komplexen Struktur des Testgebiets Controlling (vgl. Abbildung 3.15). Der mit dem Testitem 2 angesprochene Inhaltsbereich ist hingegen viel weiter gefasst. Zur Berechnung der Kennzahlen sind die Einzelvariablen aus der Erfolgsrechnung zu identifizieren und in der Formel der Kennzahl zu verarbeiten. Der Kennzahlwert ist anschließend ökonomisch und inhaltlich zu interpretieren. Die Lernenden/Auszubildenden sollen auf Basis ihrer Interpretation zu Modifizierungen bzw. Neuentwicklungen von Indikatoren als

Hinweissysteme der unternehmensinternen Steuerung gelangen. Hierfür ist es notwendig, ein betriebswirtschaftliches Konzept in vollem Umfang zu validieren und es in seiner praktischen und wissenschaftlichen Bedeutung zu erfassen.

Testitem 2 unterscheidet sich von Testitem 1 folglich in folgenden Punkten: (1) Es ist zeitlich bzw. inhaltlich strukturell dem Testitem 1 nachgeordnet und baut damit auf den kognitiven Ressourcen der Lernenden/Auszubildenden aus vorangegangenen Lernprozessen auf; (2) es umfasst darüber hinaus einen größeren Inhaltsbereich und stellt damit auf das Verständnis eines betriebswirtschaftlichen Konzepts in voller Breite ab.

Die verschiedenen Inhaltskategorien charakterisieren zunächst nur die Anforderungssituation; sie stellen noch keine Hierarchie im Hinblick auf den Schwierigkeits- bzw. Komplexitätsgrad einer Anforderungssituation dar. Dies lässt sich jedoch dadurch erreichen, dass einerseits die charakteristischen Merkmale der inhaltlichen Komplexität sowie andererseits das erwartete Antwortverhalten der Lernenden/Auszubildenden auf eine gemeinsame Skala transformiert werden (vgl. Abbildung 3.15).

Abbildung 3.15 *Inhaltliche Komplexität*



Auf Seiten der Bedingungen für die Itemkonstruktion (linke Seite der Abbildung) werden Faktoren genannt, mit deren Hilfe ein Testitem einfacher bzw. komplexer gestaltet werden kann: Bezieht sich der Inhalt des Testitems auf Konzepte, Zusammenhänge und Prozesse aus zwei oder mehreren curricularen Bereichen, wird davon ausgegangen, dass im Vergleich zu einem Testitem, mit dem ausschließlich Fakten, einfache Berechnungen und Zusammenhänge eines isolierten curricularen Bereichs erfasst werden sollen, weniger Lernende/Aus-

zubildende es lösen können. Die Lösungswahrscheinlichkeiten reduzieren sich weiter, wenn es im Rahmen der Theoriekonstruktion und der Anwendung von wissenschaftlichen Konzepten und Methoden um Entscheidungsfindungen und Modellformulierungen unter Rückgriff auf zwei oder mehrere curriculare Bereiche geht. Im Hinblick auf das Antwortverhalten und damit auf die Leistungsfähigkeiten der Lernenden/Auszubildenden (rechte Seite der Abbildung) erfasst die Skala der inhaltlichen Komplexität das Einbringen grundlegender Wissensbestände am unteren Leistungsrand sowie die Verwendung wissenschaftlicher Arbeitstechniken und -prozeduren zur Lösung eines Testitems am oberen Ende des Leistungsspektrums.

In diesem Zusammenhang sind zwei Einschränkungen zu berücksichtigen: Die Testitems müssen (1) so beschaffen sein, dass die Lernenden/Auszubildenden auch die Möglichkeit haben, ihr volles Leistungsvermögen zu präsentieren; es muss (2) ferner gewährleistet sein, dass die Testitems, die hinsichtlich ihrer inhaltlichen Komplexität vergleichend analysiert werden, identische kognitive Prozesse ansprechen.

3.3.3 *Das Prinzip der Kognitiven Taxonomie bzw. Art der kognitiven Beanspruchung*

Insbesondere in Anlehnung an die Arbeiten von Embretson (u. a. 1998; 2002) kann davon ausgegangen werden, dass sich Lern- und Arbeitsanforderungen im Hinblick auf die Art der kognitiven Auseinandersetzung voneinander unterscheiden. Je intensiver die kognitive Erbringungsleistung, umso komplexer und anspruchsvoller wird ein Testitem bewertet. Während sich das Prinzip der funktionalen Modellierung mit den notwendigen Bearbeitungsschritten und Modellierungsleistungen vor dem Hintergrund des Abstraktionsgrades einer Anforderungssituation beschäftigt und das Prinzip der inhaltlichen Komplexität auf spezifische Zusammenhänge und den Umfang von fachlichen Inhaltsbereichen abstellt, werden mit dem Prinzip der kognitiven Taxonomierung Lernfähigkeiten entsprechend der in Lern- und Arbeitsanforderungen angeregten kognitiven Prozesse unterschieden (vgl. hierzu auch Goldman & Pellegrino, 1984; Carroll, 1987).

Unter kognitiver Taxonomierung (cognition) ist mit Rückgriff auf Bloom und Kollegen (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956) die schematische Darstellung des Lernens in der kognitiven Dimension zu verstehen. Die Bedeutung der Bloomschen Lernzieltaxonomie liegt aber nicht allein in der Möglichkeit, Lernen abzubilden und zu beschreiben, sondern insbesondere darin, Ansatzpunkte und Kriterien für Prozesse des Lehrens und des Assessments abzuleiten, die vor dem Hintergrund einer outputorientierten Kompetenzdiskussion geeignet sind, Lern- und Arbeitsanforderungen zu charakterisieren:

„[...] the Taxonomy is perhaps best viewed as a conceptual framework that can be used within virtually any philosophical framework. [...] the Taxonomy Table is best seen as aiding the necessary transition from curriculum to instruction. The Taxonomy Table does not define curriculum; only people can do that“ (Anderson & Krathwohl, 2001, p. 241).

Mittlerweile ist die Taxonomie nach Bloom von zahlreichen Autoren adaptiert, revidiert und weiterentwickelt worden; diese Arbeiten orientieren sich im Hinblick auf ihre gedankliche Struktur sowie ihre methodische Umsetzung größtenteils eng an der 1956 erschienenen *Taxonomy of educational objectives: Handbook I: Cognitive Domain* (eine detaillierte Gegenüberstellung verschiedener Taxonomien findet sich bei Anderson & Krathwohl, 2001, pp. 259-286). Um darzustellen, welchen Einfluss kognitive Prozesse auf den Schwierigkeits- bzw. Komplexitätsgrad von Anforderungssituationen haben und welche Einschränkungen hierbei zu beachten sind, werden im Folgenden zwei aktuelle Taxonomien kurz skizziert: the *Taxonomy for Learning, Teaching, and Assessing* (Anderson & Krathwohl, 2001) sowie the *New Taxonomy of Educational Objectives* (Marzano & Kendall, 2007; 2008). Bei beiden Beispielen handelt es sich um Revisionen der Bloom'schen Taxonomie vor dem Hintergrund der Entwicklungen im Bereich der angewandten Kognitionspsychologie.

The Taxonomy for Learning, Teaching, and Assessing

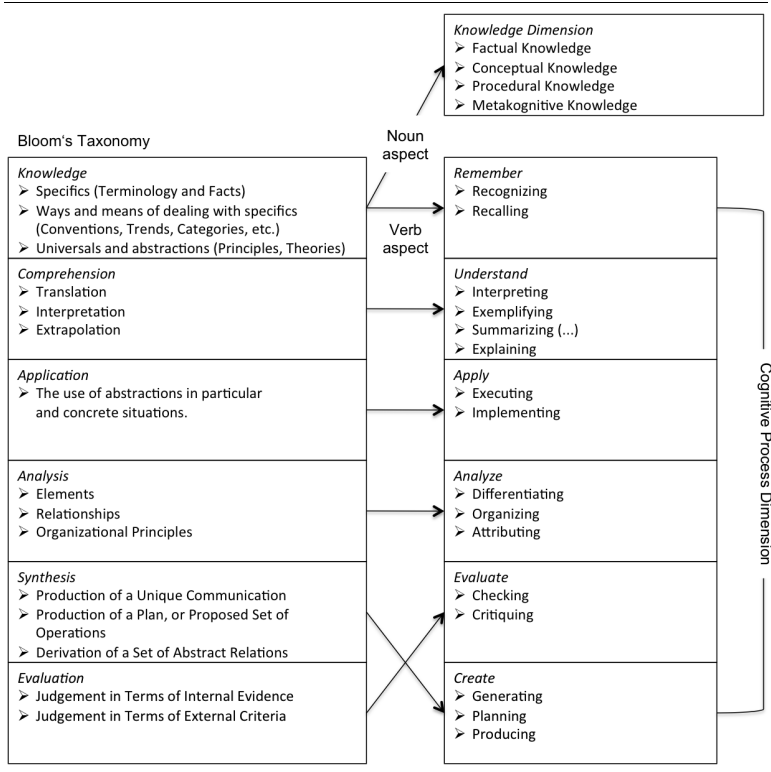
Die Kategorien einer Taxonomie suggerieren kumulative Hierarchien zwischen den Kategorien. In der kategorialen Struktur von Bloom reicht das Kontinuum von „Wissen“ (einfach) bis „Bewerten“ (komplex), wobei Bloom und Kollegen unter der Wissenskategorie und deren Subkategorien insbesondere Verhaltensattribute subsumieren.

„Knowledge as defined here includes those behaviors and test situations which emphasize the remembering, either by recognition or recall, of ideas, materials or phenomena“ (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956, p. 62).

In der Revision von Anderson und Krathwohl (2001) wird die Wissenskategorie der Hierarchie entnommen und als eigene Dimension, die sich auf den Inhalt des Lehrens und Lernens bezieht, den kognitiven Prozessen gegenüber gestellt (vgl. Abbildung 3.16). Dies ist die wohl augenscheinlichste Veränderung im Vergleich zur Originaltaxonomie. Insgesamt werden von Anderson und Krathwohl zwölf substantielle Unterschiede formuliert: jeweils vier im Hinblick auf die theoretische Akzentuierung, die Terminologie und die Struktur (Anderson & Krathwohl, 2001, pp. 305-312). Die Revision betont (1) die Nutzung der Taxonomie für Prozesse des Lernens und Lehrens unter Berücksichtigung der Assessmentkomponente, während im Original Fragen des Assessments das stärkste Gewicht haben und Aspekte der Instruktion sehr randständig bleiben. Eng damit verbunden ist, dass sich die Revision (2) gezielt an Lehrende richtet und ihnen mit der Taxonomie ein Hilfsmittel der Unterrichtsplanung und -analyse zur Verfügung stellt. Dies geschieht, indem (3) zahlreiche Beispiele gegeben werden, mit deren Hilfe die verschiedenen Kategorien der Wissensdimension und der Dimension der kognitiven Prozesse illustriert werden. Hierfür ist (4) eine genaue Definition der einzelnen Subkategorien notwendig. In der Revision werden unter der

Prämisse der Unterrichtstauglichkeit die Hauptkategorien über die einzelnen Subkategorien definiert. Dies hat den Vorteil, dass curriculare Übersetzungsleistungen direkt in die instruktionale Unterrichtsgestaltung einfließen können.

Abbildung 3.16 *Bloom's Taxonomie und deren Revision*
(Anderson & Krathwohl, 2001, p. 310)



Diese Änderungen in der konzeptionellen Ausrichtung der Taxonomie werden in der Terminologie konsequent aufgegriffen. So werden (5) die Wissenskategorien als Bezugselement zum Unterrichtsinhalt als Substantive und (6) die Kategorien der kognitiven Prozesse als Abbildung des Lehr- und Lernvorgangs als Verben in die Taxonomie eingebunden. Dabei erfolgt (7) eine Umbenennung und Reorganisation der Subkategorien der Wissensdimension in Anlehnung an aktuelle Entwicklungen der Kognitionswissenschaft. Als Subkategorien werden aufgenommen: Factual knowledge, conceptual knowledge, procedural knowledge sowie metacognitive knowledge. Diese Einteilung entspricht der in der Literatur sehr verbreiteten Unterscheidung zwischen declarative, procedural und conditional/interpretative knowledge (vgl. Gelman & Greeno, 1989; Greeno, 1998;

Byrnes, 2008).⁶ Alexander, Schallert und Hare (1991) haben die verschiedenen Terminologien des Wissensbegriffs vergleichend analysiert. Im Ergebnis lassen sich aus den Variationen der Begrifflichkeiten Wissenskategorien identifizieren, die (a) sach- und konzeptbezogen auf einen spezifischen Inhalt bezogen sind (deklaratives Wissen), die (b) Handlungen, Prozeduren und Arbeitstechniken zur Bearbeitung eines spezifischen Inhaltsbereiches umfassen (prozedurales Wissen) und mit deren Hilfe (c) ein spezifischer Inhaltsbereich im Hinblick auf die einzusetzenden Wissensbestände und Fähigkeiten interpretiert und in der Folge als Sequenz von entscheidungsbasierten Handlungen erfasst werden kann (konditionales bzw. strategisches Wissen).

„Conditional knowledge requires that individuals evaluate the context and bring their declarative and procedural knowledge to bear as required by the conditions or circumstances“ (Murphy & Alexander, 2006, p. 41).

Im Hinblick auf die Differenzierung der kognitiven Prozesse enthält die Revision zwei Umbenennungen: Comprehension wird zu Understand und Synthesis zu Create. Mit diesen Veränderungen in der Terminologie werden (8) die Arbeiten im Bereich des Problemlösens sowie des kritischen Denkens implementiert.

Die Struktur der Taxonomie weist in diesem Zusammenhang (9) auch einen Wechsel in der Reihenfolge der kognitiven Prozesse auf: Synthesis/Create und Evaluation/Evaluate sind in ihren Positionen vertauscht. Die Struktur der revidierten Taxonomie ist darüber hinaus (10) durch die Unterscheidung von Wissensdimension und Dimension des kognitiven Prozesses geprägt, die (11) in einer zweidimensionalen Darstellungsform der Taxonomie zum Ausdruck kommt. Hierbei wird angenommen, dass (12) die kognitiven Prozesse von unterschiedlicher Komplexität sind. Die Bloomsche Taxonomie geht von einer kumulativen Hierarchie aus, die besagt, dass die Bewältigung eines höheren kognitiven Prozesses die Bewältigung niederer kognitiver Prozesse voraussetzt, und in der davon ausgegangen wird, dass die kognitiven Prozesse voneinander eindeutig separiert werden können. Im Hinblick auf die Konstruktion von Testitems bedeutete dies, dass sich unter Rückgriff auf kumulativ hierarchisch strukturierte kognitive Prozesse Handlungsmodelle für spezifische Inhaltsbereiche differenzieren ließen, die als Niveaustufen im Rahmen von Kompetenzmodellen zu interpretieren wären (vgl. u. a. Blum, Neubrand, Ehmke, Senkbeil, Jordan, Ulfig & Carstensen, 2004). Dies wäre möglich, wenn die ranghöheren Taxonomieklassen bzw. Prozessstufen die Beherrschung bzw. das Durchlaufen der darunterliegenden kumulativ implizieren würden und insofern einen höheren Schwierigkeits- bzw. Komplexitätsgrad aufwiesen. Diese linearen Entwicklungsverläufe können jedoch von den empirischen Befunden nicht belegt werden. Kreitzer und Madaus (1994) zei-

⁶ Diese Dreiteilung der Fähigkeitsstruktur findet über die Kategorien conceptual competence, procedural competence und interpretative competence auch im Rahmen des Kompetenzmodells für die kaufmännische Bildung Berücksichtigung (vgl. hierzu Kapitel 2.4.2).

gen im Rahmen einer Metaanalyse, dass sich eine kumulative Hierarchie nur für den mittleren Bereich der kognitiven Prozesse annehmen lässt; für die niederen und höheren kognitiven Prozesse liegen hierfür jedoch keine Befunde vor. Witt (2006; Müller, Fürstenau & Witt, 2007) zeigt für die kaufmännische Ausbildung, dass Prozesse der Anwendung teilweise als einfacher empfunden werden als Prozesse des Verstehens und dass sich die einzelnen Prozesskategorien nicht konsequent voneinander unterscheiden lassen. In der Revision der Bloomschen Taxonomie wird in diesem Zusammenhang darauf verwiesen, dass von einer tendenziellen Zunahme an Komplexität in der Dimension der kognitiven Prozesse auszugehen ist, die jedoch (a) nicht hierarchisch kumulativ angelegt ist, die (b) auch Überschneidungen zwischen den einzelnen kognitiven Prozessen zulässt und die (c) vorrangig aus der instruktionalen Arbeit des Lehrenden resultiert.

Eine Adaption der Taxonomie von Anderson und Krathwohl wurde beispielsweise im Rahmen der ULME III-Studie (Untersuchung von Leistungen, Motivation und Einstellungen in der beruflichen Bildung; Lehmann & Seeber, 2007) zur Klassifizierung des kognitiven Anspruchsniveaus von Aufgabensätzen herangezogen (vgl. Brand, Hofmeister & Tramm, 2005; Hofmeister, 2005). Auch im DFG-Projekt „Integrierte Kompetenzförderung in den beruflichen Fächern des Wirtschaftsgymnasiums“ (Achtenhagen & Winther, 2006; Winther, 2006) sind auf Basis der Gegenüberstellung von Wissensdimension und Dimension der kognitiven Prozesse Testitems zur Erfassung wirtschaftsberuflichen Vorwissen konstruiert und im Hinblick auf die angeregten kognitiven Verarbeitungsprozesse ausgewertet worden. Erfolgt die Konstruktion von Testitems jedoch unter der Prämisse einer zunehmenden Komplexität der kognitiven Beanspruchung ist unter Bezugnahme auf die Taxonomie von Bloom und deren Revisionen Folgendes zu bedenken: Sowohl Bloom's Taxonomie (Bloom et al., 1956) als auch die Revision von Anderson und Krathwohl (2001) beschreiben generelle Kategorien der Informationsverarbeitung; sie sind jedoch nicht dazu geeignet, Leistungs- bzw. Lösungsverhalten in spezifischen Lern- und Arbeitsanforderungen vorherzusagen (vgl. Rohwer & Sloane, 1994). Hierfür sind theoretische Entwicklungen notwendig, die das hierarchische System der kognitiven Auseinandersetzung auf die Verarbeitungs- und Lösungsprozesse spezifischer Lern- und Arbeitsanforderungen transformieren (Marzano & Kendall, 2007, p. 17).

The New Taxonomy of Educational Objectives

Die Autoren der New Taxonomy of Educational Objectives (Marzano & Kendall, 2007; 2008) greifen grundlegend auf die gleichen Vorstellungen zurück, die auch der Taxonomie von Bloom (Bloom et al., 1956) und deren Revision von Anderson und Krathwohl (2001) zugrunde liegen. Die Taxonomien folgen den Vorschlägen Tylers (1946):

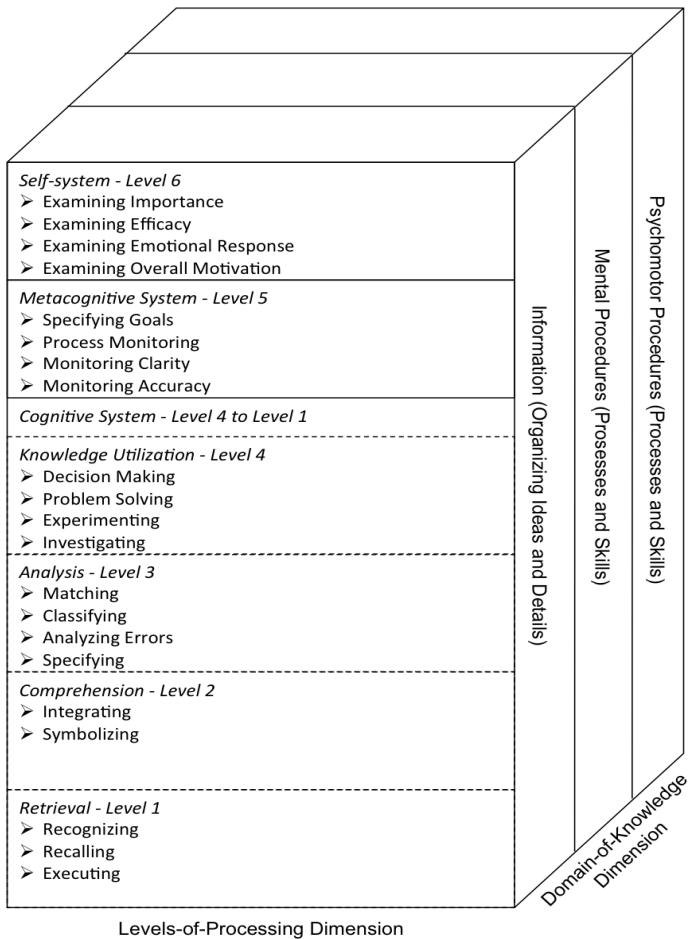
„The most useful form for stating objectives is to express them in terms which identify the kind of behavior to be developed in the student and

the content [...] in which the behavior is to operate" (Tyler, 1946, zit. n. Marzano & Kendall, 2007, p. 17).

Ein Vergleich zwischen der Bloomschen Taxonomie (einschließlich der Revisionen) und der Taxonomie nach Marzano und Kendall zeigt allerdings zwei zentrale Unterschiede auf: (1) Die Einbindung und die Struktur der kognitiven Prozesse erfolgt nicht im Hinblick auf eine zunehmende Schwierigkeit, sondern wird über kognitive Bewertungs- und Kontrollprozesse erfasst. (2) Für die Berücksichtigung der Wissensdimension werden spezifische Wissensbereiche im Sinne von Verarbeitungs- und Handlungsmodellen definiert (Abbildung 3.17). Ad (1): Generell lassen sich mit Hilfe der Taxonomien Lern- und Arbeitsanforderungen klassifizieren, indem die Art des Wissens und die kognitiven Prozesse identifiziert werden, auf die die Instruktion hin ausgerichtet ist. Marzano und Kendall greifen hierfür auf kognitionspsychologische Modelle der Informationsverarbeitung sowie auf Modelle mentaler Prozesse zurück und binden die Annahmen eines zunehmenden Komplexitätsgrades der kognitiven Prozesse über diese Modelle in ihre Taxonomie ein. Kognitive Prozesse unterscheiden sich demnach hinsichtlich zweier Faktoren: Die Komplexität der kognitiven Prozesse wird bestimmt (a) über die Anzahl der eingebundenen und miteinander interagierenden Bearbeitungsschritte und (b) über den Grad der Vertrautheit der zugrunde liegenden Lern- und Arbeitsanforderung. Marzano und Kendall schlussfolgern daraus, dass sich kognitive Prozesse hierarchisch nicht nach ihrer Schwierigkeit, sondern nach unterschiedlichen Kontrollprozessen ordnen lassen:

„Although mental processes cannot be ordered hierarchically in terms of difficulty, they can be ordered in terms of control: Some processes exercise control over the operation of other processes“ (Marzano & Kendall, 2007, p. 11).

Zur Klassifizierung dieser Kontrollprozesse basiert die Taxonomie auf drei mentalen Systemen: das System des Selbst (self-system), das metakognitive System und das kognitive System. Diese Systeme beschreiben unterschiedliche Stufen des Bearbeitungs- bzw. Lösungsverhaltens (levels-of-processing dimension). Das System des Selbst entscheidet über Art und Umfang der Auseinandersetzung mit den spezifischen Lern- und Arbeitsanforderungen. Hierfür sind kognitive Bewertungen der Anforderungssituation notwendig, die sich auf motivationale, emotionale und volitionale Aspekte des Lernens beziehen lassen (vgl. hierzu Winther, 2006). Über das System des Selbst wird dem Umstand Rechnung getragen, dass ein ausgeprägtes Verstehen von Begriffen und Konzepten den Grad der Auseinandersetzung mit einer konkreten Anforderungssituation zwar fördert, dass jedoch selbst bei einer ausreichenden Wissensbasis eine erfolgreiche Umsetzung des Wissens in konkrete Handlungen scheitern kann, wenn die dazu erforderlichen motivationalen und volitionalen Vermittlungsleistungen nicht bestehen (vgl. dazu auch Artelt, Demmrich & Baumert, 2001).

Abbildung 3.17 *The New Taxonomy (Marzano & Kendall, 2007, p. 13)*

Im Hinblick auf die Entwicklung von Testinstrumenten sind in diesem Zusammenhang die Anstrengungen der Forschungen zu sehen, Testinhalte möglichst authentisch im Hinblick auf reale und persönlich bedeutsame Lern- und Arbeitsanforderungen zu gestalten und Testformate zu wählen, die medial vielseitig emotionale und motivationale Aufmerksamkeit erzeugen können. Das metakognitive System stellt die Verbindung zwischen dem System des Selbst und dem kognitiven System dar. Metakognition beschreibt das Wissen und die Kontrolle über das eigene kognitive System, die Auseinandersetzung mit eigenen ko-

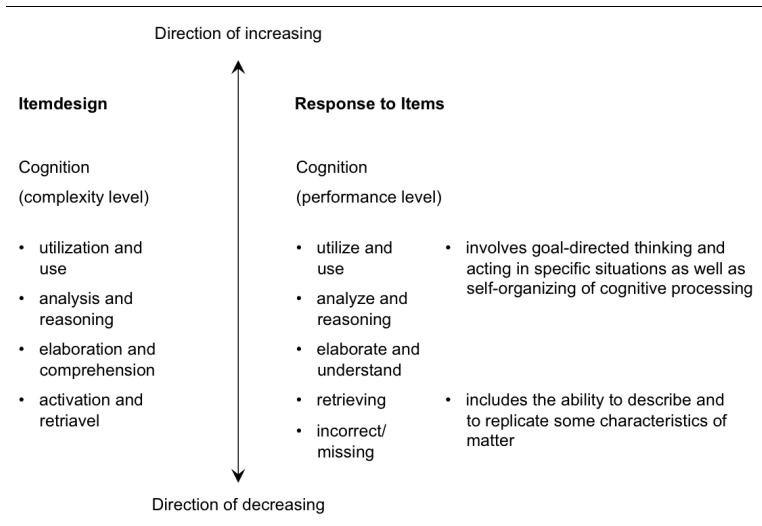
gnitiven Prozessen und den Versuch, diese zu verstehen und mit diesem Verständnis sinnvoll umzugehen (Hasselhorn, 1998; 2000). Metakognition ist damit eng mit der „Bewusstheit“, d. h. mit einer wesentlichen Steuerungskomponente des Handelns verknüpft. Diese impliziert, dass das Handeln und die Anstrengung zu lernen willentlich beeinflusst und damit gelenkt werden können. Dieser Kontrollprozess ist strikt kontextspezifisch zu interpretieren. Die Lernenden/Auszubildenden nehmen die Lern- und Arbeitsanforderung über verschiedene Prozesse des Monitorings – Aufgabenformulierung, Schwierigkeitsgrad, Präsentationsform etc. – wahr, überwachen und evaluieren sie (perception, monitoring, and evaluation of context) (Pintrich, 2000; 2004). Durch die Prozesse des Monitoring interagiert das metakognitive System permanent mit dem kognitiven System. Das kognitive System bezieht sich auf die kognitiven Prozesse, die die Art der kognitiven Beanspruchung im Rahmen einer spezifischen Lern- bzw. Arbeitsanforderung beschreiben. Hier sind die Parallelen zur Taxonomie von Bloom und deren Revisionen am deutlichsten: Das kognitive System besteht aus vier Subsystemen, die ausgehend vom Abrufen spezifischer Wissensbestände und prozeduraler Fähigkeiten Prozesse des Verstehens, der Analyse und der Verwertung umfassen. Eine wesentliche Neuerung ist, dass der Prozess der Evaluation in der Taxonomie nicht als eigenständiger kognitiver Prozess enthalten ist, sondern in die vier Subkategorien jeweils integriert wurde.

Ad (2): Als vierte, von den Kontrollprozessen separierte Komponente wird die Wissensdimension aufgenommen. Wissen wird über drei generelle Kategorien organisiert: Information sowie mentale und psychomotorische Prozeduren. Diese Kategorien werden als Verarbeitungs- bzw. Handlungsmodelle beschrieben, die den mentalen Systemen zugrunde liegen und aus diesen resultieren können (domains-of-knowledge dimension).

Der Bereich der Information wird als deklaratives Wissen und der Bereich der mentalen Prozeduren als prozedurales Wissen verstanden. Unter psychomotorischen Prozeduren verstehen Marzano und Kendall physische Aktivitäten zur Bewältigung von Anforderungssituationen – für Arbeitsprozesse können so Prozeduren identifiziert werden, die ähnlich wie mentale Prozeduren im Gedächtnis gespeichert sind und die als Handlungswissen in spezifischen Anforderungssituationen abgerufen werden können (vgl. Anderson, 1983; 1993). Mit dieser Art der Aufbereitung von Wissenskategorien wird deutlich, dass Wissen über spezifische Zugriffsmuster, mit denen Lernende/Auszubildende Inhaltsbereiche erfassen, definiert wird. Diese Zugriffe implizieren eine strukturierte, jedoch *nicht in ihrer Komplexität variierende Wissensbasis*, aus der heraus die kognitiven Prozesse in Abhängigkeit von der gegebenen Anforderungssituation generiert werden. Dabei ist weniger entscheidend, in welcher Terminologie verschiedene Wissensvarianten in den Taxonomien aufgenommen werden; entscheidend ist, dass die Wissenskategorien als Hilfsmittel genutzt werden können, um Lern- und Arbeitsanforderungen zu charakterisieren und aufzuzeigen, über welches Wissen ein Lernender/Auszubildender im Hinblick auf eine zugrunde liegende Anforderungssituation verfügen sollte, um diese Anforderungssituation bewältigen zu können.

Wissensdimension und Dimension der kognitiven Prozesse stellen Kriterien zur Verfügung, mit denen sich Anforderungssituationen charakterisieren lassen. Um jedoch im Rahmen der Itemkonstruktion den Schwierigkeits- bzw. Komplexitätsgrad einer Lern- und Arbeitsanforderung a priori zu schätzen, sind diejenigen Elemente der kognitiven Taxonomien zu identifizieren, die sich möglichst trennscharf auf einem Kontinuum von geringer bis hoher kognitiver Beanspruchung verorten lassen (vgl. Abbildung 3.18).

Abbildung 3.18 Kognitive Taxonomierung



Während sich die verschiedenen Wissensarten ausschließlich im Hinblick auf ihre Zugriffsoptionen unterscheiden, stellen die vier Subkategorien des kognitiven Systems unter Rückgriff auf kognitive Modelle differenzierbare Verarbeitungs- und Handlungsprozesse dar, die sich verschiedenen Komplexitätsstufen zuordnen lassen: (1) die Aktivierung und der situationsspezifische Abruf von Wissen, (2) die Elaboration von Wissen und damit die Darstellung von Verstehensprozessen, (3) die schlussfolgernde Auslegung von Wissen im Hinblick auf die Plausibilität einer situationsspezifischen Lösung und (4) die Verwertung von Wissen beispielsweise im Rahmen von Problemlösungen oder Transferaufgaben. Um Aussagen über das Schwierigkeitsniveau der Testitems treffen zu können, sind diese itemcharakteristischen Bedingungen der kognitiven Beanspruchung wiederum auf einer gemeinsamen Skala mit dem erwarteten Antwortverhalten der Lernenden/Auszubildenden abzutragen. Die Leistungsfähigkeiten der Lernenden/Auszubildenden können demnach zwischen der Fähigkeit, Wissensbestände situationsangemessen abzurufen, und der Fähigkeit, Bearbeitungsprozesse eigenständig zu steuern, variieren. Auf der Stufe geringer Leistungsfähigkeit

bedeutet dies, dass relevantes Wissen im Hinblick auf einen spezifischen Inhaltsbereich identifiziert werden kann; auf der Ebene der höchsten Leistungsfähigkeit können die identifizierten Wissensbestände und prozeduralen Fähigkeiten im Sinne eines Verwertungsgedankens in den Bearbeitungsprozess eingebracht werden, um neue Erkenntnisse zu erzeugen und bestehende Erkenntnisse zu erweitern.

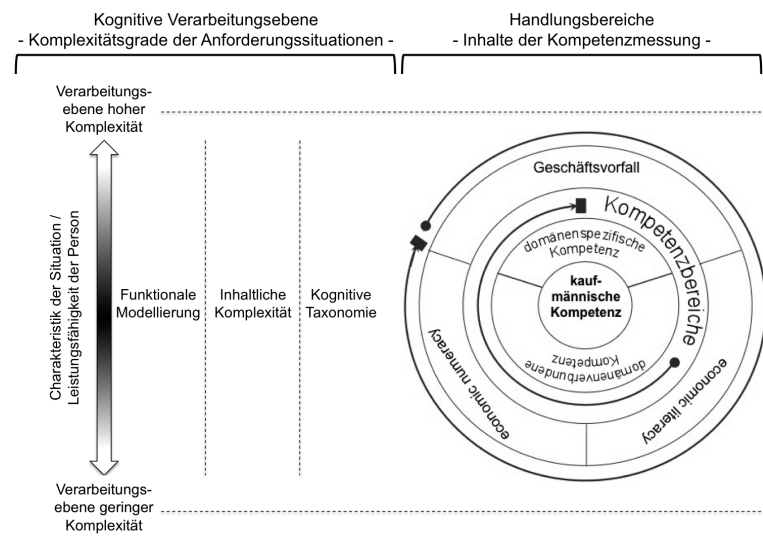
3.3.4 Modell der Instrumentenentwicklung für die kaufmännische Bildung

Im Modell der Instrumentenentwicklung für die kaufmännische Bildung werden Komplexitätsgrade der Anforderungssituationen auf die Inhalte der Kompetenzmessung bezogen. Über die Handlungsbereiche, die sich als Kompetenzstrukturen interpretieren lassen, wird die inhaltlich-fachdidaktische Ebene der Kompetenzmodellierung (vgl. hierzu die Ausführungen zum Kompetenzstrukturmodell Kapitel 2.4.2) mit der kognitiven Verarbeitungsebene kombiniert, über die einerseits die Charakteristik der Lern- und Arbeitssituationen beschrieben werden kann und die andererseits eine Schätzung der Leistungsfähigkeit der Lernenden/Auszubildenden ermöglicht (Abbildung 3.19). Dem Kompetenzstrukturmodell für die kaufmännische Bildung folgend, lassen sich die Handlungsbereiche jeweils über Handlungsabfolgen beschreiben: über die Bearbeitung einer definierten konkreten Lern- und Arbeitsanforderung durch aktivierte deklarative Wissensbestände (conceptual competence) vor dem Hintergrund eines ganzheitlichen und funktionalen Verständnisses zentraler domänenbezogener Ideen, über die Selektion und Ausführung domänenbezogener Handlungen und Prozeduren (procedural competence) sowie über die kognitive Erfassung der Anforderungssituation und die angemessene Interpretation und Validation der gefundenen Lösung (interpretative competence).

Die Beziehung zwischen kognitiver und inhaltlicher Ebene wird durch Kriterien begründet, die systematische Variationen des Anspruchsniveaus der Anforderungssituationen und damit Variationen der Lösungswahrscheinlichkeiten für die Bewältigung dieser Anforderungssituationen zulassen: funktionale Modellierung, inhaltliche Komplexität und kognitive Taxonomierung. Ein wesentliches Element ist, dass diese Kriterien aus kognitiven Modellen des Lehrens und Lernens hergeleitet wurden; sie fokussieren auf interne Verarbeitungsprozesse, von denen angenommen werden kann, dass sie sich (a) empirisch trennen lassen und sich (b) im Hinblick auf die Verarbeitung von Wissensbeständen und prozeduralen Fähigkeiten substantiell unterscheiden. Diese Bedingungen gelten als maßgeblich für die Beschreibung von kontextualisierten Lern- und Arbeitsanforderungen zur Erfassung von Kompetenzen in verschiedenen Kompetenzbereichen. Während also über die kognitive Verarbeitungsebene ein Kontinuum verschiedener Schwierigkeits- bzw. Komplexitätsgrade abgebildet wird, werden auf der inhaltlich-fachdidaktischen Ebene verschiedene inhaltliche Testbereiche und die Kompetenzstrukturen dargestellt. Damit wird dem Umstand Rechnung getragen, dass Kompetenz in einer Domäne keine eindimensionale kognitive Disposition ist und es folglich im Hinblick auf Kompetenzerwerb und -entwicklung nicht genügen

wird, nur auf eine spezifische kognitive Struktur zu fokussieren (Kilpatrick, Swafford & Findell, 2001; Brand, Hofmeister & Tramm, 2005).

Abbildung 3.19 Modell der Instrumentenentwicklung für die kaufmännische Bildung



Für die Betrachtung von Kompetenzen aus Perspektive inhaltlicher Handlungsbereiche ergibt sich folglich, dass die Kompetenzstruktur bezogen auf die drei Inhalte economic literacy, economic numeracy und Geschäfts-vorfall *prozessual und nicht hierarchisch* ist. Dies impliziert, dass keine Zunahme an Komplexität im Hinblick auf die Kompetenzstrukturen angenommen wird; die verschiedenen Kompetenzstrukturen charakterisieren ausschließlich (1) den Zugriff auf spezifische Inhaltsbereiche und stellen (2) mögliche sich unterscheidende Fähigkeitsstrukturen dar, die zur Anforderungsbewältigung einzusetzen sind.

Die Kompetenzstruktur ist im Hinblick auf die verschiedenen Handlungsbereiche als kohärentes Ganzes ausgestaltet, um domänentypische Lern- und Arbeitsanforderungen repräsentieren zu können. Testitems, die in den verschiedenen Handlungsbereichen verortet sind, sind folglich so zu konstruieren, dass die differenzierte kognitive Struktur im Sinne von Handlungsabfolgen auch tatsächlich getestet werden kann. Im Hinblick auf die psychometrische Modellierung lassen sich durch die Unterscheidung zwischen kognitiver Verarbeitungsebene und Inhaltsebene insbesondere zwei Annahmen vorgeben: Während die kognitive Bearbeitungsebene die Kalibrierung von Testitems stützt und Einfluss auf die Schätzung von Itemschwierigkeiten und Personenfähigkeiten nimmt, werden über die Inhalts- und Zugriffsebene die Dimensionen der kaufmännischen Kompetenz abgebildet.

4 Analysing Measures: Einführung in die Item Response-Theorie

Die im Rahmen der Konstruktrepräsentation (Kapitel 3.3) getroffenen Annahmen stellen heuristische Zusammenhänge zwischen den Testitems und den Personenmerkmalen dar. So wird beispielsweise angenommen, dass sich Testitems im Hinblick auf ihr Anspruchsniveau unterscheiden und dass ein höheres Anspruchsniveau zu niedrigeren Lösungswahrscheinlichkeiten führt. Umgekehrt gilt, dass Personen, die eine hohe Fähigkeit aufweisen, Testitems mit einer höheren Wahrscheinlichkeit lösen können als Personen mit einer geringer ausgeprägten Fähigkeit. Diese Beziehungen lassen sich im Hinblick auf kognitive Theorien plausibel erklären. Im folgenden Kapitel werden diese Zusammenhänge unter Rückgriff auf die probabilistische Testtheorie (Item Response-Theorie) mathematisch beschrieben. Der Fokus liegt dabei auf Fragen der Anwendbarkeit der Theorie und der Auswahl geeigneter Testmodelle. Bei komplexen Tests und bei angenommenen komplexen Zusammenhängen von Testverhalten und zu erfassendem Personenmerkmal erscheint es nicht sinnvoll, ein einziges generelles Testmodell zu wählen, mit dessen Hilfe sich alle angenommenen Zusammenhänge beschreiben lassen. Das Ziel ist es eher, solche Modelle zu entwickeln, die auf einzelne zentrale Zusammenhänge abstellen und die die Empirie in Ausschnitten vereinfacht beschreiben. Die im Folgenden dargestellten Modelle werden im Rahmen von Kompetenzmessungen international eingesetzt. Die Auswahl der Modelle beschränkt sich dabei auf jene, deren Eigenschaften insbesondere für Large Scale-Assessments bedeutsam sind und die auch im Rahmen der Studien zur Erfassung kaufmännischer Kompetenz (vgl. dazu Kapitel 5) zum Einsatz gekommen sind. Das folgende Kapitel ist als Einführung konzipiert. Es werden die allgemeinen Annahmen und Einsatzmöglichkeiten der Item Response-Theorie aufgezeigt. Dies erfordert zum Teil auch einen Rückgriff auf die mathematischen Notationen; auf die mathematische Beweisführung der Gültigkeit der Modelle wird hingegen weitgehend verzichtet. Diese Einführung richtet sich an Leser, die über Grundkenntnisse der Klassischen Testtheorie verfügen und sich über die Möglichkeiten der probabilistischen Testtheorie vor dem Hintergrund der Verfahren und Methoden von Kompetenzmessungen informieren möchten. Für eine tiefer gehende Auseinandersetzung wird auf folgende Literatur verwiesen:

- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. 2. Aufl. Bern: Huber.
- Wright, B. D. & Stone, M. H. (1979). *Best Test Design*. Chicago, IL: MESA Press.

Ausgehend von den Hintergründen und den Intuitionen von Messungen im Bereich von Kompetenzen und Leistungsfähigkeiten werden verschiedene psychometrische Modelle beschrieben, die Schätzungen der Item- und Personenparameter erläutert und zentrale Erweiterungen der grundlegenden Theorie skizziert.

4.1 Personenmerkmal, Testitem und Messwert: Zur Aussagekraft von Tests

Die folgenden Ausführungen beziehen sich auf kognitive Tests und auf die grundlegende Überlegung, wie Beobachtungen von Testleistungen in Messwerte der Leistungsfähigkeit übersetzt werden können. Unter einem *Test* versteht man:

„[...] ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (Lienert, 1969, S. 7).

Das *Merkmal* beschreibt folglich das interessierende Untersuchungsziel. Im Rahmen kognitiver Tests können das spezifische Leistungsfähigkeiten, Kompetenzen und Fertigkeiten sein. Die Beschreibung dieses Persönlichkeitsmerkmals basiert auf beobachtbaren Verhaltensweisen, die sich in der Bearbeitung von *Testitems* zeigen, von denen angenommen werden kann, dass sie das interessierende Merkmal repräsentieren. Das Testverhalten stellt damit das manifeste Kriterium dar, mit dessen Hilfe auf das zu untersuchende Personenmerkmal zurückgeschlossen werden kann. Das Personenmerkmal selbst wird als latent hypothetisches Konstrukt definiert, da es sich einer direkten Beobachtung entzieht und seine Beschreibung von dem beobachtbaren Testverhalten abhängig ist (vgl. Rost, 2004, S. 28). Die Beschreibung und Zuordnung des Testverhaltens auf das zu untersuchende Personenmerkmal geschieht über die Quantifizierung der Merkmalsausbildungen. Erfolgt die Quantifizierung nach bestimmten mathematischen Regeln, wird in diesem Zusammenhang von *Messen* gesprochen; der dem Probanden zugeordnete quantifizierte Merkmalswert ist der *Messwert*.

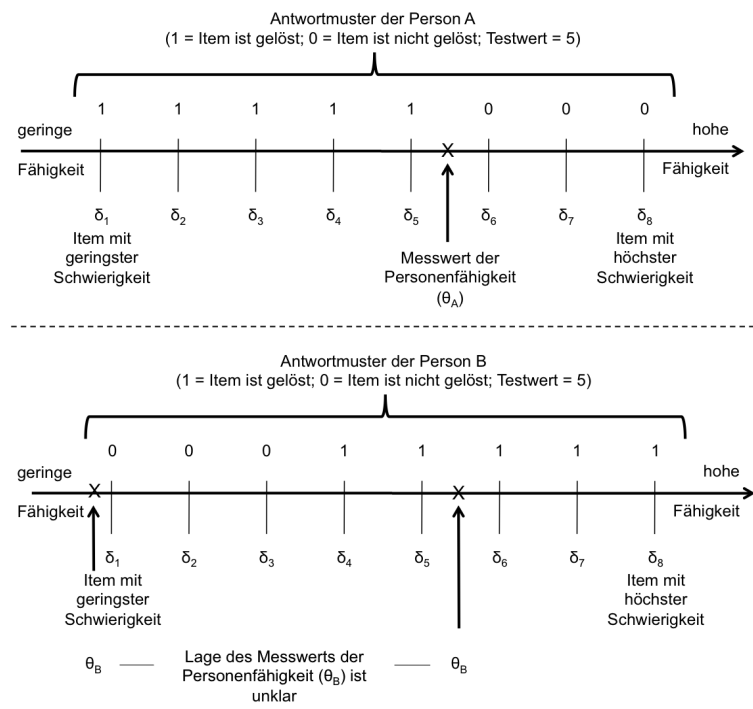
Die Methoden, mit denen im Rahmen kognitiv psychologischer Tests die Messung erfolgt, werden als psychometrisch bezeichnet; äquivalent dazu sind *psychometrische Modelle*, jene Messmodelle, die die generierten Messwerte zusammenfassen und auf das zu untersuchende Personenmerkmal transferieren können. Psychometrische Modelle, unabhängig davon, ob es sich um Modelle der Klassischen Testtheorie oder der Item Response-Theorie handelt, dienen folglich dazu, das in Testsituationen beobachtbare Verhalten in interpretierbare Outputs zu übersetzen, indem mathematische Regeln darüber aufgestellt werden, wie sich die Wahrscheinlichkeiten, eine konkrete Anforderungssituation zu bewältigen oder in dieser zu scheitern, auf die Fähigkeit, Kompetenz oder Fertigkeit des Probanden beziehen lassen.

Modelle der Klassischen Testtheorie und der Item Response-Theorie unterscheiden sich exakt in der Ausformulierung dieser mathematischen Regeln und folglich darin, wie sich die Beziehungen zwischen den Messwerten und dem Personenmerkmal einerseits und die Zusammenhänge zwischen Testitem und Personenmerkmal andererseits gestalten.

4.1.1 Validierung des Testverhaltens

Die Validierung des Testverhaltens bezieht sich auf die Beschreibung der Beziehungen zwischen Personenmerkmal, Testitem und Messwert. Für die folgenden Ausführungen wird angenommen, dass das Personenmerkmal als latentes eindimensionales Konstrukt hinreichend definiert ist und sich in der Tradition kognitiver Tests auf einem Kontinuum zwischen geringer und hoher Ausprägung abbilden lässt. Ein Beispiel hierfür ist die economic literacy-Kompetenz. Ferner sei angenommen, dass acht Testitems konstruiert worden sind, mit denen das zu untersuchende Personenmerkmal erfasst werden kann. Es liegen folglich acht Testitems vor, die eine reliable Skala der economic literacy-Kompetenz abbilden. Diese Testitems unterscheiden sich in ihrer Schwierigkeit (δ_i). Items mit hohem Anforderungscharakter beschreiben das Personenmerkmal auf Seiten der hohen Ausprägung; Items, die selbst von Personen mit geringer Ausprägung des Personenmerkmals gelöst werden können, definieren die gegenüberliegende Seite. Jedes δ_i markiert folglich die Lage eines Testitems auf dem Kontinuum des Personenmerkmals (vgl. Abbildung 4.1).

Abbildung 4.1 Validierung des Testverhaltens (in Anlehnung an Wright & Stone, 1979, p. 9)



Um auf diesem Kontinuum des Personenmerkmals nun eine spezifische Person zu verorten, ist diese Person mit den Items, über die das Personenmerkmal repräsentiert ist, zu testen. Das Antwortverhalten ausgedrückt als kognitive Fähigkeit der Person (θ_v) markiert die Lage der Person auf dem Kontinuum des zu untersuchenden Personenmerkmals. Um also beispielsweise zu erfahren, wie stark die Kompetenz im Bereich der economic literacy bei einer Person ausgeprägt ist, werden Informationen darüber benötigt, wie die Person die Testitems bearbeitet und wie sich die Testwerte/Scores in den Messwert der Personenfähigkeit überführen lassen (vgl. u. a. die Beispiele in Wright & Masters, 1982).

Wie Antwortmuster der Probanden die Aussagekraft von Tests beeinflussen

Ein intuitiv logischer Zusammenhang zwischen Testitem und Messwert der Personenfähigkeit ist der, dass die Fähigkeit einer Person θ_v als Ort auf dem Kontinuum des Personenmerkmals ausgedrückt werden kann, an dem gleichsam auch das schwierigste von der Person gelöste Item und das erste von der Person nicht gelöste Item platziert sind. In Abbildung 4.1 liegt dieser Ort für die Person A zwischen den Items, die über die Itemschwierigkeiten δ_5 und δ_6 beschrieben sind. Dabei wird davon ausgegangen, dass Items mit einem leichteren Schwierigkeitsgrad als δ_6 von der Person gelöst werden können, Items ab einem Schwierigkeitsgrad von δ_6 jedoch nicht. Das Beispiel in Abbildung 4.1 zeigt solch ein logisches Antwortmuster – die Person A hat die ersten fünf Items lösen können, die drei schwierigsten Items hingegen nicht.

In der praktischen Anwendung von Tests zur Messung von Leistungsfähigkeiten zeigt sich dieses logische Antwortmuster jedoch eher selten und nicht in dieser Eindeutigkeit. Ein Extremfall ist ebenfalls in Abbildung 4.1 dargestellt. Die Person B hat wie die Person A fünf Items lösen können. Es handelt sich hierbei jedoch um Items mit hohem Schwierigkeitsgrad. In diesem Zusammenhang stellt sich die Frage, wo genau die Fähigkeit der Person auf dem Kontinuum des Personenmerkmals zu verorten ist. Oder anders formuliert: Wie stark ist die Kompetenz im Bereich der economic literacy der Person B tatsächlich ausgeprägt? Jede Position auf der Skala ist unbefriedigend und benötigt eine differenzierte Erklärung. Insgesamt kann man feststellen, dass der Test für die Person B kein aussagekräftiges Ergebnis liefern kann. Dies könnte einerseits daran liegen, dass die Items nicht korrekt kalibriert sind, oder andererseits daran, dass die Person B den Test nicht so bearbeitet hat, wie von den Entwicklern intendiert.

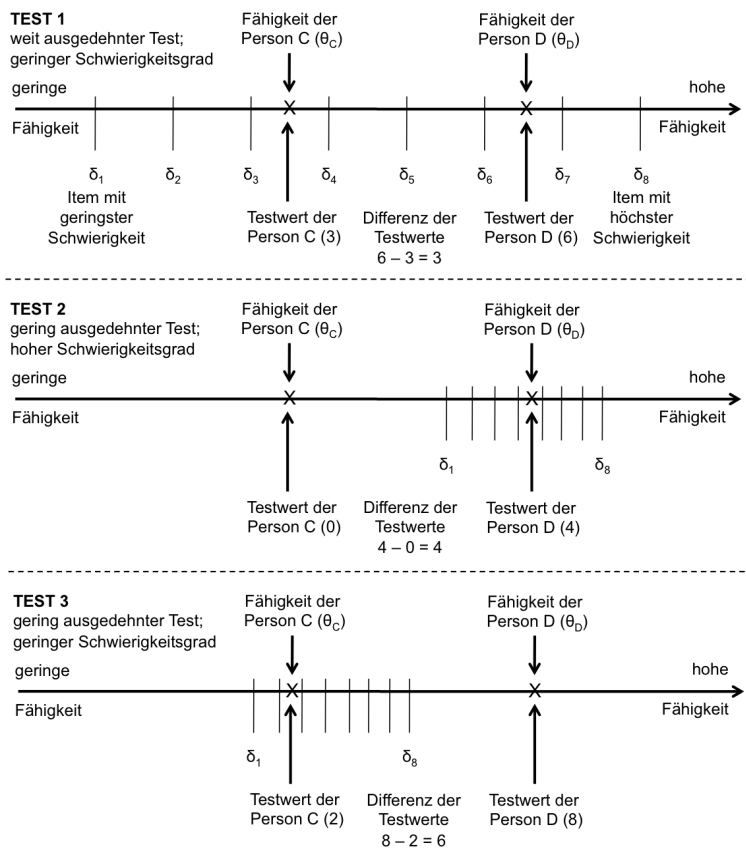
Die Gegenüberstellung der Antwortmuster der Personen A und B kann sehr anschaulich vermitteln, dass selbst dann, wenn Items konstruiert werden konnten, die das zu untersuchende Personenmerkmal valide und reliabel repräsentieren, die Antwortmuster der Probanden zu analysieren sind, bevor ihre Testwerte/Scores dazu genutzt werden können, die Ausprägung der Fähigkeit zu schätzen. Ein Test ist vor diesem Hintergrund nur dann aussagekräftig, wenn die Itemkalibrierung durch eine zuverlässige Anzahl an Personen validiert werden kann.

te, die vergleichbare Antwortmuster wie die Person A in der Abbildung 4.1 aufweisen.

Wie Schwierigkeitsgrad und Ausdehnung von Tests die Schätzung des Personenmerkmals beeinflussen

Ein weiteres Problem, das die Aussagkraft von Tests erheblich beeinflusst, ist die Abhängigkeit der Testwerte von dem Schwierigkeitsgrad der Testitems und der Ausdehnung, mit der ein Test das Kontinuum des Personenmerkmals repräsentiert.

Abbildung 4.2 Nichtlinearität der Testwerte (in Anlehnung an Wright & Stone, 1979, p. 14)



Dieser Zusammenhang ist bei Lernenden/Auszubildenden genau wie bei Lehrenden nur allzu gut bekannt und stellt eine alltägliche Herausforderung im Hin-

blick auf die Gestaltung von Leistungstests und deren Bewertung dar. Es ist direkt einsehbar, dass eine 80-Prozent-Lösungsquote bei einem sehr leichten Test nicht oder nur schwer vergleichbar ist mit einer 80-Prozent-Lösungsquote, die bei einem sehr schwierigen Test erreicht wurde. Zur Veranschaulichung dieses Problems wird in Abbildung 4.2 die Fähigkeit einer Person C als bekannt vorausgesetzt. Hierfür wird die Fähigkeit θ_C auf dem Kontinuum des Personenmerkmals fixiert und die relative Position der Fähigkeit θ_C zu den Items verschiedener Tests beschrieben. Die Tests unterscheiden sich hinsichtlich des Schwierigkeitsgrades der Items (geringer und hoher Schwierigkeitsgrad) und im Bezug auf ihre Ausdehnung auf dem Kontinuum des Personenmerkmals (weite und geringe Ausdehnung).

Die Abbildung 4.2 verdeutlicht, dass bei einer konstanten Fähigkeit θ_C die Person C in den verschiedenen Tests unterschiedliche Testwerte/Scores erreicht. So beträgt der Testwert der Person C bei Tests mit geringem Schwierigkeitsgrad drei bzw. zwei Punkte in Abhängigkeit davon, ob der Test einen breiten Bereich auf dem Kontinuum des Personenmerkmals abdeckt oder dieser Bereich stärker reduziert ist. Der Test mit hohem Schwierigkeitsgrad wird hingegen mit null Punkten abgeschlossen. Dieser Vergleich zeigt, dass selbst wenn die Validität der Itemkalibrierung und die Validität des Antwortverhaltens gegeben sind, keine Sicherheit dahingehend besteht, dass über die Testwerte zuverlässig auf die Ausprägung des Personenmerkmals geschlossen werden kann. Insbesondere in den Fällen, in denen keine Items (Test 2 in der Abbildung 4.2) oder aber alle Items gelöst werden, existiert keine zufriedenstellende Möglichkeit, aus dem Testwert der Person auf die Ausprägung des zu untersuchenden Personenmerkmals zu schließen. In diesen Fällen besitzt der angewandte Test keine Aussagekraft und es ist nach einem Test zu suchen, der besser zu der Fähigkeit des Probanden passt.

Die Darstellung der Abbildung 4.2 verweist auf ein weiteres Problem. Es geht um die Frage, wie kognitive Entwicklungen auf Basis von Testwerten/Scores zu beschreiben und zu vergleichen sind. Hierfür sei angenommen, dass die Tests 2 und 3 der Abbildung 4.2 zeitlich nach dem Test 1 eingesetzt wurden, um aufzuzeigen, wie sich die Fähigkeit der Person C im Zeitverlauf oder aber als Ergebnis einer Intervention verändert hat. Es wird deutlich, dass hier ein Vergleich unter Rückgriff auf die Testwerte nicht zielführend sein kann, da diese eine Veränderung hinsichtlich des Messwerts des zu untersuchenden Personenmerkmals implizieren, die gar nicht vorhanden ist.

Zu einer vergleichbaren Schlussfolgerung gelangt man, wenn die Abbildung 4.2 dazu genutzt wird, um die Aussagekraft von Tests im Hinblick auf die Beschreibung von Unterschieden zwischen Personen zu analysieren. Hierfür ist zusätzlich zur Fähigkeit der Person C die Fähigkeit einer Person D (θ_D) auf dem Kontinuum des Personenmerkmals fixiert worden. Der Abstand zwischen den Personen C und D ausgedrückt über die Differenz ihrer Fähigkeiten ist jeweils gleich. Werden diesen Personen Tests unterschiedlicher Schwierigkeit und Ausdehnung vorgelegt, zeigt sich, dass die Differenzen der Testwerte erheblich vari-

ieren. So ergeben sich bei den drei verschiedenen Tests auch drei verschiedene Differenzwerte (drei, vier sowie sechs). Diese implizieren, wechselnde Unterschiede der Fähigkeiten der Personen C und D – dies ist jedoch im Beispiel der Abbildung 4.2 nicht der Fall.

Im Hinblick auf die Aussagekraft von Tests resultieren aus obigen Überlegungen folglich zwei Fragen: (1) Wie lassen sich Testwerte/Scores aus verschiedenen Tests dazu nutzen, um Entwicklungen von Fähigkeiten zu beschreiben, und (2) wie lassen sich Differenzen der Testwerte zwischen Personen dazu nutzen, Unterschiede in den Fähigkeiten der Personen zu untersuchen? Hinsichtlich dieser Fragen kann zusammenfassend geschlussfolgert werden, dass Testwerte/Scores stets testgebunden sind. Sollen also Entwicklungen beschrieben werden, müssen neben der Berücksichtigung von Floor- und Ceiling-Effekten insbesondere die Itemkalibrierungen und die Antwortmuster zwischen den Tests vergleichbar sein, um valide auf die Ausprägung des Personenmerkmals schließen zu können. Hierfür sind insbesondere theoretische Vorarbeiten notwendig, die begründet Kriterien der Itemschwierigkeiten und Verhaltensmerkmale bei der Bearbeitung von unterschiedlich schwierigen Items auf einer Skala verorten können (vgl. Konstruktrepräsentation, Kapitel 3.3). Diese theoretischen Konstruktionsarbeiten ersetzen allerdings die empirische Analyse nicht. Es ist ferner festzuhalten, dass über Testwerte/Scores, die in der Form, wie in Abbildung 4.2 dargestellt, vorliegen, die Darstellung von Entwicklungen oder von Vergleichen zwischen Personen zu keinem aussagekräftigen Ergebnis führen kann. Die so ermittelten Testwerte verhalten sich nicht linear zu dem zu untersuchenden Personenmerkmal. D. h., um Testwerte als Basis der Messung zu nutzen, müssen Möglichkeiten gefunden werden, Testwerte in annähernd lineare Messwerte des zu untersuchenden Personenmerkmals zu transformieren.

Für aussagefähige Tests werden folglich (1) Messwerte benötigt, die testunabhängig (test-free) sind, um die Fähigkeit einer Person generell zu lokalisieren, und (2) Messwerte, die sich linear verhalten, um Entwicklungen und Unterschiede in der Leistungsfähigkeit zwischen Personen zu analysieren. Darüber hinaus wäre es (3) hilfreich, Parameter für die Itemschwierigkeiten zu schätzen, die möglichst unabhängig von der Stichprobe (sample-free) sind (Wright & Stone, 1979).

4.1.2 *Kurve des Testverhaltens*

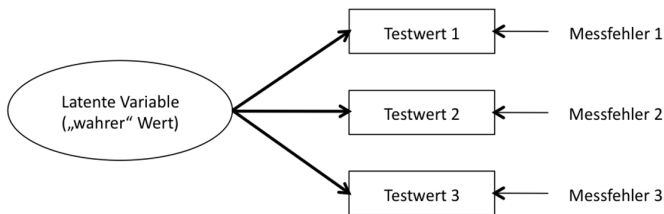
Die Forderung nach testunabhängigen Schätzern für die Fähigkeiten von Personen und nach stichprobenunabhängigen Schätzern für die Itemschwierigkeiten eines Tests verlangt nach einer Methode, in der das Testverhalten einer Person bei der Bearbeitung eines beliebigen Items exakt auf die Schwierigkeit dieses Items bezogen werden kann. Dies impliziert die Annahme, dass ausschließlich die Itemschwierigkeit δ_i das Antwortverhalten der Person beeinflusst. Umgekehrt gilt, dass ausschließlich die Fähigkeit einer Person θ_v vorgibt, wie unterschiedlich schwierige Items bearbeitet werden. Diese Annahme kann durchaus kontrovers diskutiert werden, da das Antwortverhalten von verschiedenen Faktoren abhän-

gig sein kann: Motivations- und Volitionseffekte, soziale Interaktion oder auch einfach nur die Möglichkeit, Antworten zu erraten. Hinter dieser Annahme verbirgt sich jedoch auch eine große Chance hinsichtlich der Gestaltung von Leistungsmessungen in Lehr-Lern-Kontexten. Letztlich sollen mit kognitiven Tests tatsächlich bestehende Personenfähigkeiten gemessen werden. Dies bedeutet für den Testentwickler, dass die interferierenden Einflussfaktoren zu kontrollieren und zu minimieren sind, um die Aussagefähigkeit von Tests zu erhöhen. Hinsichtlich der Konstruktion von Testitems besteht mit dieser Annahme die Verpflichtung, Schwierigkeitsparameter kognitiv begründet zu identifizieren und zum Ausgangspunkt der Entwicklungsprozesse zu machen.

Abbildung 4.3 Unterschied zwischen den Testtheorien
(in Anlehnung an Embretson, 1999, p. 6)

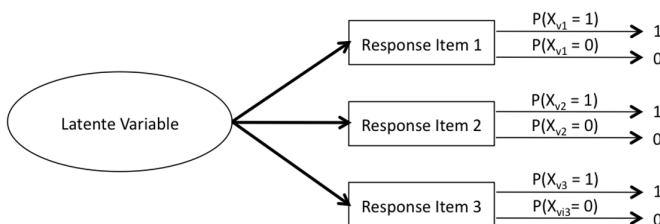
Klassische Testtheorie:

Beobachteter Wert = „wahrer“ Wert + Messfehler



Probabilistische Testtheorie (IRT):

Wahrscheinlichkeit $P(X_{vi} = 1) = \text{Funktion}[(\text{Personenfähigkeit}) - (\text{Itemschwierigkeit})]$



Mit der Klassischen Testtheorie und der Item Response-Theorie liegen zwei Möglichkeiten vor, wie aus dem Testverhalten auf das zu untersuchende Personenmerkmal (latente Variable) geschlossen werden kann (vgl. Abbildung 4.3): In der Klassischen Testtheorie ergibt sich der beobachtbare Testwert als Summe des „wahren“ Testwerts und eines Messfehlers. Dabei gilt, dass Messfehler eine uniforme Normalverteilung aufweisen, einen Erwartungswert von Null haben und unkorreliert mit allen anderen Variablen sind. Nach Embretson (1999, pp.

5/6) weist die Messfehlertheorie der Klassischen Testtheorie trotz der Messfehlerannahmen einige Limitationen auf, die sich insbesondere auf die Forderung nach test- und stichprobenunabhängigen Schätzern beziehen lassen:

- Der Messfehler lässt sich nur auf Items eines spezifischen Tests beziehen, da keine Annahmen über mögliche Variationen der Itemparameter im Rahmen der Klassischen Testtheorie getroffen werden können.
- Die Messfehlertheorie basiert auf zwei unabhängigen Variablen (dem „wahren“ Wert und dem Messfehler). Diese lassen sich jedoch nicht wirklich trennscharf separieren.

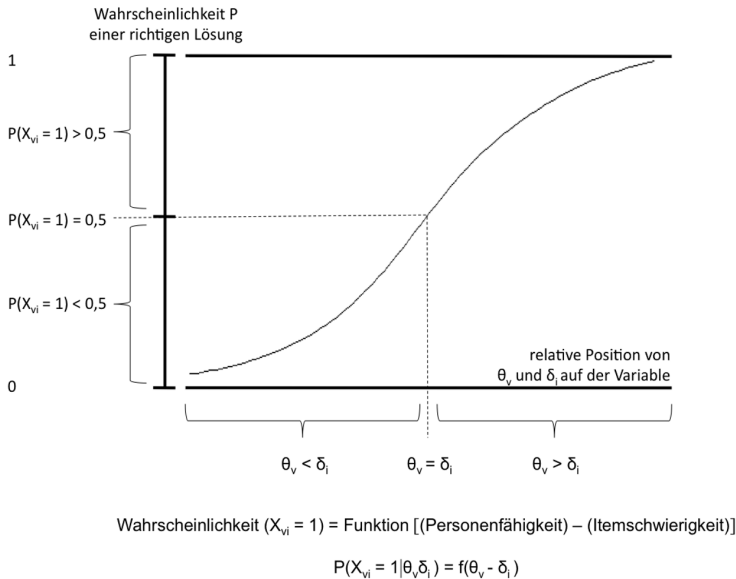
Die Itemeigenschaften sind nicht an das Testverhalten gebunden und müssen daher über Zusatzstatistiken abgesichert werden (Varianzen, Reliabilitäten).

Im Rahmen der Item Response-Theorie wird nach einer mathematischen Relation gesucht, die den Zusammenhang zwischen Personenfähigkeit θ_v und Itemschwierigkeit δ_i im Hinblick auf das Antwortverhalten X_{vi} beschreibt (vgl. u. a. Wright & Stone, 1979; Wilson, 2004c). Dabei wird davon ausgegangen, dass θ_v und δ_i Orte auf dem Kontinuum des Personenmerkmals repräsentieren und folglich auf eine gemeinsame Skala transformiert sind. Ihre natürliche Beziehung lässt sich damit als Differenz ($\theta_v - \delta_i$) ausdrücken. Diese Differenz von Personenfähigkeit θ_v und Itemschwierigkeit δ_i lässt sich im Folgenden logisch auf das Antwortverhalten X_{vi} beziehen. Ist die Personenfähigkeit größer als die Itemschwierigkeit, wird angenommen, dass das Item gelöst werden kann; ein negativer Differenzwert deutet umgekehrt darauf hin, dass das Item nicht gelöst wird. Da es in der Bearbeitung von Tests dennoch dazu kommen kann, dass Personen mit einer geringen Ausprägung der Personenfähigkeit ein in Relation dazu schwieriges Item lösen können, greift die Item Response-Theorie nicht auf deterministische Relationen, sondern auf die Annahme von Wahrscheinlichkeiten zurück (probabilistische Testtheorie). Die Wahrscheinlichkeit, dass eine Person ein Item bewältigen kann oder aber bei der Bearbeitung scheitert, ist damit ausschließlich abhängig von der Fähigkeit der Person und der Itemschwierigkeit. Bei einem positiven Differenzwert besteht demnach eine mehr als 50-prozentige Wahrscheinlichkeit, das Item lösen zu können; ein negativer Differenzwert steht für eine mehr als 50-prozentige Wahrscheinlichkeit, das Item nicht zu lösen. Entspricht die Personenfähigkeit exakt der Itemschwierigkeit, ist die Wahrscheinlichkeit hingegen genau $P = 0,5$.

In Abbildung 4.4 ist die Kurve des Testverhaltens in Abhängigkeit der Personenfähigkeit und der Itemschwierigkeit als itemcharakteristische Funktion für jede mögliche Kombination dargestellt. Die Beziehung zwischen Personenfähigkeit und Itemschwierigkeit lässt sich jedoch auf zwei verschiedene Arten beschreiben: Einerseits kann das Antwortverhalten von Personen verschiedener Fähigkeitsausprägungen auf ein Item bezogen werden; andererseits lassen sich verschiedene Itemschwierigkeiten, mit denen die Fähigkeit einer Person gemessen wird, abbilden. Wenn die Kurve ein Item auf Basis des Antwortverhaltens von Personen mit unterschiedlichen Fähigkeitsausprägungen beschreibt, handelt es sich um eine Item Characteristic Curve (ICC; Abbildung 4.4); wird über verschie-

dene Itemschwierigkeiten die Fähigkeit einer Person dargestellt, wird von einer Person Characteristic Curve (PCC; Abbildung 4.5) gesprochen (vgl. Wright & Stone, 1979, p. 12).

Abbildung 4.4 Kurve des Testverhaltens (in Anlehnung an Wright & Stone, 1979, p. 14)



Die Kurve des Testverhaltens beschreibt den auf Wahrscheinlichkeiten basierenden mathematischen Zusammenhang zwischen Personenfähigkeit θ_v und Itemschwierigkeit δ_i bezogen auf das beobachtbare Antwortverhalten X_{vi} . Auf Basis dieser mathematischen Relation lässt sich der Messwert der Person durch Schätzung von θ_v und die Kalibrierung des Items durch Schätzung von δ_i bestimmen. Über die in Abbildung 4.4 dargestellte Beziehung können die Schätzung der Personenfähigkeiten testunabhängig und die Schätzung der Itemschwierigkeiten stichprobenunabhängig erfolgen (Embretson & Reise, 2000; van der Linden & Hambleton, 1997; vgl. Kapitel 4.3).

Der Kurvenverlauf der Abbildung 4.4 lässt sich über verschiedene mathematische Funktionen beschreiben. Für Modelle der Item Response-Theorie werden typischerweise logistische Funktionen gewählt. Alternativ existieren Modelle, die sich auf die Normalverteilung beziehen lassen – deren Funktionen sind jedoch ungleich komplexer, so dass in der praktischen Anwendung logistische IRT-Modelle überwiegen. Eine möglichst einfache Form, den Kurvenverlauf der Ab-

bildung 4.4 in einer mathematischen Relation auszudrücken, ist mit der nachfolgenden Exponentialfunktion gegeben:

$$P(X_{vi} = 1 | \theta_v, \delta_i) = \frac{\exp(\theta_v - \delta_i)}{1 + \exp(\theta_v - \delta_i)} = \frac{e^{(\theta_v - \delta_i)}}{1 + e^{(\theta_v - \delta_i)}}$$

mit $(\theta_v - \delta_i) = 0$ folgt:

$$P(X_{vi} = 1 | \theta_v, \delta_i) = \frac{e^0}{1 + e^0} = \frac{1}{1 + 1} = \frac{1}{2}$$

mit $(\theta_v - \delta_i) = -x$; $x \neq 0$ folgt:

$$P(X_{vi} = 1 | \theta_v, \delta_i) = \frac{e^{-x}}{1 + e^{-x}} = \frac{\frac{1}{e^x}}{1 + \frac{1}{e^x}} = \frac{1}{1 + e^x} < \frac{1}{2}$$

mit $(\theta_v - \delta_i) = x$; $x \neq 0$ folgt:

$$P(X_{vi} = 1 | \theta_v, \delta_i) = \frac{e^x}{1 + e^x} > \frac{1}{2}$$

Diese Exponentialfunktion wurde vor dem Hintergrund folgender Bedingungen gewählt: Die Differenz $(\theta_v - \delta_i)$ kann beliebige Werte zwischen $-\infty$ und $+\infty$ annehmen. Die Wahrscheinlichkeit eines erfolgreichen Antwortverhaltens ($X_{vi} = 1$) soll hingegen zwischen 0 und 1 variieren. Zum Erreichen dieser Bedingungen wird (1) die Differenz $(\theta_v - \delta_i)$ als Exponent der Eulerschen Zahl e geschrieben. Damit wird erreicht, dass der Exponentialausdruck zwischen 0 und $+\infty$ variiert. Um ihn im Intervall zwischen 0 und 1 zu halten, wird (2) der entsprechende Quotient gebildet. Mit Hilfe dieser mathematischen Funktion gelingt es, die Personenfähigkeit θ_v und die Itemschwierigkeit δ_i unabhängig voneinander zu schätzen. Damit liegt ein logistisches Modell vor, das sowohl die Linearität der Skala als auch die Generalität der Messwerte möglich macht. Der dänische Mathematiker Georg Rasch (1960) hat dieses Modell erstmals für Fragen der psychometrischen Modellierungen genutzt – das Rasch-Modell ist das einfachste Modell der IRT-Familie und ist in der dargestellten Form ausschließlich für dichotome Antwortmuster anwendbar.

4.2 Modelle der Item Response-Theorie

Nachfolgend werden ausgewählte IRT-Modelle vorgestellt. Aus der Breite der verfügbaren Modelle werden jedoch nur diejenigen genauer betrachtet, die in den Studien zur Messung kaufmännischer Kompetenzen zur Anwendung gekommen sind und deren Einsatz im Hinblick auf ein international vergleichendes Large-Scale Assessment in der beruflichen Bildung als sinnvoll erachtet werden kann. Die Modelle lassen sich danach sortieren, ob mit ihnen nominal- oder ordinalskalierte Daten berechnet werden sollen und ob eine eindimensionale oder mehrdimensionale Struktur der Daten vorliegt.

Nominal- und ordinalskalierte Daten sind typisch für kognitive Erhebungen – sowohl im Wissens- als auch im Handlungsbereich. Wird beispielsweise untersucht, inwieweit Lernende/Auszubildende in der Lage sind, eine Preiskalkulation durchzuführen, ist es aus betrieblicher Perspektive einerseits legitim, direkt nach dem korrekten Ergebnis zu fragen, da nur dieses für die weiteren Geschäftsprozesse von Belang ist bzw. ein falsches Ergebnis zu fehlerbesetzten betrieblichen Entscheidungen führen kann. In diesem Fall geben dichotome Daten eine hinreichend genaue Auskunft. Andererseits kann es aus Perspektive individueller Lern- und Entwicklungsprozesse hilfreich sein, detaillierter nach Verstehensstufen im Rahmen der Preiskalkulation zu suchen und danach zu fragen, wie das ermittelte Ergebnis zustande gekommen ist. Wie wurden die Kostenpreise ermittelt? Konnte eine ökonomisch sinnvolle Beziehung zu den Marktpreisen hergestellt werden? In diesem Fall ergeben sich ordinale Antwortkategorien. Die Modelle der IRT verarbeiten die unterschiedliche Datenqualität durch verschiedene mathematische Relationen – dies bedeutet, dass die Wahl der Skala und die angenommenen Beziehungen, die zwischen den Antwortkategorien bestehen, die Auswahl der IRT-Modelle beeinflussen.

Die verschiedenen Modelle für nominal- und ordinalskalierte Daten können im Hinblick auf die Fragen der Dimensionalität erweitert werden. Das im Kapitel 2.4.2 vorgestellte Kompetenzmodell für die kaufmännische Bildung impliziert eine mehrdimensionale Struktur. Es wird beispielsweise generell erwartet, dass sich domänenspezifische und domänenverbundene Kompetenzbereiche empirisch separieren lassen, wobei diese Bereiche wiederum verschiedene Dimensionen aufweisen können. So kann im domänenverbundenen Bereich zwischen „economic literacy“ und „economic numeracy“ differenziert werden; im domänenspezifischen Bereich sind Bereichsdefinitionen mit Bezug auf verschiedene Geschäftsprozesse (Wertschöpfungs-, Steuerungs- und Unterstützungsprozesse; vgl. Kapitel 3.2.3) und/oder hinsichtlich des Abstraktionsgrades der Arbeitsprozesse vorstellbar. Der Abstraktionsgrad des Arbeitsprozesses bestimmt sich durch die Nähe/Distanz des in der Testsituation inszenierten Arbeitsprozesses zu realen Arbeitsprozessen. Werden also beispielsweise über betriebliche Simulationen domänenspezifische Handlungen und konkrete Arbeitsschritte erfasst, zeigt diese Testumgebung einen geringeren Abstraktionsgrad auf als eine Testumgebung, die den inhaltlich gleichen Arbeitsprozess darüber erfasst, dass Ler-

nende/Auszubildende die zur Bewältigung des Arbeitsprozesses notwendigen domänenspezifischen Handlungen und konkreten Arbeitsschritte beschreiben. Diese Beschreibungen können dabei in unterschiedlicher Form vorliegen: von Essays über Verstehensfragen bis hin zu Multiple-Choice-Tests. Im Falle der Studien zur Messung kaufmännischer Kompetenz beziehen sich die unterschiedlichen Dimensionen auf verschiedene kognitive Strukturen, die sich auf Basis unterschiedlicher Inhalts- und Zugriffsebenen der beruflichen Lern- und Anforderungssituationen ergeben und folglich mögliche Kompetenzstrukturen darstellen. Über mehrdimensionale Modelle der IRT lassen sich diese Kompetenzstrukturen empirisch validieren.

4.2.1 IRT-Modelle für nominalskalierte Daten

Im folgenden Kapitel werden das 1-parametrische logistische Modell (1PL) oder Rasch-Modell sowie das mehrdimensionale Rasch-Modell vorgestellt (für eine Beschreibung des 2PL- und 3PL-Modells sowie für einen Überblick weiterer Modelle vgl. van der Linden & Hambleton, 1996).

Das 1-parametrische logistische Modell

Das 1-parametrische logistische Modell oder Rasch-Modell beschreibt die Wahrscheinlichkeitsverteilung von Antworten auf dichotome Items in Abhängigkeit von zwei Variablen – dem Fähigkeitsparameter der Person θ_v und dem Schwierigkeitsparameter des Items δ_i . Da das Itemmerkmal nur über die Schwierigkeit in das Modell eingeht, wird das Rasch-Modell auch als 1-parametrisches logistisches Modell bezeichnet (Embretson & Reise, 2001, pp. 67/68).

Die abhängige Variable im Rasch-Modell ist folglich die dichotome Antwort einer spezifischen Person auf ein ausgewähltes Testitem; die unabhängigen Variablen sind die Fähigkeit der Person θ_v und die Itemschwierigkeit δ_i . Die Beziehung zwischen den unabhängigen Variablen wird als latente Additivität bezeichnet. Diese drückt aus, dass Person- und Itemparameter linearisiert sind und sich auf einer gemeinsamen Skala oder Dimension befinden. Damit ergibt sich $(\theta_v - \delta_i)$ als natürliche Relation der unabhängigen Variablen: Der Itemparameter wird vom Personenparameter subtrahiert und beschreibt damit die Itemschwierigkeit. Die Beziehung zwischen der Differenz $(\theta_v - \delta_i)$ und der Antwort auf das dichotome Item hängt davon ab, wie die abhängige Variable modelliert wird. Sie lässt sich als natürlicher Logarithmus des Wettquotienten (log odd) ausdrücken oder aber über Wahrscheinlichkeiten beschreiben.

Werden Wahrscheinlichkeiten genutzt, ergibt sich die in Kapitel 4.1.2 dargestellte Modellform: Die abhängige Variable beschreibt die Wahrscheinlichkeit, dass eine Person v das Item i erfolgreich lösen kann ($P(X_{vi} = 1|\theta_v, \delta_i)$). Allgemeiner lässt sich das 1-parametrische logistische Modell wie folgt formulieren: Beschrieben wird die Wahrscheinlichkeit, mit der eine Person v mit der Fähigkeit θ_v eine Antwort X_{vi} auf ein Item i mit der Schwierigkeit δ_i gibt. Die Antwort X_{vi} kann dabei nur zwei mögliche Werte annehmen: den Wert $X_{vi} = 1$ für eine erfolgreiche Bear-

beutung des Items und den Wert $X_{vi} = 0$, wenn das Item nicht gelöst werden kann (Wright & Stone, 1979, p. 17). Mathematisch lässt sich das Rasch-Modell ausgedrückt über Wahrscheinlichkeiten wie folgt darstellen:

$$P(X_{vi} | \theta_v, \delta_i) = \frac{\exp[X_{vi}(\theta_v - \delta_i)]}{1 + \exp[X_{vi}(\theta_v - \delta_i)]} = \frac{e^{X_{vi}(\theta_v - \delta_i)}}{1 + e^{(\theta_v - \delta_i)}}$$

mit $X_{vi} = 1$ folgt:

$$P(X_{vi} = 1 | \theta_v, \delta_i) = \frac{\exp(\theta_v - \delta_i)}{1 + \exp(\theta_v - \delta_i)} = \frac{e^{(\theta_v - \delta_i)}}{1 + e^{(\theta_v - \delta_i)}}$$

mit $X_{vi} = 0$ folgt:

$$P(X_{vi} = 0 | \theta_v, \delta_i) = \frac{1}{1 + \exp(\theta_v - \delta_i)} = \frac{1}{1 + e^{(\theta_v - \delta_i)}}$$

Die zweite Variante, das Rasch-Modell zu beschreiben, bezieht sich auf die Darstellung des logarithmierten Wettquotienten (log odd). Der Wettquotient ist das Verhältnis der Anzahl der Erfolge und der Anzahl der Misserfolge. Ist der Wettquotient beispielsweise 4/1 bedeutet dies, dass eine Person bei fünf möglichen Versuchen viermal erfolgreich sein wird und einmal scheitert.

Wettquotient:

$$\frac{P(X_{vi} = 1 | \theta_v, \delta_i)}{P(X_{vi} = 0 | \theta_v, \delta_i)} = \frac{\frac{\exp(\theta_v - \delta_i)}{1 + \exp(\theta_v - \delta_i)}}{\frac{1}{1 + \exp(\theta_v - \delta_i)}} = \frac{\frac{e^{(\theta_v - \delta_i)}}{1 + e^{(\theta_v - \delta_i)}}}{\frac{1}{1 + e^{(\theta_v - \delta_i)}}} = e^{(\theta_v - \delta_i)}$$

Logarithmierter Wettquotient:

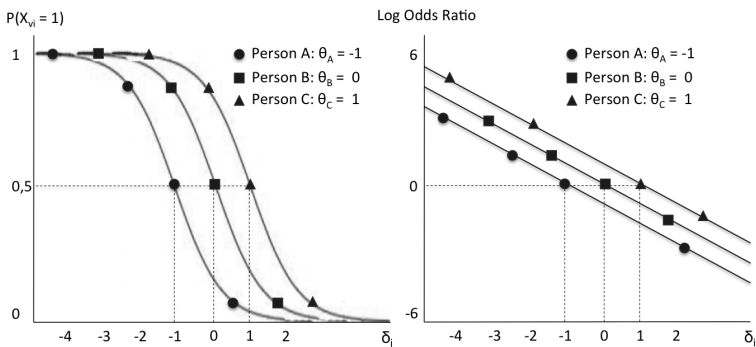
$$\ln \frac{P(X_{vi} = 1 | \theta_v, \delta_i)}{P(X_{vi} = 0 | \theta_v, \delta_i)} = \theta_v - \delta_i$$

Der Wettquotient lässt sich natürlich auch als das Verhältnis von Wahrscheinlichkeiten ausdrücken. Die Wahrscheinlichkeit des Erfolgs dividiert durch die Wahrscheinlichkeit des Misserfolgs beträgt im gewählten Beispiel 0,80/0,20. Der Wettquotient beschreibt damit nichts anderes als das Verhältnis der Lösungswahrscheinlichkeit zu ihrer Gegenwahrscheinlichkeit; der logarithmierte Wettquotient ist eine lineare Funktion der Personenfähigkeit und der Itemschwierig-

keit. Sind Personenfähigkeit und Itemschwierigkeit gleich, ergibt sich ein log odd von 0. Dies bedeutet, dass die Chancen der Person v mit der Fähigkeit θ_v , das Item i der Schwierigkeit δ_i zu lösen oder bei der Lösung zu scheitern, gleich verteilt sind (0,50/0,50).

In Abbildung 4.5 sind für einen Vergleich die log odds sowie die Wahrscheinlichkeiten von drei Personen (als Person Characteristic Curves) mit unterschiedlichen Fähigkeitsausprägungen im Hinblick auf Itemschwierigkeiten dargestellt.

Abbildung 4.5 Wahrscheinlichkeiten und Log Odds (vgl. Embretson & Reise, 2001, p. 50)



Der Zusammenhang zwischen Person- und Itemparameter lässt sich als lineare Funktion zur Vorhersage der log odds und als nichtlineare Funktion (S-Shape) zur Vorhersage von Wahrscheinlichkeiten beschreiben. Die Abbildung 4.5 zeigt für die nichtlineare Funktion, dass Änderungen der Itemschwierigkeiten dann die größten Effekte auf die Erfolgswahrscheinlichkeiten haben, wenn die Itemschwierigkeiten um den Personenparameter variieren (hier ist der Anstieg der nichtlinearen Funktion am größten). Die Person A hat eine Personenfähigkeit von -1. Bei einer Itemschwierigkeit von ebenfalls -1 besteht eine 50-prozentige Chance, das Item richtig zu lösen. Diese Lösungswahrscheinlichkeit sinkt erheblich, wenn die Itemschwierigkeit von -1 auf 0 erhöht wird; eine weitere Erhöhung der Itemschwierigkeit von 0 auf 1 zeigt hingegen einen deutlich geringeren Effekt. Für die lineare Funktion gilt, dass je höher die Itemschwierigkeit ist, desto geringer die log odds sind. Entspricht die Itemschwierigkeit der Personenfähigkeit sind die log odds = 0. Für den Fall, dass die Personenfähigkeit größer als die Itemschwierigkeit, ist ein Erfolg bei der Bearbeitung des Items wahrscheinlicher als ein Misserfolg und vice versa.

Insbesondere die mathematische Relation der log odds und die folgende lineare Beziehung zwischen Personen- und Itemparameter verdeutlichen die beson-

deren Eigenschaften des Rasch-Modells (vgl. Embretson & Reise, 2001, pp. 49/50):

- Bei einer bekannten Fähigkeitsausprägung können die log odds des Erfolgs für jedes kalibrierte Item ermittelt werden. Umgekehrt gilt, dass ausgehend von jedem Item, dessen Schwierigkeit bekannt ist, ein Schätzer der Personenfähigkeit gefunden werden kann. Im 1-parametrischen Modell wird folglich die Itemschwierigkeit auf die Leistungsfähigkeit bezogen. In mehr komplexen Modellen ist es möglich, die Itemmerkmale auszuweiten. So können Parameter der Itemdiskrimination (2PL Modell), Guessing-Effekte (3PL Modell) oder selbst die Beschaffenheit von Alternativen berücksichtigt werden (vgl. hierzu die Beiträge der Originalautoren in van der Linden & Hambleton, 1997).⁷
- Personen- und Itemparameter werden parallel skaliert. Wird folglich die Ausprägung der Personenfähigkeit um einen fiktiven festen Wert erhöht, steigt damit die Antwortwahrscheinlichkeit im gleichen Ausmaß, als würde die Itemschwierigkeit um diesen fiktiven Wert reduziert werden.
- Das Antwortverhalten wird ausschließlich durch die Itemschwierigkeit und die Personenfähigkeit bestimmt. Damit liegt mit dem Rasch-Modell ein komplettes Verhaltensmodell vor, das für die Person und für das Testitem durch differenzierte Parameter zu beschreiben ist. Diese *Separabilität* von Personen- und Itemmerkmalen hat zur Folge, dass die Aussagen über die Personen, die einen Test bearbeitet haben, und über die Items dieses Tests nur von den Merkmalen der Personen und Items abhängen, nicht jedoch von den Methoden, mit denen diese Merkmale gemessen worden sind.
- Personen- und Itemparameter sind unabhängige Variablen, die separat geschätzt werden können. Rasch beschreibt diese Eigenschaft als *spezifische Objektivität* und misst ihr eine essentielle messtheoretische Bedeutung bei (Rasch, 1960, pp. 104/105). Spezifische Objektivität im Rasch-Modell bedeutet, dass die Personenparameter von der Stichprobe der Items und die Itemparameter von der Stichprobe der Personen unabhängig sind. Nur so sind Vergleiche zwischen Personen sowie Vergleiche zwischen Items zulässig (Stichprobenunabhängigkeit; vgl. Kapitel 4.1.1). Die Personenparameter sind demnach unabhängig (1) von den Items, mit denen sie gemessen worden sind, und (2) von den Merkmalen anderer Personen; die Itemparameter sind hingegen unabhängig (1) von dem Antwortverhalten der Personen und (2) von den Merkmalen anderer Items des Tests.

Wie die Darstellungen im Kapitel 4.1.1 bereits gezeigt haben, sind die Eigenschaften der Separabilität und der spezifischen Objektivität sowie das Vorhandensein von Messwerten, die sich linear verhalten, unabdingbar, wenn Unterschiede zwischen Personen analysiert werden sollen. Die Möglichkeit, Messwerte für Personen zu ermitteln, die unabhängig sind von den Items, die die Personen

⁷ Die komplexeren IRT-Modelle für dichotome Daten gehen in das Rasch-Modell über, wenn die zusätzlich berücksichtigten Parameter beispielsweise der Trennschärfe, des Schätzverhaltens etc. auf 1 fixiert werden.

bearbeiten, eröffnet insbesondere mit Blick auf ein Large Scale-Assessment in der beruflichen Bildung die Möglichkeit vor dem Hintergrund begrenzter Testzeit, nicht allen Personen alle Items vorzulegen – die Eigenschaft der spezifischen Objektivität würde in diesem Fall die Vergleichbarkeit der Messwerte gewährleisten. Insbesondere vor diesem Hintergrund ist auch die Expertise von Rost im Rahmen der Machbarkeitsstudie „Berufsbildungs-PISA“ zu lesen (Rost, 2006). In dieser Expertise zum Einsatz der Item Response-Theorie für die Messung berufsbezogener Kompetenzen kommt Rost aufgrund der zu erwartenden Testparameter – Längsschnittdesign, Verlinkung der Teilstichproben sowie Effekte auf die Itemschwierigkeiten – zu dem Ergebnis, den Einsatz von IRT-Modellen zu präferieren.

Das explorative mehrdimensionale Rasch-Modell

Bei der Wahl mehrdimensionaler IRT-Modelle (MIRT) ist, vergleichbar mit der Faktoranalyse der Klassischen Testtheorie, im Vorfeld darüber zu urteilen, ob explorative oder konfirmatorische Modelle zu wählen sind. Für diese Entscheidung gelten zwischen Klassischer Testtheorie und der IRT vergleichbare Kriterien: Ein exploratives Modell wird gewählt, wenn in einem Datensatz nach Strukturen zwischen den Daten gesucht wird und im Vorfeld der Analyse der Zusammenhang zwischen den Variablen nicht bekannt ist. Ein exploratives Modell dient demnach dem Ziel der Hypothesengenerierung. Ein konfirmatorisches Modell wird im Gegensatz dazu für die Hypothesenüberprüfung eingesetzt – also für Daten, in denen bereits unterschiedliche Dimensionen vermutet werden, die sich theoriebezogen begründen lassen.

Die Beschreibung des explorativen Rasch-Modells für dichotome Daten geht auf McKinley und Reckase (1982; zit. n. Embretson & Reise, 2001, p. 83) zurück:

$$P(X_{vi} = 1 | \theta_{vm}, \delta_i) = \frac{\exp(\sum_m \theta_{vm} + \delta_i)}{1 + \exp(\sum_m \theta_{vm} + \delta_i)}$$

Die Gleichung weist im Vergleich zum eindimensionalen Rasch-Modell folgende Veränderungen auf: (1) Im Hinblick auf den Itemparameter wird im mehrdimensionalen Raschmodell nicht die Schwierigkeit, sondern die Leichtigkeit des Items abgebildet. Dies drückt sich dadurch aus, dass die Relation zwischen Personen- und Itemparameter nicht durch eine Differenz, sondern durch eine Summe beschrieben wird. (2) Anstatt des Personenparameters θ_v werden im mehrdimensionalen Modell mehrere gleich gewichtete Fähigkeitsparameter für m verschiedene Dimensionen zusammengestellt. θ_{vm} beschreibt dann folglich den Fähigkeitsparameter der Person v in der Dimension m . Diese Aufgliederung stellt sicher, dass nur dann, wenn die multiplen Dimensionen nicht gleich in jedem Item gewichtet sind, eine mehrdimensionale Struktur aufgezeigt werden kann.

Bei einem gleichen Gewicht ließen sich für die verschiedenen Dimensionen keine Unterschiede im Antwortverhalten der Probanden abbilden. Dies führt dazu, dass die verschiedenen Fähigkeitsparameter der m Dimensionen nicht separat geschätzt werden können und damit das Modell nicht passt. In diesem Fall ist zu prüfen, ob das eindimensionale Rasch-Modell die Daten fittet (vgl. Embretson & Reise, 2001, p. 83).

Insgesamt lässt sich urteilen, dass explorative Modelle für kognitive Tests im Rahmen der IRT kaum zum Einsatz kommen. Dies liegt insbesondere daran, dass im Vorfeld von Assessments detaillierte, kognitiv determinierte Überlegungen darüber angestellt werden, wie Items zu konstruieren sind und welche kognitiven Fähigkeiten diese Items bei den Probanden ansprechen können und sollen. Diese Überlegungen umfassen folglich auch eine Abbildung möglicher Antwortmuster und die Fragen, wie Probanden auf die Items reagieren werden und wie sich diese Reaktionen in Messwerte der Personenfähigkeit transformieren lassen (vgl. Assessmentmodelle Kapitel 3.1).

Im Rahmen mehrdimensionaler IRT-Modelle ist die Anzahl der Modelle besonders groß – hier spielt die Anwendung der Modelle eine Rolle und die Frage, wie sich Dimensionen unterscheiden lassen und welche Parameter diese Unterschiede mathematisch differenziert durch die Kombination mit einem eindimensionalen IRT-Modell am besten hervorbringen. So gibt es beispielsweise Modelle, die sich mit der Erfassung von Lernleistungen im Zeitverlauf in Folge von instruktionalen Implikationen befassen (MRMLC: Multidimensional Rasch Model for Learning and Change, Embretson, 1991), oder Modelle, in denen Itemparameter auf Basis unterschiedlicher Item Stimuli und deren Effekte auf das Antwortverhalten geschätzt werden (MLTM: Multicomponent Latent Trait Model, Whitely, 1980). Moderne konfirmatorische mehrdimensionale IRT-Modelle zeichnen sich dadurch aus, dass sie neben dichotomen auch ordinale Daten verarbeiten können. Beispiele dafür werden im folgenden Kapitel vorgestellt.

4.2.2 IRT-Modelle für ordinalskalierte Daten

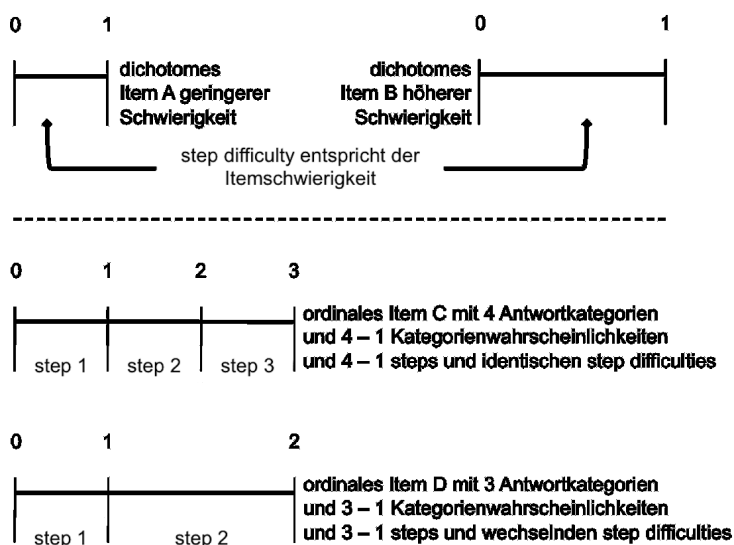
Für kognitive Tests ist es typisch, dass Items mehr als zwei Antwortkategorien aufweisen. Eine Preiskalkulation kann „vollständig richtig“, „teilweise richtig“ oder „falsch“ sein. Im Rahmen der IRT existieren verschiedene Modelle, in denen das Rasch-Modell auf ordinale Daten erweitert wurde. Im Folgenden wird das Partial Credit Model (PCM; Masters, 1982) als Beispiel für ein eindimensionales IRT-Modell für ordinale Daten und das Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM; Adams, Wilson & Wang, 1997) zur Abbildung mehrerer Fähigkeitsdimensionen vorgestellt.

Das Partial Credit Model

Das für Kompetenzmessungen intuitiv am besten anzuwendende Modell ist das Partial Credit Model (PCM; Masters, 1982). Die Bezeichnung Partial Credit bezieht sich darauf, dass es möglich ist, bei einem Item nicht die volle Punktzahl

(full credit) zu erreichen, jedoch aber eine Teilpunktzahl (partial credit). Das PCM hat gegenüber anderen IRT-Modellen den Vorteil, dass der Abstand zwischen den vergebenen Punkten variieren kann. Abbildung 4.6 stellt diese Grundidee dar: Bei dichotomen Antwortkategorien beschreibt der Abstand zwischen 0 und 1 die Schwierigkeit des Items. Je größer der Abstand, umso schwieriger ist das Item (vgl. dichotome Items A und B).

Abbildung 4.6 Graphische Darstellung von Itemschwierigkeit und step difficulties



Das PCM zerlegt ein ordinales Item mit verschiedenen Antwortkategorien in unterschiedliche Schwellenschwierigkeiten, wobei die Zwischenräume zwischen den Antwortkategorien die Schwierigkeit, von einer Antwortkategorie zur nächsten zu gelangen, beschreiben (step difficulties). Es wird folglich davon ausgegangen, dass zum Erreichen der nächst höheren Antwortkategorie die davorliegende Kategorie gelöst sein muss. Die Besonderheit des PCM ist, dass im Hinblick auf ein Item die Zwischenräume der Antwortkategorien variieren können (vgl. ordinales Item D in Abbildung 4.6) und dass die Abstände zwischen den Antwortkategorien beliebig für verschiedene Items auslegbar sind. Das PCM zerlegt folglich ein ordinales Item mit x Antwortkategorien in $(x-1)$ Kategorienwahrscheinlichkeiten für $(x-1)$ dichotome Items, die sich über $(x-1)$ Schwellenschwierigkeiten (step difficulties) beschreiben lassen (vgl. Wright & Masters, 1992, pp. 43/44).

„The logic underlying the PC model is analogous to modeling a polytomous item as if it were a series of dichotomous items, one for every step

in the polytomous item, each successive one of which being conditional on having answered correctly the one for the preceding step" (Moore, 1996, p. 221).

Dieses Konzept ist für die Abbildung von Fähigkeitsstrukturen unumgänglich, da es berücksichtigt, dass die Kategorien eines Items unterschiedliche Schwierigkeits"hindernisse" aufweisen können. Zur Veranschaulichung sei folgendes Item gegeben:

In Ihrem Ausbildungsbetrieb wurden bei Aufnahme der Geschäftstätigkeit die Herstellkosten ohne Berücksichtigung der Maschinenstundensätze berechnet, so dass sich ein Fertigungsgemeinkostenzuschlag von über 800 Prozent ergab. Bewerten Sie die Höhe des Zuschlagssatzes ökonomisch!

Die Antwortkategorien sind wie folgt gewählt: 2 = ökonomisch sinnvolle Interpretation mit Bezug auf das betriebswirtschaftliche Konzept; 1 = ökonomisch sinnvolle Interpretation mit Bezug auf die betriebliche Praxis; 0 = ökonomisch nicht sinnvoll. Im Hinblick auf das Antwortverhalten sind jetzt folgende Variationen möglich: Das Erreichen der Antwortkategorie 1 setzt einen hohen Fähigkeitsparameter voraus. Ist die Antwortkategorie jedoch erreicht, fällt es den Probanden relativ leicht, auch auf ein betriebswirtschaftliches Konzept im Rahmen der Interpretation zu verweisen. Eine andere Möglichkeit ist, dass unter Bezugnahme auf die betriebliche Praxis dieses Item einen leichten Schwierigkeitsparameter aufweist, also die Antwortkategorie 1 mit einem geringen Fähigkeitsparameter erreicht werden kann. Die Antwortkategorie 2 und damit das Herstellen einer direkten betriebswirtschaftstheoretischen Beziehung stellt jedoch eine große Schwierigkeits"barriere" dar und setzt damit einen hohen Fähigkeitsparameter voraus.

Letztlich ist empirisch zu ermitteln, welche Schwierigkeitsparameter die einzelnen Steps/Schwellen aufweisen – dadurch, dass aber ein Item verschiedene Schwierigkeitsparameter im PCM aufweisen kann, ist es möglich, die kognitiven Strukturen der Lernenden/Auszubildenden gezielter zu untersuchen und entsprechende Hinweise für die Gestaltung von Lehr-Lernprozessen zu geben.

Mathematisch wird diese Eigenschaft als *divide-by-total* beschrieben. Das PCM ist damit anders als andere IRT-Modelle für ordinale Daten ein *direktes* Modell, d. h. die Antwortwahrscheinlichkeit einer Kategorie (Kategorienwahrscheinlichkeit) wird als Exponential direkt ins Verhältnis zur Summe aller Exponentiale der Kategorienwahrscheinlichkeiten gesetzt. Damit bleiben die besonderen Eigenschaften des Rasch-Modells (Separabilität, spezifische Objektivität) erhalten (vgl. Embretson & Reise, 2001, p. 106).

Der Term δ_{ij} ($j = 1, \dots, m_i$) beschreibt die step difficulties bzw. Schwellenschwierigkeiten. Je größer δ_{ij} ausfällt, desto schwieriger ist ein Step von einer

Antwortkategorie zur nächsten relativ zu den anderen Steps innerhalb des Items. Für das ordinale Item D mit drei Antwortkategorien der Abbildung 4.6 gilt, dass zwei step difficulties ermittelt werden und dass Step 2 relativ zu Step 1 eine höhere Schwierigkeit aufweist.

$$P(X_{vi} | \theta_v, \delta_{ij}) = \frac{\exp \sum_{j=0}^s (\theta_v - \delta_{ij})}{\Psi}$$

für drei Antwortkategorien gilt:

$$P(X_{vi} = 0 | \theta_v, \delta_{i0}) = \frac{1}{1 + \exp(\theta_v - \delta_{i1}) + \exp(2\theta_v - \delta_{i1} - \delta_{i2})}$$

$$P(X_{vi} = 1 | \theta_v, \delta_{i1}) = \frac{\exp(\theta_v - \delta_{i1})}{1 + \exp(\theta_v - \delta_{i1}) + \exp(2\theta_v - \delta_{i1} - \delta_{i2})}$$

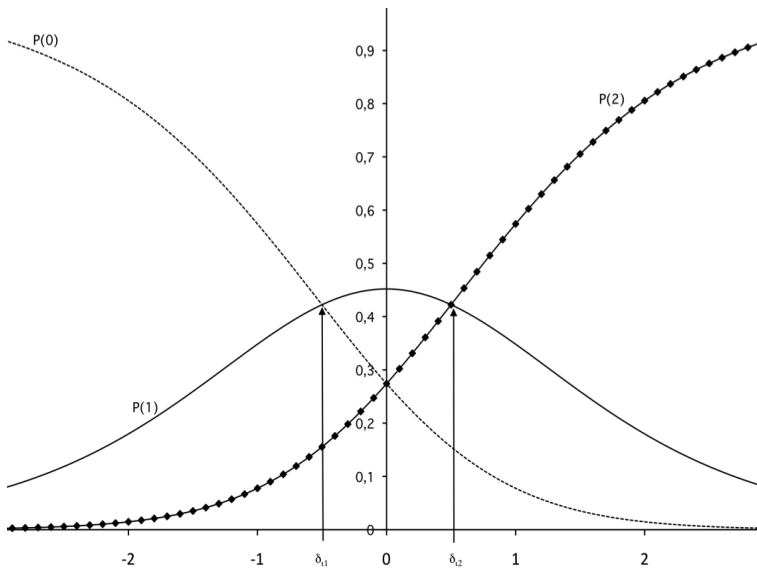
$$\begin{aligned} P(X_{vi} = 2 | \theta_v, \delta_{i2}) &= \frac{\exp[(\theta_v - \delta_{i1}) + (\theta_v - \delta_{i2})]}{1 + \exp(\theta_v - \delta_{i1}) + \exp(2\theta_v - \delta_{i1} - \delta_{i2})} \\ &= \frac{\exp(2\theta_v - \delta_{i1} - \delta_{i2})}{1 + \exp(\theta_v - \delta_{i1}) + \exp(2\theta_v - \delta_{i1} - \delta_{i2})} \end{aligned}$$

Die step difficulties innerhalb eines Items repräsentieren folglich *nicht* den Punkt auf dem Kontinuum der Merkmalsausprägung, an dem ein Proband das Item bzw. die Itemschritte mit einer Wahrscheinlichkeit von $P = 0,5$ lösen kann, sondern die relative Schwierigkeit eines jeden Steps j innerhalb des Items i . Mit δ_{ij} wird folglich beschrieben, an welcher Stelle des Kontinuums der Merkmalsausprägung sich zwei benachbarte Kategorienkurven schneiden. Diese Koordinaten, die die latente Variable in Schwellenabschnitte zerlegen, bestimmen die Kategorienvahrscheinlichkeiten (vgl. Abbildung 4.7).

Die Schwellenschwierigkeiten δ_{ij} werden folglich über die Schnittpunkte der einzelnen nicht-kumulativen Wahrscheinlichkeitsfunktionen der Antwortkategorien bestimmt. Bei diesem Vorgehen ist es – beispielsweise durch einen flacheren Verlauf der Wahrscheinlichkeitsfunktion $P(1)$ – möglich, dass die Schwellenschwierigkeiten eine umgekehrte Reihenfolge aufweisen. Solche, für die Leistungsmessung nicht ungewöhnlichen Ergebnisse zeigen auf, dass es (1) einfacher sein kann, die dritte Antwortkategorie statt der zweiten zu erreichen, und dass (2) manche Items der Leistungsmessung, von denen beispielsweise eine dreikategoriale ordinale Struktur (0, 1 und 2) angenommen wurde, doch nur zwei Antwortkategorien (0 und 2) aufweisen. Stimmen die Schwellenschwierigkeit δ_{ij} mit dem Fähigkeitsparameter θ_v einer Person v überein, beträgt die Wahr-

scheinlichkeit, dass das Antwortverhalten der Person in eine der beiden benachbarten Antwortkategorien fällt, $P = 0,5$.

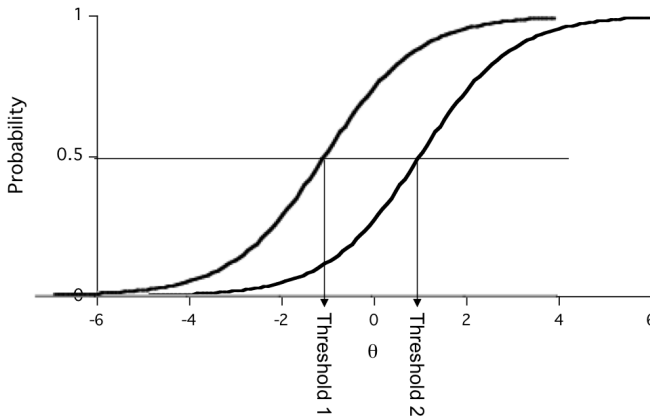
Abbildung 4.7 Graphische Darstellung verschiedener Schwellenschwierigkeiten



Um in der Darstellung des eindimensionalen Rasch-Modells zu bleiben, lassen sich die einzelnen Kategorienwahrscheinlichkeiten in Form einer itemcharakteristischen Funktion (item characteristic curve; ICC) darstellen. Die Schnittpunkte, an denen die kumulierten Wahrscheinlichkeitskurven die $P = 0,5$ -Linie kreuzen, werden als *Item Thresholds* bezeichnet (vgl. Abbildung 4.8). Item Thresholds beschreiben damit den Ort auf dem Kontinuum der Merkmalsausprägung, bei dem eine Person v mit einer Personenfähigkeit θ_v , die dem Threshold 1 entspricht, mit einer Wahrscheinlichkeit von $P = 0,5$ in der Antwortkategorie 1 oder höher antwortet. Bei einer Personenfähigkeit θ_v , die unterhalb des Thresholds 1 liegt, ist es hingegen wahrscheinlicher, dass diese Person das Item nicht beantworten kann (Antwortkategorie = 0). Die Darstellung der Parameter des Partial Credit Modells mit Hilfe der Item Thresholds hat insbesondere folgende Vorteile: (1) Die Kumulierung der Wahrscheinlichkeitsfunktionen stellt sicher, dass die Thresholds der Antwortkategorien stets in der vom Kodierer intendierten Reihenfolge abgetragen werden. (2) Die Fähigkeitsparameter und Itemparameter des PCM können auf Basis eines normierten Wahrscheinlichkeitswerts ($P = 0,5$) aufeinander bezogen werden. Die graphische Darstellung der Thresholds zeigt darüber hinaus, dass beim PCM aus einem ordinalen Item mit x Antwortkategorien ($x-1$) Pseudo-Items

konstruiert werden. Für ein Item mit drei Antwortkategorien ergeben sich damit zwei itemcharakteristische Funktionen (vgl. Abbildung 4.8).

Abbildung 4.8 Graphische Darstellung der Item Thresholds



Konfirmatorische mehrdimensionale IRT-Modelle

Für Leistungstests, wie z. B. die PISA-Studien, ist es üblich, verschiedene Fähigkeitsstrukturen anzunehmen. Die Items eines Tests sind daher so konstruiert, dass mit ihnen verschiedene latente Variablen gemessen werden können. Besteht eine theoretisch begründete Annahme darüber, welche latenten Variablen ein Test erfasst, werden konfirmatorische mehrdimensionale IRT-Modelle benötigt. Im Folgenden wird aus der Vielzahl der Modelle (vgl. u. a. Walker & Beretvas, 2000; Ackerman, 1994) das Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM; Adams, Wilson & Wang, 1997; Briggs & Wilson, 2004; für den eindimensionalen Fall vgl. Adams & Wilson, 1996) vorgestellt. Diese Auswahl wird wie folgt begründet:

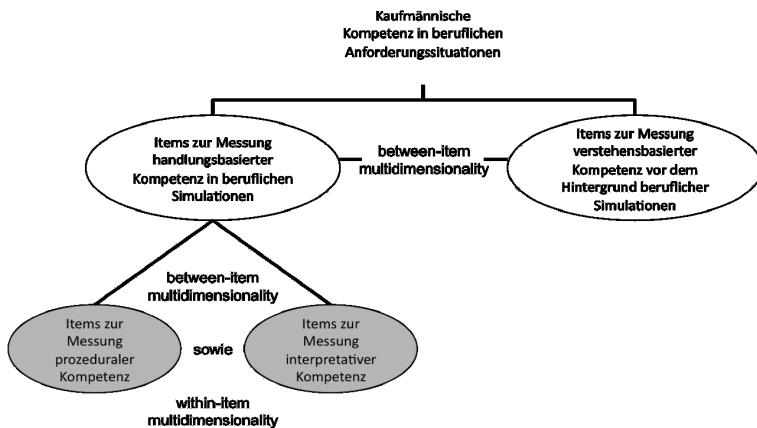
- Das MRCML-Modell ist ein IRT-Modell der Rasch-Familie; es kann damit sowohl das 1-parametrische logistische Modell als auch das Partial Credit Model repräsentieren. Es ist folglich möglich, *dichotome und ordinale Daten* in dem Modell zu verarbeiten.
- Die Schätzung der Personen- und Itemparameter des MRCML-Modells basieren auf *Maximum Likelihood-Techniken*, die ebenfalls für die eindimensionalen Modelle angewendet werden (vgl. Kapitel 4.3).
- Das MRCML-Modell erlaubt eine Prüfung hinsichtlich *between-item- und within-item-Dimensionalitäten*. Between-item-Mehrdimensionalität bedeutet, dass ein Item nur eine der möglichen latenten Variablen misst; within-

item-Mehrdimensionalität sagt hingegen aus, dass ein Item mehrere Personenmerkmale messen kann.

- Das MRCML-Modell ist als generelles mehrdimensionales IRT-Modell in die *Software ConQuest* (Wu, Adams & Wilson, 1997) implementiert. Für die Auswertungen der in Kapitel 5 beschriebenen Studien zur Messung kaufmännischer Kompetenz wird im Rahmen der IRT auf diese Software zurückgegriffen.

Die Studien zur Messung kaufmännischer Kompetenz machen sehr anschaulich deutlich, aus welchen Gründen mehrdimensionale Modelle insbesondere für kognitive Tests gefunden werden müssen. Im Rahmen des DFG-Projekts „Integrierte Kompetenzförderung in den beruflichen Fächern des Wirtschaftsgymnasiums“ wurde ein eindimensional ausgerichteter Test zur Erfassung des Vorwissens der Lernenden konstruiert (Winther, 2006; vgl. Kapitel 5.1). Vertiefende Analysen unter Rückgriff auf die Klassische Testtheorie zeigten jedoch, dass der Test zwei reliable latente Dimensionen erfasst – economic literacy und economic numeracy.

Abbildung 4.9 Mehrdimensionalität des Tests zur Messung kaufmännischer Kompetenz



Neben dem Gesamtscore für das Vorwissen wurden folglich auch Werte für die Subskalen economic literacy und numeracy ausgegeben. Die Fähigkeiten der Lernenden in den Teilbereichen economic literacy und numeracy haben den Erfolg der Lernenden in den Unterrichtssequenzen am Wirtschaftsgymnasium unterschiedlich stark beeinflusst (vgl. Winther, 2007). Das Fehlen von testunabhängigen und stichprobenunabhängigen Schätzern der Personenfähigkeiten hat in diesem Zusammenhang eine weitere Auswertung dieser Effekte stark eingeschränkt. In einer Re-Analyse dieses Vorwissenstests werden in Kapitel 5.1 zwei latente

Variablen angenommen und die Personen- und Itemparameter im Rahmen des konfirmatorischen IRT-Modells unabhängig voneinander geschätzt. Im Rahmen des BMBF-Projekts „Computergestützte Messung berufsfachlicher Kompetenz“ (Winther & Achtenhagen, 2009b; 2010; vgl. Kapitel 5.2) wurde der Test hingegen auf Basis eines Kompetenzstrukturmodells für die kaufmännische Bildung (vgl. Kapitel 2.4.2 sowie 3.3.4) bereits in einer mehrdimensionalen Struktur angelegt. Der Test enthält Items im Hinblick auf eine between-item-Mehrdimensionalität und einige Itemkonstellationen, die das Zusammenspiel zwischen prozeduralen und interpretativen Kompetenzbereichen im Sinne einer within-item-Mehrdimensionalität abbilden. Hat also die Leistungsfähigkeit eine mehrdimensionale Struktur, stellt die Modellierung der einzelnen, jedoch nicht notwendiger Weise unverbundenen, Dimensionen mit Hilfe der IRT die Bedingung dafür dar, dass Personen- und Itemparameter valide geschätzt und die Relationen zwischen den Dimensionen klar beschrieben werden können (Briggs & Wilson, 2004).

Das MRCML-Modell ist eine Erweiterung des Rasch-Modells; ein Vergleich des logarithmierten Wettquotienten (log odds) macht dies besonderes deutlich: Hierfür sei angenommen, das ein Item i mit k ($k = 1, \dots, K$) Antwortkategorien eine oder mehrere latente Dimensionen d ($d = 1, \dots, D$) messen kann. Das Antwortverhalten der Personen wird mit n ($n = 1, \dots, N$) indiziert. Der Wettquotient ermittelt sich aus der Kategorienwahrscheinlichkeit des Antwortverhaltens für das Item i (P_{ik}) im Vergleich zur Kategorienwahrscheinlichkeit $k-1$ ($P_{i(k-1)}$). Der logarithmierte Wettquotient ist eine lineare Funktion der Personenfähigkeit in der Dimension d (θ_d) und der relativen Itemschwierigkeit der Kategorie k (δ_{ik}).

$$\ln \left[\frac{P_{ik}}{P_{i(k-1)}} \right] = \theta_d - \delta_{ik}$$

Jede Person v ist folglich über ihr Profil der Personenparameter θ_v ($\theta_{v1}, \dots, \theta_{vD}$) für jede gemessene latente Dimension d beschrieben. Der Itemparameter δ_{ik} des Items i für die Antwortkategorie k wird in Anlehnung an die Ausführungen zum Partial Credit Model als Schwellenschwierigkeit (step difficulty) bezeichnet; er drückt die relative Schwierigkeit aus, in der Kategorie k anstatt in der Kategorie $k-1$ geantwortet zu haben. Für den Fall, dass between-item-Mehrdimensionalität vorliegt, misst das Item i nur die latente Dimension d und der Itemparameter δ_{ik} bildet ab, ob es relativ einfacher oder schwieriger für eine Testperson ist, in der Antwortkategorie $k-1$ im Vergleich zu Antwortkategorie k klassifiziert zu werden. Mathematisch formal wird das MRCML Model wie folgt beschrieben: Jedes Item i ($i = 1, \dots, n$) hat ($K_i + 1$) Antwortkategorien ($k = 1, \dots, K_i$) und kann eine oder mehrere Dimensionen d ($d = 1, \dots, D$) messen. Es wird davon ausgegangen, dass ein Set von d Personenfähigkeiten (traits) das Antwortverhalten der Testperson bestimmt, so dass die Position einer Person v in dem d -dimensionalen latenten

Raum durch den Vektor $\theta = (\theta_1, \dots, \theta_d)$ lokalisiert werden kann. θ_{vd} ist damit der Personenparameter der Person v in der Dimension d . Der Personenparameter wird durch den Vektor $X_i = (X_{i1}, \dots, X_{iK_i})$ ausgedrückt, der die $(K_i + 1)$ Antwortmöglichkeiten des Item i anzeigt; die Anzahl der Zeilen entsprechen den hypothetischen Dimensionen des Testinstruments.

$$P(X_{ik} = 1; A, B, \xi | \theta_{vd}) = \frac{\exp(b_{ik}\theta_{vd} + a'_{ik}\xi)}{\sum_{k=1}^{K_i} \exp(b_{ik}\theta_{vd} + a'_{ik}\xi)}$$

mit:

$$X_{ik} = \begin{cases} 1 & \text{if response to item } i \text{ is in category } k \\ 0 & \text{otherwise} \end{cases}$$

Die Item- und Kategorienparameter werden durch den Vektor $\xi' = (\xi_1, \dots, \xi_p)$ beschrieben. Um im Modell der Antwortwahrscheinlichkeiten die empirische Charakteristik der Antwortkategorien eines jeden Items zu beschreiben, werden Linearkombinationen gebildet. Die Linearkombinationen sind über Designvektoren a'_{ik} ($i = 1, \dots, n$, $k = 1, \dots, K_i$), jeder mit einer Länge p , abgebildet, die in eine Designmatrix $A = (a_{11}, \dots, a_{1K_1}, a_{21}, \dots, a_{2K_2}, \dots, a_{1K_n})$ überführt werden können. Die Matrix enthält folglich zusammenfassende Aussagen über die Schwierigkeitsparameter eines Items. Die Zuordnung der Items zu den latenten Dimensionen erfolgt mit Hilfe einer Scoringmatrix B . Diese Matrix enthält für jedes Item die Information darüber, welche Dimension(en) es misst (Zeilen), und für jede Dimension ist ersichtlich, durch welche Items sie repräsentiert ist (Spalten). Die einzelnen Scores der d -Dimensionen werden als Spaltenvektor b_{ik} geschrieben, aus denen sich Scoring-Submatrizen der Items i ergeben, die in der Scoringmatrix B für den gesamten Test zusammengefasst werden (Adams, Wilson & Wang, 1997, pp. 2ff.).

Die Unterschiede im Hinblick auf ein eindimensionales IRT-Modell zeigen sich (a) in der Darstellung der Personenfähigkeit, die im eindimensionalen Modell als Skalarprodukt und im mehrdimensionalen Modell als $d \times 1$ -Spaltenvektor beschrieben ist, und (b) darin, dass die Scoringfunktion im mehrdimensionalen Modell zusätzlich Informationen über die Zuordnung der Items zu den d Dimensionen enthält (vgl. Briggs & Wilson, 2004, p. 326).

Im Rahmen des 5. Kapitels werden die hier vorgestellten Modelle am Beispiel konkreter Tests zur Messung kaufmännischer Kompetenz in ihrer Anwendung beschrieben, und es wird gezeigt, unter welchen Bedingungen die Modelle zum Einsatz kommen und welche Ergebnisse sich mit ihnen erzielen lassen. Für eine Interpretation der Ergebnisse sind neben dem Verständnis der grundlegenden Modellstrukturen auch Kenntnisse im Hinblick auf die Aussagekraft der einzelnen Modellparameter sowie deren Schätzung von Nutzen. Die Anzahl der Parameter

in den einzelnen Modellen variiert; sie lässt sich jedoch generell auf Item- und Personenparameter reduzieren: Die Schwellenschwierigkeiten der ordinalen Items im Partial Credit-Modell werden vergleichbar wie die Itemschwierigkeiten im Rasch-Modell geschätzt, und die Grundlagen der Schätzung der Personenparameter sind für ein- und mehrdimensionale Modelle vergleichbar.

4.3 Schätzung der Modellparameter

Über die Wahl des Messmodells werden Entscheidungen darüber getroffen, wie eine Variable mathematisch definiert wird und wie sich Personen auf dieser Variable messen lassen. Hierzu ist zu fragen, wo Items und Personen auf der interessierenden Variable positioniert sind und welche Schlussfolgerungen sich daraus für die Güte des gewählten Modells ableiten lassen. Im Folgenden sind diese Fragen unter einer anderen Akzentuierung zu stellen: Im Rahmen der Parameterschätzung ist von Interesse, (a) wie gut die Daten auf das gewählte Messmodell passen (Modellfit), (b) wie gut die beobachteten Antworten auf ein Item i mit den erwarteten Werten des Messmodells übereinstimmen (Itemfit) und (c) wie gut das Testverhalten einer Person v die erwarteten Werte des Messmodells repräsentiert (Personenfit). Diese Fragen erfassen die Anforderungen der Validitätsprüfung in vollem Umfang, da mit ihnen beantwortet werden kann, ob die Items eines Tests die zu messende Variable tatsächlich abbilden und ob die Testpersonen die Items in dem vom Testentwickler intendierten Sinne bearbeiten.

„Even the most inspired and disciplined selection and administration of best possible items to appropriate and cooperative persons cannot guarantee the definition of a useful variable nor the successful measurement of a person“ (Wright & Masters, 1976, p. 90).

Im Folgenden werden Verfahren der Parameterschätzung vorgestellt. Hierbei werden jene Verfahren genauer betrachtet, die für die in Kapitel 5 dargestellten Studien zur Messung kaufmännischer Kompetenz von Bedeutung sind und die in der Software *ConQuest* (Wu, Adams & Wilson, 1997) implementiert sind. Vorrangig sind dies Schätzverfahren nach dem Maximum Likelihood-Prinzip. Zur besseren Veranschaulichung wird zunächst das grundlegende Verfahren am Beispiel des Rasch-Modells vorgestellt, bevor auf die Item- und Personenparameterschätzung bei Modellen der Rasch-Familie eingegangen wird.

4.3.1 Maximum Likelihood-Schätzung

Verfahren der Maximum Likelihood-Schätzung für IRT-Parameter zeichnen sich unter der Bedingung einer großen Stichprobe insbesondere durch folgende positive Eigenschaften aus: (1) Konsistenz in dem Sinne, dass der Schätzwert mit zunehmender Stichprobengröße gegen den wahren Wert konvergiert; (2) Effizienz, die in einem kleinen Standardfehler zum Ausdruck kommt, und (3) die Annahme der Normalverteilung des Schätzfehlers. Hinzukommt, dass die Gesamtl likelihood

der Daten in eine Beziehung zu einer Chi-Square-Verteilung gestellt werden kann, was einen Vergleich der Modellgüte auf Basis des Verhältnisses der Likelihood einzelner Modelle erlaubt (vgl. Embretson & Reise, 2001, p. 200).

Die besondere Schwierigkeit der Parameterschätzung besteht darin, dass sowohl die Personenparameter als auch die Itemparameter unbekannt sind. Gerade vor diesem Hintergrund ist es (mathematisch) hilfreich, dass für die Modelle der Rasch-Familie das Prinzip der Separabilität (d. h. die Trennbarkeit der Personenfähigkeit und der Itemschwierigkeit) Gültigkeit besitzt. Hieraus folgt, dass für Rasch-Modelle der Testscore eine suffiziente Statistik zur Schätzung der Personenfähigkeit darstellt und respektive die Anzahl der Personen, die ein Item des Tests erfolgreich lösen, eine suffiziente Statistik für die Schätzung der Itemschwierigkeit ist. Für Rasch-Modelle gilt daher, dass Personen mit demselben Testscore über eine identische Fähigkeitsausprägung verfügen, und dies unabhängig davon, welche spezifischen Items tatsächlich gelöst wurden. Zudem gilt auch, dass Personen, die ein bestimmtes Item gelöst haben oder aber an der Lösung gescheitert sind, keinen Einfluss auf den Schätzer der Itemschwierigkeit haben (für die math. Beweisführung vgl. Baker, 1992). Das Vorliegen einer suffizienten (erschöpfenden) Statistik macht es möglich, bei der Schätzung der Itemparameter die Likelihood ausschließlich unter der Bedingung der Antwortmuster bei gleichen Testscores zu ermitteln. Vor diesem Hintergrund wird das Prinzip des Maximum Likelihood-Verfahrens unter der einschränkenden Bedingung, dass die Personenparameter als bekannt vorausgesetzt werden können, illustriert. Es sei herausgestellt, dass diese Illustration einzig der Veranschaulichung dient und keine direkte praktische Anwendbarkeit besitzt, da in der Realität sowohl Item- als auch Personenparameter unbekannt sind.

Die Likelihoodfunktion beschreibt die Wahrscheinlichkeit der Daten unter der Annahme, dass das gewählte Modell Gültigkeit besitzt. Die Likelihoodfunktion ist damit eine Funktion mehrerer Veränderlicher, die das Modell beschreiben. Die Anzahl der veränderlichen Größen – die Anzahl der Modellparameter – ist abhängig von der Wahl des Messmodells. Beim Rasch-Modell sind zwei Modellparameter zu schätzen: die Personenfähigkeit θ_v und die Itemschwierigkeit δ_i . Die Maximum Likelihood-Methode besagt, dass alle Modellparameter auf den Wert festgelegt werden, an dem die Likelihoodfunktion ihr Maximum besitzt (u. a. Rost, 2004). Unter der Annahme, dass der Wert der Personenfähigkeit determiniert ist, wird im Folgenden ausgehend von der Definition der Wahrscheinlichkeiten einer Antwort auf das Item i die Gesamtlikelihood der Antworten aller Personen auf alle Items eines Tests definiert. Die Notation orientiert sich dabei an Embretson und Reise (2001) sowie an Wilson (2005).

Für dichotome Items lassen sich basierend auf dem Rasch-Modell zwei Antwortwahrscheinlichkeiten bestimmen: P_{vi} für das Lösen und Q_{vi} für das Scheitern. Die Wahrscheinlichkeiten für eine korrekte und inkorrekte Antwort lassen sich durch einen „mathematischen Trick“ in eine Gleichung des beobachteten Antwortverhaltens, X_{vi} , überführen. Der Trick besteht darin, eine Gleichung zu finden, in der über den Exponenten kontrolliert werden kann, ob die Wahrscheinlichkeit

für korrekte oder inkorrekte Antworten definiert wird. Diese Gleichung ist die Likelihoodfunktion für die Antwortwahrscheinlichkeit eines Items. Gibt eine Person v eine korrekte Antwort auf das Item i , $X_{vi} = 1$, wird die Wahrscheinlichkeit als $P_{vi}^1 Q_{vi}^{1-1}$, also P_{vi} , definiert; bei einer inkorrekten Antwort ergibt sich $P_{vi}^0 Q_{vi}^{1-0}$, also Q_{vi} .

Antwortwahrscheinlichkeit für eine korrekte Antwort:

$$P(X_{vi} = 1 | \theta_v, \delta_i) = P_{vi} = \frac{\exp(\theta_v - \delta_i)}{1 + \exp(\theta_v - \delta_i)}$$

Antwortwahrscheinlichkeit für eine inkorrekte Antwort:

$$P(X_{vi} = 0 | \theta_v, \delta_i) = Q_{vi} = \frac{1}{1 + \exp(\theta_v - \delta_i)}$$

Likelihoodfunktion für die Antwortwahrscheinlichkeit eines Items:

$$P(X_{vi} | \theta_v, \delta_i) = P_{vi}^{X_{vi}} Q_{vi}^{1-X_{vi}}$$

Die Likelihoodfunktion für das Antwortverhalten einer Person v auf alle Items (X_{v1}, \dots, X_{vI}) ergibt sich als Produkt der Antwortwahrscheinlichkeiten eines jeden Items i . Äquivalent dazu ergibt sich die Likelihoodfunktion des Antwortverhaltens aller Personen als Produkt des Antwortmusters einer jeden Person.

Likelihoodfunktion des Antwortverhaltens (alle Items):

$$P(X_{v1}, \dots, X_{vI} | \theta_v, \delta) = P(X_v | \theta_v, \delta) = \prod_i P_{vi}^{X_{vi}} Q_{vi}^{1-X_{vi}}$$

Likelihoodfunktion aller Daten (alle Items und alle Personen):

$$P(X_{11}, \dots, X_{NI} | \theta, \delta) = \prod_v P(X_v | \theta_v, \delta) = \prod_v \prod_i P_{vi}^{X_{vi}} Q_{vi}^{1-X_{vi}}$$

Zur Bestimmung des Maximums der Likelihoodfunktion aller Daten ist die erste partielle Ableitung zu ermitteln. Um dieses Vorgehen zu vereinfachen, ist es gängige Praxis, die Likelihoodfunktion vor dem partiellen Differenzieren zu logarithmieren. Hierdurch ändert sich der Ort ihres Maximums nicht, jedoch aber der Wertebereich der Funktion. Die Likelihoodfunktion aller Daten kann stets nur Werte zwischen 0 und 1 annehmen, da sie als Produkt von Wahrscheinlichkeiten

definiert ist. Die Werte der Log-Likelihoodfunktion sind hingegen negativ, da der Logarithmus von Zahlen zwischen 0 und 1 im Wertebereich von $-\infty$ und 0 liegt. Für die Parameterschätzung werden geeignete Computerprogramme verwendet, mit denen die Schätzer iterativ gewonnen werden.

Log-Likelihoodfunktion aller Daten (alle Items und alle Personen):

$$\ln(P(X|\theta, \delta)) = \sum_v \sum_i \ln(P_{vi}^{X_{vi}} Q_{vi}^{1-X_{vi}})$$

Die Software *ConQuest* stellt den Wert der Log-Likelihoodfunktion auf Modell-ebene als Deviance-Wert zur Verfügung. Die Deviance ist ein Wert der Modellgüte und berechnet sich, indem die Log-Likelihood mit -2 multipliziert wird. Die Genauigkeit der Schätzung der Deviance ist dabei abhängig vom gewählten Abbruchkriterium des Modells. Als Abbruchkriterium kommen (1) eine festgelegte Anzahl von Iterationen, bei deren Erreichen der Schätzvorgang abgebrochen wird, oder (2) Differenzwerte in Betracht, die angeben, um welchen Wert sich die Log-Likelihood von einer Iteration zur nächsten unterscheiden soll. Wird dieser Wert unterschritten, ist der Algorithmus abzubrechen. Rost (2004) weist darauf hin, dass die Wahl des Abbruchkriteriums von der Beschaffenheit der Likelihoodfunktion abhängig sein sollte (Rost, 2004, S. 307): Eine Funktion, die einen flachen Verlauf oder mehrere lokale Maxima aufweist, setzt folglich zur Absicherung des „wahren“ Maximums mehr Iterationsvorgänge und damit sensitivere Genauigkeitskriterien voraus als eine Funktion mit eindeutigem Maximum. Für die Praxis der IRT-Modellberechnung spielen diese Überlegungen kaum mehr eine Rolle, da in die zur Berechnung verwendete Software generell sehr selektive Genauigkeitskriterien implementiert sind.

In Abbildung 4.10 ist auf Basis des Datensatzes des BMBF-Projekts „Computergestützte Messung berufsfachlicher Kompetenz“ (Winther & Achtenhagen, 2009b; Winther & Achtenhagen, 2010) die Schätzung des Modellfits auf Basis eines Partial Credit-Modells dargestellt. Aus dem Wert der Deviance lässt sich die Log-Likelihood ableiten; sie beträgt im Beispiel -7.201,7933. Der Modellfit wurde in 29 Iterationen geschätzt. Als Abbruchkriterien sind sowohl die Zahl der maximalen Iterationsvorgänge als auch Veränderungsgrenzen festgelegt. Im gegebenen Beispiel würde der Iterationsprozess bei 1000 durchgeführten Iterationen oder aber spätestens dann abgebrochen, wenn von einer Iteration zur nächsten die Deviance sich um weniger als 0,0001 verändert. Die (Log-)Likelihoodfunktion beschreibt die Wahrscheinlichkeit der beobachteten Testdaten unter der Bedingung, dass das gewählte Testmodell gilt. Auf Basis des Wertes der (Log-)Likelihoodfunktion kann folglich festgestellt werden, wie wahrscheinlich die Testdaten in dem gewählten Modell tatsächlich sind. Der Wert der Funktion (in *ConQuest* die Deviance) kann genutzt werden, um festzustellen, ob bei Wahl eines

anderen Modells höhere Wahrscheinlichkeiten vorliegen und dieses Modell damit einen besseren Modellfit aufweist. Die Prüfung der Modellgüte im Vergleich zwischen Modellen erfolgt auf Basis der Chi-Square-Statistik. Für den Vergleich der Modellgüte in *ConQuest* gilt, dass die Differenz der Deviance-Werte mittels der Chi-Square-Statistik zu prüfen ist. Ist der Differenzwert statistisch signifikant, fittet das Modell mit der geringeren Deviance (respektive der größeren Log-Likelihood) die Daten besser (vgl. hierzu Kapitel 5.1.3.2 sowie Kapitel 5.2.4.2).

Abbildung 4.10 *Modellfit auf Basis der Log-Likelihoodfunktion in ConQuest*

```

=====
ConQuest: Generalised Item Response Modelling Software
SUMMARY OF THE ESTIMATION
=====

Estimation method was: Gauss-Hermite Quadrature with 225 nodes
Assumed population distribution was: Gaussian
Constraint was: NONE
The Data File: ALUSIM_ALL.dat
The format: id 1-7 responses 8-67
The regression model:
Grouping Variables:
The item model: item+item*step
Sample size: 265
Final Deviance: 14403.58654
Total number of estimated parameters: 83
The number of iterations: 29
Termination criteria: Max iterations=1000, Parameter Change= 0.00010
Deviance Change= 0.00010

Die Log-Likelihood
auf Modellebene beträgt -7201,7933.

Die Log-Likelihood
wurde in 29 Iterationen geschätzt.

```

Im Folgenden wird aufgezeigt, wie auf Item- und Personenebene Schätzer des gewählten Modells bestimmt werden können.

4.3.2 Schätzung der Itemparameter

Mit Hilfe der Software *ConQuest* erfolgt die Itemparameterschätzung nach der marginalen Maximum Likelihood-Methode (MML). Diese erlaubt eine unabhängige Schätzung der Itemparameter von den Personenparametern. Hierfür enthält die Likelihoodfunktion neben den Itemparametern nicht die Personenparameter selbst, sondern die Verteilungsparameter der Personenvariablen. Dieser „mathematische Kniff“ basiert auf der Annahme der Separabilität beider Parameter und darauf, dass in Modellen der Rasch-Familie mit den Testscores eine suffiziente Statistik zur Personenparameterschätzung und mit der Anzahl der Personen, die ein Item des Tests korrekt gelöst haben, eine suffiziente Statistik für die Schätzung der Itemschwierigkeiten vorliegt. Die Möglichkeit der unabhängigen Schätzung verhindert, dass die Schätzung der Itemparameter unter der Ungenauigkeit der Personenparameterschätzung leiden würde (vgl. Rost, 2004, S. 310).

Die Grundidee der marginalen Maximum Likelihood-Methode ist, dass nicht die einzelnen Personenparameter, sondern nur diejenigen Parameter in der Schätzfunktion berücksichtigt werden, die die Verteilung der Personenparameter

beschreiben (i. d. R. wird eine Normalverteilung angenommen; so auch bei *ConQuest*). Das mit einem Test erfasste Personenmerkmal wird vor diesem Hintergrund als Dichtefunktion $f(\theta)$ berücksichtigt. Damit die Itemparameter bei der Schätzung nicht mehr von den Personenparametern abhängen, wird über die Dichtefunktion integriert.

Marginale Likelihoodfunktion aller Daten:

$$MML = f(\text{Daten} | \delta, f(\theta))$$

Die marginale Likelihoodfunktion des Antwortverhaltens einer Person v über alle Items i gibt folglich die Wahrscheinlichkeit des Auftretens eines Antwortverhaltens (Antwortvektor) an, indem über die Verteilung der traits integriert wird (Embretson & Reise, 2001, p. 212). Die marginale Likelihoodfunktion über alle Personen ergibt sich damit aus dem Produkt der Likelihoodfunktionen des Antwortverhaltens einer jeden Person multipliziert mit der Dichtefunktion der Personenparameter.

marginale Likelihoodfunktion des Antwortverhaltens (alle Items):

$$P(X_{v1}, \dots, X_{vi} | \theta_v, \delta) = \int P(X_v | \theta_v, \delta) f(\theta) d\theta = \int \prod_i P_{vi}^{X_{vi}} Q_{vi}^{1-X_{vi}} f(\theta) d\theta$$

marginale Likelihoodfunktion aller Daten (alle Items und alle Personen):

$$P(X_{11}, \dots, X_{Ni} | \theta, \delta) = \prod_{v=1}^N \int P(X_v | \theta_v, \delta) f(\theta) d\theta = \prod_{v=1}^N \int \prod_i P_{vi}^{X_{vi}} Q_{vi}^{1-X_{vi}} f(\theta) d\theta$$

Zur Schätzung der Itemparameter wird die marginale Likelihoodfunktion logarithmiert und ihre erste partielle Ableitung nach den Itemparametern gebildet. Für diesen Schritt sowie für die Lösung der Gleichungen stehen Computerprogramme wie *ConQuest* zur Verfügung, mit denen die Itemparameter in iterativen Verfahren geschätzt werden können (für die Beschreibung des Verfahrens vgl. Walter, 2005). Zur Interpretation der Outputwerte ist es darüber hinaus interessant zu wissen, dass die Lösung der Schätzgleichungen über numerische Integrationen erfolgt. Bei der numerischen Integration wird das latente Kontinuum in mehrere Abschnitte um einen neutralen Nullpunkt zerlegt – technisch ist dies nichts anderes, als die angenommene Normalverteilung zu segmentieren. Hierdurch ergeben sich Itemschätzwerte, die im negativen und positiven Bereich um den Nullpunkt streuen, wobei negative Schätzwerte für eine geringe Item-

schwierigkeit und positive Schätzwerte – relativ gesehen – für eine höhere Itemschwierigkeit stehen.⁸ Neben der Schätzung der Itemschwierigkeiten ist der Itemfit zu bestimmen. Während die Itemschwierigkeit die lokale Position eines Items auf der zu messenden Variablen beschreibt, gibt der Itemfit Hinweise darauf, mit welcher Qualität ein Item die zu messende Variable abbildet. Der Itemfit ist umso besser, je weniger die erwarteten Antworten im Hinblick auf ein Item $E(X_{vi})$ von den beobachteten Antworten X_{vi} abweichen, je geringer folglich die Differenz Y_{vi} zwischen dem beobachteten und dem erwarteten Wert ist. Für das Rasch-Modell gilt, dass sich der Erwartungswert aus der Summe der Antwortwahrscheinlichkeiten $P(X_{vi})$ einer Person v auf ein Item i ergibt. Für den Fall ordinaler Daten ist der Erwartungswert die Summe der Antwortwahrscheinlichkeiten über die jeweiligen Antwortkategorien k eines Items i . Das Antwortverhalten X_{vi} ist durch seine Varianz Var_{vi} und Kurtosis $Kurt_{vi}$ beschrieben.

Residualgröße des beobachteten und erwarteten Antwortverhaltens:

$$Y_{vi} = X_{vi} - E(X_{vi}) = X_{vi} - \sum_{k=0}^m k P(X_{vi})$$

Varianz von X_{vi} :

$$Var_{vi} = \sum_{k=0}^m (k - E(X_{vi}))^2 P(X_{vi})$$

Kurtosis von X_{vi} :

$$Kurt_{vi} = \sum_{k=0}^m (k - E(X_{vi}))^4 P(X_{vi})$$

Standardisierte Residualgröße:

$$Z_{vi} = Y_{vi} / \sqrt{Var_{vi}}$$

Die Varianz Var_{vi} ist gleichsam auch die Varianz der Differenz von beobachtetem und erwartetem Antwortverhalten und kann dazu genutzt werden, die Residualgröße zu standardisieren. Mit Hilfe der standardisierten Residualgröße ist es möglich, für jede Person v und jedes Item i Differenzen vom Erwartungswert $E(X_{vi}) = 0$ in Form von Standardabweichungen anzugeben.

Darüber hinaus beziehen sich Aussagen zum Itemfit auch darauf, wie ein Item generell über alle Personen hinweg zum gewählten Messmodell passt. Hierzu ist zu prüfen, ob die Itemkalibrierung durch eine zuverlässige Anzahl an Personen validiert werden kann. Oder anders formuliert: Je mehr Personen eine unerwartete Antwort auf ein Item i geben, desto schlechter ist dessen Itemfit. Mit sinkendem Itemfit steigt die Wahrscheinlichkeit, dass das entsprechende Item im Gegensatz zu der Mehrzahl der Items in einem Test eine andere latente Variable

⁸ Die Software *ConQuest* normiert die Segmentierung der angenommenen Normalverteilung im Bereich von -6 bis +6 auf der Logit-Skala.

misst (vgl. Wright & Masters, 1982, p. 100). Eine Möglichkeit, den Fit eines Items im Hinblick auf ein gewähltes Messmodell zusammenzufassen ist, jede der standardisierten Residualgrößen zu quadrieren und über die Anzahl aller Personen N zu mitteln. Die so entstehende Größe der ungewichteten Abweichungsquadrate (unweighted Mean Square; MNSQ) beschreibt den OUTFIT eines Items i bezogen auf das gewählte Messmodell. Wright und Masters (1982, p. 99) weisen darauf hin, dass dieser Kennwert des Itemfits extrem sensitiv auf Ausreißer in den Daten reagiert, so dass bereits dann, wenn nur sehr vereinzelt (extrem) unerwartete Antworten auf ein Item i gegeben werden, ein Missfit des Items zum Modell angezeigt wird.

Mean Square (MNSQ) / OUTFIT: Varianz des MNSQ:

$$MNSQ = \frac{\sum_{v=1}^N Z_{vi}^2}{N} \qquad \qquad \qquad Var(MNSQ) = \sum_{v=1}^N \frac{(Kurt_{vi} / Var_{vi}^2)}{N^2} = \frac{1}{N}$$

Als Alternative wird vorgeschlagen, die quadrierten Residuen zu gewichten, um den Einfluss von Antworten der Personen, für die das Item sehr ungeeignet ist, zu reduzieren. Formal ist der gewichtete Mean Square (weighted MNSQ) wie folgt beschrieben:

weighted Mean Square (wMNSQ) / INFIT: Varianz des wMNSQ:

$$wMNSQ = \frac{\sum_{v=1}^N Z_{vi}^2 Var_{vi}}{\sum_{v=1}^N Var_{vi}} = \frac{\sum_{v=1}^N Y_{vi}^2}{\sum_{v=1}^N Var_{vi}} \qquad \qquad \qquad Var(wMNSQ) = \frac{\sum_{v=1}^N (Kurt_{vi} - Var_{vi}^2)}{\left(\sum_{v=1}^N Var_{vi} \right)^2}$$

Der gewichtete Mean Square ist ein Kennwert der INFIT-Statistik mit einem Erwartungswert von 1, bei dem jedes quadrierte Residuum mit seiner Varianz gewichtet wird. Die Varianz Var_{vi} ist am größten, wenn Personen- und Itemschätzer übereinstimmen, und sie sinkt folglich, wenn der Schätzer der Itemschwierigkeit ober- bzw. unterhalb der Personenfähigkeit liegt. Durch dieses Vorgehen verlieren extreme Abweichungen zwischen den Parameterschätzern relativ an Bedeutung. Der weighted MNSQ gilt als der Standardparameter für den Itemfit. Für einen akzeptablen Itemfit sollte der weighted MNSQ innerhalb des Intervalls 0,75 \leq wMNSQ \leq 1,33 liegen (Adams & Khoo, 1996; Bond & Fox, 2001).

In Abbildung 4.11 ist die Zusammenfassung der Itemparameterschätzung am Beispiel von zehn Items des Vorwissenstests zur Messung kaufmännischer Kompetenz im Rahmen des DFG-Projekts „Integrierte Kompetenzförderung in den beruflichen Fächern des Wirtschaftsgymnasiums“ als *ConQuest*-Output dargestellt. Als Itemparameter sind der Schätzer der Itemschwierigkeit sowie dessen Standardfehler gegeben. Die zur Veranschaulichung ausgewählten zehn Testitems definieren die latente Variable „kaufmännische Kompetenz“ auf einer Skala von $-4,135 \leq \delta_i \leq 3,505$ Logits. Die Skalengrenzen sind durch das leichteste Item (Item 10) und durch das schwierigste Item (Item 16) im Test festgelegt. Zur Beurteilung des Itemfits sind sowohl die ungewichteten als auch die gewichteten Abweichungsquadrate gegeben. Weist ein Item einen gewichteten MNSQ von 1,00 auf, liegt ein exakter Fit dieses Items im Hinblick auf das gewählte Messmodell vor. Ist der gewichtete MNSQ eines Items i größer als sein Erwartungswert, fittet dieses Item i das Modell schlecht. Für die Beurteilung eines Missfits gilt, dass der MNSQ nicht größer als 1,33 sein sollte (Adams & Khoo, 1996, zit. n. Wilson, 2005, p. 129). Für verschiedene Studien werden diese Grenzwerte zum Teil auch konservativer gesetzt. Für die PISA-Studien gilt beispielsweise, dass der absolute Wert des gewichteten MNSQ 1,25 nicht übersteigen sollte (PISA 2000 Technical Report; Adams & Wu, 2002).

Abbildung 4.11 Zusammenfassung der Itemparameterschätzung in *ConQuest*

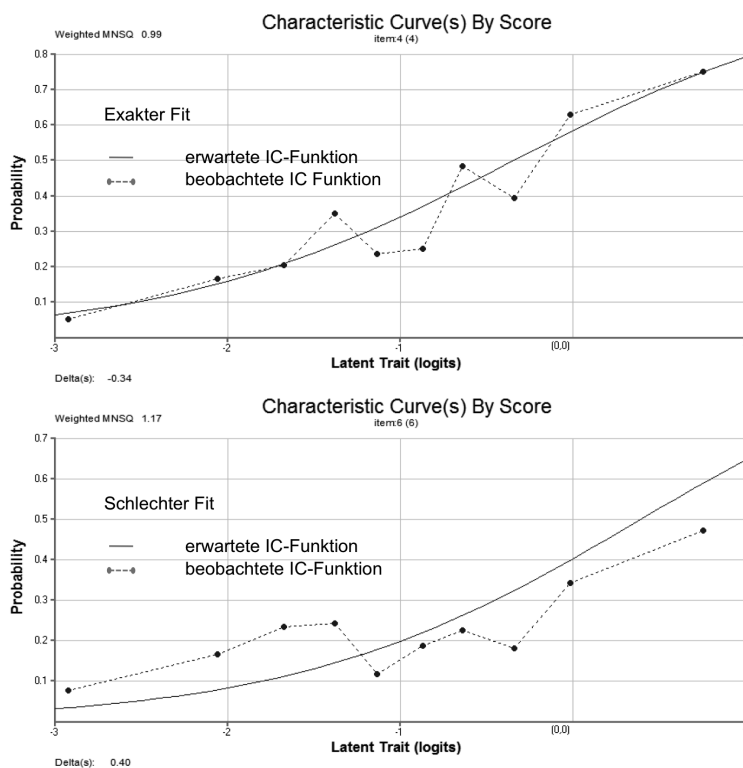
Schätzer der Itemschwierigkeit		Standardfehler				gewichteter Itemfit			
VARIABLES									
item		ESTIMATE	ERROR ^a	UNWEIGHTED FIT			WEIGHTED FIT		
				MNSQ	CI	T	MNSQ	CI	T
1	1	-2.148	0.129	0.97 (0.85, 1.15)	-0.3		0.96 (0.89, 1.11)	-0.6	
4	4	-0.340	0.124	0.99 (0.85, 1.15)	-0.0		0.99 (0.89, 1.11)	-0.2	
5	5	-1.515	0.121	0.95 (0.85, 1.15)	-0.6		0.96 (0.91, 1.09)	-0.8	
6	6	0.396	0.140	1.34 (0.85, 1.15)	4.1		1.17 (0.89, 1.21)	1.8	
10	10	-4.135	0.334	1.29 (0.73, 1.27)	2.0		1.07 (0.53, 1.47)	0.4	
11	11	-2.380	0.221	0.96 (0.73, 1.27)	-0.2		0.92 (0.82, 1.18)	-0.9	
14	14	0.287	0.274	0.90 (0.73, 1.27)	-0.7		0.97 (0.67, 1.33)	-0.1	
15	15	0.444	0.286	1.13 (0.73, 1.27)	1.0		1.07 (0.64, 1.36)	0.4	
16	16	3.505	1.010	0.25 (0.73, 1.27)	-8.1		0.97 (0.00, 2.92)	0.3	
27	27	1.475	0.303	4.99 (0.73, 1.27)	15.7		1.18 (0.69, 1.31)	1.2	

An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability = 0.961
Chi-square test of parameter equality = 1689.21, df = 27, Sig Level = 0.000

Um einschätzen zu können, ob Items eine zu geringe Trennschärfe aufweisen, ist in *ConQuest* zudem die T-Statistik einschließlich des Vertrauensintervalls gegeben. Die T-Statistik prüft, ob die Abweichungen zwischen erwarteten und beo-

bachteten Häufigkeiten statistisch signifikant sind. Da der T-Wert annähernd normalverteilt ist, lässt sich der Signifikanztest anhand der Normalverteilungstabelle vornehmen. Ist der empirisch ermittelte z-Wert größer als der für das $\alpha = 5\%$ -Niveau bei zweiseitigem Test erwartete Testwert ($z = \pm 1,96$) deutet dies auf eine signifikante Abweichung in Richtung einer zu niedrigen Trennschärfe hin (vgl. u. a. Bortz, 2005). In der Folge wäre inhaltlich zu prüfen, ob das Item aus dem Test genommen werden sollte. Für das in Abbildung 4.11 gegebene Beispiel liegt ein solches Item nicht vor.

Abbildung 4.12 IC-Funktionen zur Beurteilung des Itemfits



Ist der gewichtete MNSQ eines Items i kleiner als dessen Erwartungswert von 1, fittet das Item das Modell zu gut. Auch in diesem Fall, sind (1) die zulässige Untergrenze für einen akzeptablen Itemfit (0,75; vgl. Adams & Khoo, 1996) sowie (2) der Wert der T-Statistik zur Beurteilung eines Missfits zu prüfen und es ist gegebenenfalls inhaltlich zu entscheiden, ob Items mit einem Überfit aus dem

Test entfernt werden sollten. Der Itemfit lässt sich mit Hilfe der itemcharakteristischen Funktion (item characteristic curve; ICC) auch graphisch darstellen. In Abbildung 4.12 sind Beispiele eines exakten Fits (Item 4; $wMNSQ = 0,99$) sowie eines schlechten Fits (Item 6; $wMNSQ = 1,17$) im Vergleich gegeben.

Die Graphik zeigt für die gewählten Items eine Gegenüberstellung der erwarteten und beobachteten IC-Funktionen. Fittet ein Item das gewählte Messmodell exakt, entspricht die beobachtete IC-Funktion der erwarteten Funktion weitgehend, was in einem identischen Anstieg beider Funktionen zum Ausdruck kommt. Liegt ein schlechter Itemfit vor, hat die beobachtete im Vergleich zur erwarteten IC-Funktion einen flacheren Verlauf. Im Falle eines Überfits ergäbe sich entsprechend ein steilerer Verlauf der beobachteten IC-Funktion.

Für die Itemparameterschätzung kommen neben der marginalen Maximum Likelihood-Methode (MML) auch andere Verfahren zur Anwendung, die (1) ebenfalls von einer separierten Schätzung von Item- und Personenparametern ausgehen (bedingte (conditional) Maximum Likelihood-Methode; CML), oder in denen (2) die Item- und Personenparameter simultan geschätzt werden (unbedingte (joint) Maximum Likelihood-Methode; JML). Der zentrale, anwendungsbezogene Unterschied zwischen der MML- und der CML-Methode besteht darin, dass die CML-Schätzung vollständig ohne Personenparameter auskommt, während die MML-Methode die Personenparameter über deren Verteilung in der Likelihood berücksichtigt. Dies führt dazu, dass mit der MML-Methode im Gegensatz zur CML-Methode (1) auch die Schätzung von IRT-Modellen möglich ist, die die Eigenschaft der Separabilität von Personen- und Itemparameter nicht aufweisen, und (2) latente Verteilungen und deren Parameter direkt geschätzt werden können. Damit ist die MML-Methode für eine größere Anzahl von Modellen anwendbar und stellt insbesondere für Large Scale-Studien, in denen vor allem die Verteilungen von Personenparametern beispielsweise bezogen auf die latente Variable „Kompetenz“ von Interesse sind, das geeignete Schätzverfahren dar – dies gilt insbesondere auch für Modelle, in denen eine mehrdimensionale Fähigkeitsstruktur angenommen wird. Die JML-Methode hingegen wird in der Testpraxis für Modelle der Rasch-Familie kaum noch genutzt, da die Genauigkeit der Schätzung der Itemparameter durch die Anwesenheit der Personenparameter in der Likelihoodfunktion erheblich beeinflusst wird, so dass die messtheoretisch bedeutsame Eigenschaft der spezifischen Objektivität (vgl. Kapitel 4.2.1) angezweifelt werden kann. Zudem zeigen sich JML-Schätzungen für Itemparameter weder erwartungstreu noch konsistent (zu Vorschlägen der Bias-Korrektur vgl. Wright & Douglas, 1977; für einen detaillierten Vergleich der drei Methoden vgl. Embretson & Reise, 2001; Rost, 2004; Walter, 2005). Allerdings ist die JML-Methode zentral für die Schätzung der Personenparameter nach dem Maximum Likelihood-Prinzip.

4.3.3 Schätzung der Personenparameter

In der einschlägigen Literatur werden i. d. R. drei verschiedene Schätzer für den Parameter einer Person besprochen: der Maximum Likelihood-Schätzer (MLE), der gewichtete (weighted) Likelihood-Schätzer (WLE) und der Bayes-Schätzer

Expected a posteriori (EAP). Diese drei Schätzer stehen auch in der Software *ConQuest* zur Verfügung und werden im Folgenden dargestellt. Für die Personenparameterschätzung gilt generell, dass die zuvor geschätzten Itemparameter als ‚wahre‘ Werte angenommen und damit als gegeben in die unbedingte Likelihoodfunktion eingesetzt werden können. Das Verfahren ähnelt dem in Kapitel 4.3.1 beschriebenen Grundprinzip der Maximum Likelihood-Schätzung, mit dem Unterschied, dass jetzt die Itemparameter und nicht die Parameter der Person als bekannt vorausgesetzt werden.

Maximum Likelihood-Schätzer (MLE)

Bei der Maximum Likelihood-Schätzung wird die Wahrscheinlichkeit der Daten unter der Bedingung der Modellparameter maximiert, wobei für die Schätzung der Personenparameter die Itemparameter bereits als gegebene Größen vorliegen. Die Maximierung der Likelihood erfolgt in zwei Schritten: Im ersten Schritt werden die zuvor geschätzten Itemparameter in die Likelihoodfunktion des Antwortverhaltens einer Person eingesetzt; im zweiten Schritt wird, um die Maxima der Funktion zu bestimmen – wie zuvor beschrieben – für die logarithmierte Likelihoodfunktion die erste partielle Ableitung nach θ_v gebildet und diese gleich Null gesetzt. Da die resultierende Schätzgleichung zur Bestimmung des MLE nicht nach dem gesuchten Personenparameter aufgelöst werden kann, erfolgt eine Annäherung an den Schätzer mit Hilfe der Taylor-Serie. Hierfür ist zusätzlich die zweite partielle Ableitung der Log-Likelihood nach θ_v zu bilden, um auf Basis der Taylor-Serie die Schätzer der Person iterativ zu gewinnen. Für das Rasch-Modell stellt sich die Personenparameterschätzung wie folgt formal dar:

Likelihoodfunktion des Antwortverhaltens (alle Items):

$$P(X_v | \theta_v, \delta) \Rightarrow \max$$

$$P(X_v | \theta_v, \delta) = \prod_i P(X_{vi} | \theta_v, \delta_i) = \prod_i P_{vi}^{X_{vi}} Q_{vi}^{1-X_{vi}} = \prod_i \frac{\exp[X_{vi}(\theta_v - \delta_i)]}{1 + \exp(\theta_v - \delta_i)}$$

Schätzgleichung zur Bestimmung des MLE:

$$\frac{\partial \ln P(X_v | \theta_v, \delta)}{\partial \theta_v} \Rightarrow 0 \quad | \quad \frac{\partial \ln P(X_v | \theta_v, \delta)}{\partial \theta_v} = MLE = \sum_i P(X_{vi}) = \sum_i \frac{\exp(\theta_v - \delta_i)}{1 + \exp(\theta_v - \delta_i)} = 0$$

Taylor-Serie zur Bestimmung der Maxima:

$$\frac{\partial \ln P(X_v | \theta_v, \delta)}{\partial \theta_v} = \frac{\partial \ln P(X_v | \hat{\theta}_v, \delta_i)}{\partial \theta_v} + \Delta \hat{\theta}_v \frac{\partial \ln P(X_v | \hat{\theta}_v, \delta_i)}{\partial \theta_v^2} + \Delta \hat{\delta}_i \frac{\partial^2 \ln P(X_v | \bar{\theta}_v, \delta_i)}{\partial \theta_v \partial \delta_i} = 0$$

Die Schätzgleichung zur Ermittlung des MLE ist intuitiv verständlich: Der MLE ist die Lösungswahrscheinlichkeit einer Person v bei Item i summiert für jedes Item im Test. Dieses Ergebnis entspricht dem theoretisch erwarteten Summenscore und verdeutlicht erneut, dass im Rasch-Modell der Summenscore eine suffiziente Statistik zur Schätzung der Personenfähigkeit darstellt. Allerdings sind die berechneten Schätzer nur unter der Bedingung konsistent, dass die Anzahl der Items im Test gegen unendlich geht. Zudem lassen sich für Personen, die kein Item des Tests oder alle Items des Tests gelöst haben, keine Schätzer berechnen.

Um die Messgenauigkeit der Personenmesswerte zu definieren, stehen in der probabilistischen Testtheorie zwei Verfahren zur Verfügung: Es kann (1) eine Testinformationsfunktion $I(\theta_v)$ oder (2) die Varianz des Schätzers $\text{Var}(\text{MLE})$ angegeben werden.

Allgemeine Testinformationsfunktion:

$$I(\theta_v) = -E \left(\frac{\partial^2 \ln P(X_v | \theta_v, \delta)}{\partial \theta_v^2} \right) = -E \left(- \sum_i^K \frac{\exp(\theta_v - \delta_i)}{(1 + \exp(\theta_v - \delta_i))^2} \right)$$

$$I(\theta_v) = -E \left(- \sum_i^K P(X_{vi} = 1 | \theta_v, \delta_i) P(X_{vi} = 0 | \theta_v, \delta_i) \right)$$

Testinformationsfunktion für eine Stichprobe:

$$I(\theta_v) = \sum_i^K P(X_{vi} = 1 | \theta_v, \delta_i) P(X_{vi} = 0 | \theta_v, \delta_i)$$

Schätzfehlervarianz des MLE:

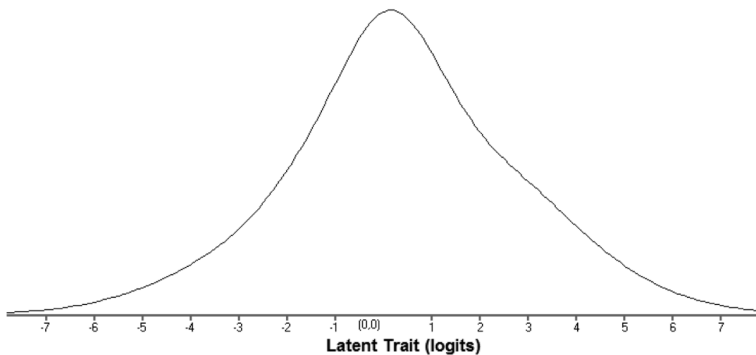
$$\lim_{K \rightarrow \infty} \text{Var}(\text{MLE}) = \frac{1}{I(\theta_v)} = \frac{1}{\sum_i^K P(X_{vi} = 1 | \theta_v, \delta_i) P(X_{vi} = 0 | \theta_v, \delta_i)}$$

Für den MLE konnte gezeigt werden, dass bei zunehmender Itemanzahl die Varianz des MLE dem Kehrwert der Testinformationsfunktion entspricht. Die Testinformationsfunktion drückt aus, welche statistische Information in den Daten hinsichtlich der Schätzung des Personenparameters enthalten ist. Formal entspricht die (allgemeine) Testinformationsfunktion dem negativen Erwartungswert der zweiten partiellen Ableitung der Log-Likelihood nach der Personenfähigkeit θ_v . Für eine Stichprobe gilt, dass sich die Testinformationsfunktion aus dem Produkt der Lösungswahrscheinlichkeit und der Gegenwahrscheinlichkeit summiert über alle Items des Tests ergibt (vgl. Rost, 2004, S. 357f.). Mit zuneh-

mender Itemanzahl strebt die Verteilung des MLE gegen eine Normalverteilung mit dem Mittelwert 0 und der Varianz $1/I(\theta_v)$.

Die Fehlervarianz eines Personenparameters veranschaulicht, dass die Messgenauigkeit der Personenschätzer abhängig von den Lösungswahrscheinlichkeiten der Items ist. Stimmen die Personenparameter mit dem Mittelwert der Itemparameter überein, ist die statistische Information in den Daten besonders groß, und die Parameter der Person können viel präziser geschätzt werden als Personenparameter, die deutlich von der mittleren Itemschwierigkeit eines Tests abweichen. Für die Fehlervarianz bedeutet das, dass ihr Wert umso kleiner wird, je stärker der Personenparameter dem Mittelwert der Itemparameter eines Tests entspricht.

Abbildung 4.13 *Testinformationsfunktion zur der Beurteilung der Messgenauigkeit der Personenmesswerte*



In Abbildung 4.13 ist auf Basis des Tests zur Messung kaufmännischer Kompetenz im Rahmen des BMBF-Projekts „Computergestützte Messung berufsfachlicher Kompetenz“ die Testinformationsfunktion für eine der Kompetenzdimensionen gegeben. Zur Berechnung der Testinformationsfunktion wurden 34 Items berücksichtigt, deren Itemparameter im Mittel 0,791 Logits betragen. Die Abbildung zeigt, dass die Testinformationsfunktion ihr Maximum entsprechend der Normalverteilungsannahme in der Nähe des Nullpunktes hat und mit zunehmender Abweichung vom Mittelwert der Informationsgehalt sinkt bzw. die Varianz zunimmt.

Gewichteter (weighted) Likelihood-Schätzer (WLE) und Expected a posteriori (EAP)

Da bei der Schätzung des MLE für Personen mit Extremscores keine endlichen Schätzer ermittelt werden können, werden zum einen absolute Wahrscheinlich-

keitsgrenzen für den MLE festgelegt und damit Werte für Personen mit Extremscores in Abhängigkeit von anderen in einem Test vertretenen Scores abgeleitet – dieses Vorgehen ist zwar begründet, aber in seinen Grundzügen willkürlich (vgl. MLE-Methode in *ConQuest*; Wu, Adams, Wilson & Haldane, 2007). Zum anderen wurde in der Vergangenheit nach Schätzern gesucht, die ähnliche positive Eigenschaften wie der MLE aufweisen, jedoch für alle auftretenden Scores Schätzer bereitstellen und darüber hinaus auch einen geringeren Bias als der MLE aufweisen.

Bei einem dieser neueren Schätzverfahren, der gewichteten (weighted) Maximum Likelihood-Methode, wird die Wahrscheinlichkeit der Personenparameter unter der Bedingung der Daten und der Itemparameter maximiert (Bayes-Ansatz der Parameterschätzung). Rost (2004) betont, dass dieses Vorgehen in seiner Logik davon ausgeht, dass bei der Schätzung diejenigen Personenparameter ermittelt werden, die bei den gegebenen Daten am wahrscheinlichsten sind. Hierfür sind im Vorfeld der Analyse Annahmen über die Verteilung der zu schätzenden Personenparameter zu treffen. Die Verteilungsannahme wird über die Dichtefunktion $f(\theta_v)$ mathematisch beschrieben. Die Bayes-Wahrscheinlichkeit ist damit proportional zu der Likelihoodfunktion des Personenparameters und dessen a priori-Verteilung (Rost, 2004, S. 313). Für die Schätzung des WLE wurden folgende Bedingungen formuliert: (1) Der Schätzer sollte möglichst durch die Testdaten beeinflusst sein und (2) auch endliche Werte für Summenscores von Null und k aufweisen. Nach Warm (1989, zit. n. Rost, 2004) erfüllt die Wurzel der Testinformationsfunktion $I(\theta_v)$ diese Bedingung, da sie (1) als a priori-Verteilung im Gegensatz zur Likelihood nur einen geringen Beitrag zu Berechnung der Bayes-Wahrscheinlichkeit leistet und (2) sehr kleine und sehr große Werte für θ_v in dieser Funktion sehr unwahrscheinlich werden (Wang, 2001). Wie in den Schätzverfahren zuvor beschrieben, wird die a posteriori-Verteilung logarithmiert und partiell nach dem Personenparameter differenziert. Die resultierenden Gleichungen werden anschließend zur Bestimmung der Maxima gleich Null gesetzt, um iterativ den Wert für den Personenparameter zu ermitteln.

Bayes-Wahrscheinlichkeit (a posteriori-Verteilung):

$$P(\theta_v | X_v, \delta) \propto P(X_v | \theta_v, \delta) f(\theta_v)$$

Bayes-Wahrscheinlichkeit nach Warm (1989):

$$P(\theta_v | X_v, \delta) \propto P(X_v | \theta_v, \delta) \sqrt{I(\theta_v)} \propto \prod_i^K P(X_{vi} | \theta_v, \delta_i) \sqrt{\sum_i^K P(X_{vi} = 1 | \theta_v, \delta_i) P(X_{vi} = 0 | \theta_v, \delta_i)}$$

Zur Ermittlung des Expected a posteriori-Schätzers (EAP) wird hingegen eine a priori-Verteilung der Personenparameter gewählt, die einen großen Beitrag zur

Schätzung der Bayes-Wahrscheinlichkeit leistet. Während also bei der gewichteten ML-Methode die Likelihood die Schätzung dominiert, basiert der Bayes-Schätzer Expected a posteriori im Gegensatz dazu auf a priori-Verteilungen, die die Likelihood der Personenparameter viel stärker beeinflussen. Die Bayes-Wahrscheinlichkeit ist dabei abhängig von den Daten einer Person X_v , den Itemparametern δ und den Parametern, die die Verteilung der Personenparameter $f(\theta_v)$ bestimmen.

Bayes-Wahrscheinlichkeit (a posteriori-Verteilung):

$$P(\theta_v | X_v, \delta, f(\theta_v)) = \frac{P(X_v | \theta_v, \delta) f(\theta_v)}{\int P(X_v | \theta_v, \delta) f(\theta_v) d\theta}$$

Expected a posteriori (EAP):

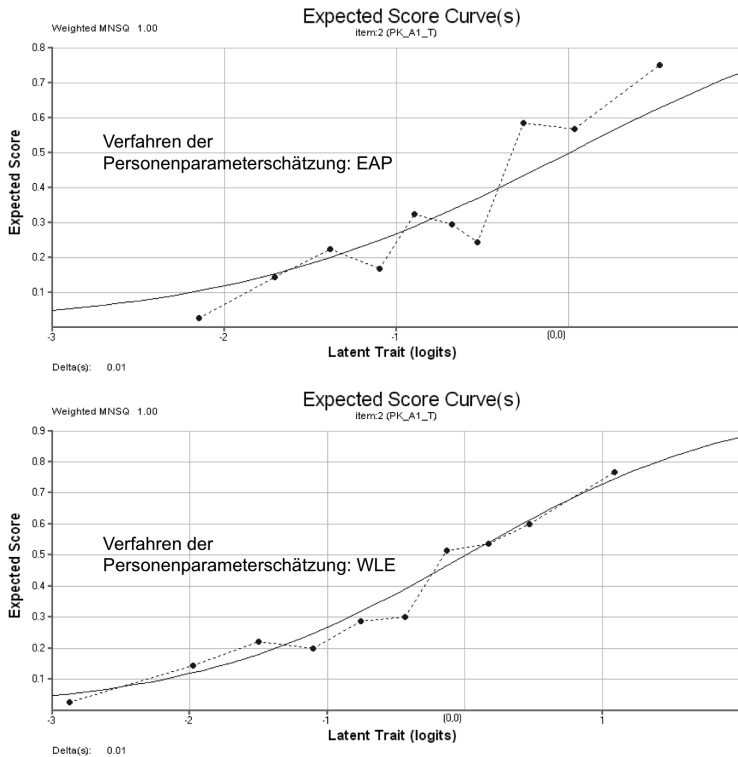
$$EAP = E(\theta_v | X_v, \delta, f(\theta_v)) = \frac{\int P(X_v | \theta_v, \delta) f(\theta_v) \theta_v d\theta}{\int P(X_v | \theta_v, \delta) f(\theta_v) d\theta}$$

Wie im Rahmen der marginalen Maximum Likelihood-Methode zur Schätzung der Itemparameter gezeigt, gibt die marginale Likelihoodfunktion des Antwortverhaltens einer Person v über alle Items i die Wahrscheinlichkeit des Auftretens eines Antwortverhaltens an, indem über die Verteilung der traits integriert wird. Zur Ermittlung der Bayes-Wahrscheinlichkeit wird daher die a posteriori-Verteilung des Personenparameters unter der Bedingung der Itemparameter und der Testdaten ins Verhältnis zur marginalen Likelihoodfunktion der Antwortpatterns gesetzt. Der EAP wird letztlich als Erwartungswert der Parameter der Person unter der Bedingung der Daten, der Itemparameter und der Verteilung der Personenparameter beschrieben. In Simulationsstudien und Large Scale-Analysen werden die Personenschätzer im Hinblick auf ihre Aussagekraft und Güte verglichen. Wang (2001) zeigt beispielsweise, dass bei einer festgelegten Testlänge im Rahmen Computer-adaptiven Testens der WLE der genaueste Punktschätzer im Vergleich der drei vorgestellten Parameterschätzer ist. Mit dem WLE wird folglich die Fähigkeitsausprägung für eine einzelne Person am besten getroffen (vgl. auch Rost, 2004; Wu, 2005).

In Abbildung 4.14 ist die Schätzung der Personenparameter am Beispiel eines Items aus dem Vorwissenstest zur Messung kaufmännischer Kompetenz des DFG-Projekts „Integrierte Kompetenzentwicklung in den beruflichen Fächern des Wirtschaftsgymnasiums“ sowohl für den EAP als auch für den WLE dargestellt. Das Item hat einen gewichteten MNSQ von 1,00 und der Schätzer der Itemschwierigkeit beträgt 0,01. Anhand dieses Beispiels ist sehr klar zu erkennen, dass der WLE die Personenfähigkeit präziser misst als der EAP. Die Funktio-

nen der erwarteten und beobachteten Scores bezogen auf ein Item i weisen beim EAP größere Abweichungen auf als beim WLE. EAP und WLE unterscheiden sich folglich vor allem in ihrem Bias. Der EAP weist i. d. R. einen größeren Bias auf, der sich auf die Verzerrung durch die a priori Verteilung zurückführen lässt (Shrinkage-Effekt). Für den EAP gilt, dass der Bias umso größer ist, je stärker der Personenparameter vom (a priori) Erwartungswert abweicht. Da in die Schätzgleichung des WLE eine a priori-Funktion einfließt, deren Informationsgehalt geringer ist, reduziert sich der Einfluss der a priori-Verteilung auf den Schätzer. Allerdings zeigt sich der EAP im Hinblick auf die Testeffizienz dem WLE überlegen (Wang, 2001).

Abbildung 4.14 IC-Funktionen zur Beurteilung des Itemfits



Wird die Personenparameterschätzung über Monte Carlo-Verfahren simuliert, zeigt sich, dass WLE und EAP dann erwartungstreu sind, wenn Personenfähigkeit

und Itemschwierigkeit relativ nah beieinander liegen. Unter dieser Bedingung ist auch der Standardfehler der Schätzer relativ klein (Kim & Nicewander, 1992).

Für die Testpraxis hat der WLE den MLE als Standardschätzer abgelöst. Der EAP kommt aktuell im Rahmen Computer-adaptiven Testens verstärkt zum Einsatz: Um zu entscheiden welches Item eine Person beim adaptiven Testen als nächstes bearbeiten soll, wird nach dem Prinzip der Minimierung des Standardfehlers das Item entsprechend der geschätzten Personenfähigkeit auf Basis des EAP ausgewählt. Über den EAP werden folglich die Items selektiert, die die Standardabweichung des Personenparameters reduzieren (Embretson & Reise, 2001).

Im Folgenden werden zwei Studien zur Messung kaufmännischer Kompetenz auf Basis der Modelle und Schätzverfahren der Item Response-Theorie beschrieben, so dass die bislang zum Teil abstrakten und sehr theoretischen Beschreibungen der IRT direkt auf reale Daten bezogen werden können. Mit diesem Vorgehen sollen die Anwendbarkeit einzelner IRT-Modelle sowie die Interpretationen des Outputs an Beispielen illustriert werden, die (1) den Nutzen der IRT für wirtschaftspädagogische Fragestellungen betonen und (2) Möglichkeiten aufzeigen, wie die Komplexität beruflichen Handelns in psychometrischen Messmodelle abgebildet werden kann.

5 Interpreting Measures: Studien zur Messung kaufmännischer Kompetenz

Die nachfolgenden Beschreibungen zweier Studien zur Messung kaufmännischer Kompetenz orientieren sich an dem in Kapitel 2.4.2 vorgestellten *Kompetenzstrukturmodell* für die kaufmännische Bildung und umfassen im Hinblick auf die Inhaltsauswahl wesentliche Elemente des in Kapitel 3.2.3 entwickelten *Domänenmodells*. Die erste Studie beinhaltet einen Test, der ausgerichtet an dem Curriculum der niedersächsischen Fachgymnasien Wirtschaft, konzeptuale und prozedurale Kompetenzen der Lernenden über die Skalen „economic literacy“ und „economic numeracy“ als domänenverbundene Vorwissensleistung erfasst. Ziel der Erhebung, die im Rahmen des DFG-Projekts „Integrierte Kompetenzentwicklung in den berufsbezogenen Fächern des Wirtschaftsgymnasiums“ durchgeführt wurde, war es, berufsfachliche Eingangsvoraussetzungen auf kognitive Leistungsfähigkeiten, motivationale und volitionale Kapazitäten sowie auf den Einsatz metakognitiver Strategien in inhaltlich unterschiedlich gestalteten Lehr-Lernprozessen zu beziehen (Winther, 2006; Achtenhagen & Winther, 2007). Im Rahmen des DFG-Projekts „Systemisches Verstehen von Geschäftsprozessen als kaufmännische Basiskompetenz“ wurde der Test zur Erfassung domänenverbundener Vorwissensleistungen auf Basis des 1-parametrischen logistischen Modells re-analysiert.

Die zweite Studie wurde im Auftrag des BMBF durchgeführt. Sie hatte zum Ziel zu prüfen, ob sich authentische, an der Arbeitsrealität von Auszubildenden zum Industriekaufmann/zur Industriekauffrau orientierte Computersimulationen als Testform für ein international vergleichendes Large-Scale Assessment in der beruflichen Bildung eignen. Das Projekt „Konstruktvalidität von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz“ war dabei vorrangig auf die Entwicklung domänenspezifischer Geschäftsvorfälle sowie auf deren fachdidaktische und technische Umsetzung in innovative Testformate ausgerichtet (Winther & Achtenhagen 2009; Achtenhagen & Winther, 2009).

Das Kompetenzmodell sowie das Domänenmodell für die kaufmännische Bildung repräsentieren, theoretisch in hohem Maße abstrakt, idealtypische und sehr komplexe Struktur- und Inhaltsannahmen. Doch gerade dieses zum Teil als „über“komplex zu beurteilende Design, gibt den Modellen die Flexibilität, die notwendig ist, um ein Modell auf die zentralen Fragestellungen einer spezifischen Untersuchung anpassen zu können. Die zwei dargestellten Untersuchungen sind als exemplarisch für die Adaptivität der (heuristischen) Modellierung zu sehen. Die Untersuchungen unterscheiden sich zentral in folgenden Punkten:

- Die Erhebung berufsfachlicher Vorwissensleistung fokussiert ausschließlich auf domänenverbundene Aspekte kaufmännischer Kompetenz. Hierzu werden die Kompetenzskalen „economic literacy“ und „economic numeracy“ analysiert. Im Gegensatz dazu wird bei der computerbasierten Messung domänenspezifische Kompetenz über ausgewählte Geschäftsvorfälle erfasst. Im Hinblick auf das Kompetenzstrukturmodell bedeutet dies, dass mit den vorliegenden Studien zunächst keine Aussagen über das Zusam-

menwirken domänenverbundener und domänenspezifischer Kompetenz getroffen werden können.

- Die Erhebung berufsfachlicher Eingangsleistung bezieht sich auf einen vollzeitschulischen Ausbildungsgang, während die computerbasierte Messung in einem Ausbildungsgang des Dualen Systems erfolgte. Entsprechend der dem Domänenmodell zugrunde liegenden Annahmen hat dieser Unterschied Einfluss auf die Wahl des externen Validitätskriterium: Für vollzeitschulische Bildungsprogramme werden die Curricula und für Ausbildungsprogramme im Rahmen des Dualen Systems neben den Curricula zusätzlich die betrieblichen Ausbildungs- und Arbeitsplatzanforderungen gewählt.

5.1 Studie I: Kaufmännische Kompetenz in vollzeitschulischen Ausbildungsgängen

Im Rahmen des DFG-Projekts „Integrierte Kompetenzentwicklung in den beruflichen Fächern des Wirtschaftsgymnasiums“ wurde zur Erfassung der Vorwissensleistung ein Test konstruiert, der sich (1) an wirtschaftlichen Alltagsbezügen und (2) an den Rahmenrichtlinien der niedersächsischen Fachgymnasien Wirtschaft orientiert. Dieser Test diene insbesondere dazu, kaufmännische Vorwissensleistung der Lernenden als möglichen Prädiktor für unterschiedliche Lernverläufe und motivationale Entwicklung in Lernprozessen zu erfassen. Der Test wurde im Rahmen des DFG-Projekts „Systemisches Verstehen von Geschäftsprozessen als kaufmännische Basiskompetenz“ auf Basis der IRT re-analysiert. Hierbei standen insbesondere folgende Fragestellungen im Zentrum: (1) Wie sind die konstruierten Items skalierbar und welche Fähigkeitsstrukturen lassen sich durch die Items des Tests messen? – Hierfür konnte auf Ergebnisse der Testanalyse auf Basis der klassischen Testtheorie zurückgegriffen werden. Diese zeigen, dass der Test aus zwei trennscharfen Skalen – economic literacy und economic numeracy – besteht, mit denen domänenverbundene Fähigkeitsstrukturen im Sinne von „konzeptueller Kompetenz“ und „prozeduraler Kompetenz“ erfasst werden. (2) Wie valide ist der Test im Hinblick auf die Heterogenität der Lernenden/Auszubildenden im Bereich der Berufsbildung? – Die grundsätzliche Annahme ist, dass die Konstruktion von Tests für berufliche Ausbildungsgänge die unterschiedlichen Bildungsbiographien und beruflichen Vorerfahrungen der Lernenden/Auszubildenden angemessen berücksichtigen muss, um zu validen Messungen zu gelangen. Mögliche Konzepte sind die Literacy/Numeracy-Annahmen der internationalen Vergleichsstudien sowie Kriterien der curricularen Validität, um die Domäne, in der gemessen werden soll, valide zu beschreiben.

Die Entwicklung des Tests zur Erfassung kaufmännischer Kompetenz in vollzeitschulischen Ausbildungsgängen ist an der entwicklungspsychologisch ausgerichteten Kompetenzkonzeption von Greeno und anderen (Greeno, Riley & Gelman, 1984; Gelman & Greeno, 1989) orientiert, die im Unterschied zum Kompetenz-Performanz-Modell Chomskys (1980) nicht auf eine Gegenüberstellung und die damit verbundene Unterscheidung von Leistung und Kompetenz abstellt, sondern zur Beschreibung kompetenter Handlungen primär zwischen domänen-

spezifischen und domänenverbundenen Variablen differenziert (Gelman & Greeno, 1989, p. 172). Greeno und Kollegen verwenden den Kompetenzbegriff im Sinne kognitiver Leistungserbringung auf Basis der Konstruktion domänenspezifischer Schemata. Demnach verfügen Lernende umso mehr über ein Verständnis der Domäne, je stärker sie ihr Wissen durch domänenspezifische Prinzipien/Termini charakterisieren können (Gelman & Greeno, 1989, p. 130; vgl. im Detail Kapitel 2.2.2). Die Frage der Wissensorganisation ist eng verbunden mit der in der Kognitionswissenschaft noch offenen Frage der Vorhersage von Expertise in relevanten Anforderungssituationen. Als plausibel gilt in der aktuellen Literatur ein Schwellenwertmodell, demzufolge der Aufbau von Expertise von kognitiv-dispositionalen Voraussetzungen abhängig ist (Gruber & Mandl, 1996; Schneider, 1997; Gruber, 1999). Kompetenz ist demnach weniger auf übergreifende kognitive Fähigkeiten und Dispositionen zurückführbar als vielmehr auf mentale Netzwerke des Domänenwissens und auf domänenbezogene Techniken. Die Konstruktion des Vorwissenstests orientiert sich vor diesem Hintergrund insbesondere darauf, solches Domänenwissen zu erfassen, von dem angenommen werden kann, dass es (1) die Lernverläufe in den berufsbezogenen Fächern des Fachgymnasiums Wirtschaft beeinflusst und mit dem (2) die unterschiedlichen Zugangswege in den Ausbildungsgang am Fachgymnasium ausgeglichen werden können. Hierbei bestand Konsens darüber, dass der Test – im Verständnis von Greeno und Kollegen – auf domänenverbundenen Inhalten basieren sollte, da hiermit (1) der theoretischen Annahme gefolgt werden kann, dass domänenverbundene Kompetenz die Bewältigung von einzelnen Anforderungen in der Domäne unterstützt (Gelman, Greeno & Riley, 1989, p. 172) und sich folglich mögliche Effekte auf nachfolgende domänenspezifische Lernprozesse nachweisen ließen. Darüber hinaus dürfte eine Fokussierung auf domänenverbundene Testinhalte (2) bei einer möglicherweise vorhandenen kaufmännischen Vorbildung weniger zu einer systematischen Verzerrung der Ergebnisse führen, als dies bei domänenspezifischen Aufgaben zu erwarten wäre.

5.1.1 Kaufmännisches Vorwissen als Kategorie des Domänenwissens

Die Entwicklung des Vorwissenstests basierte auf kognitionspsychologischen Vorstellungen des Domänenwissens sowie auf fachinhaltlichen Analysen der Curricula und Lehrbücher des Ausbildungsganges am Fachgymnasium Wirtschaft. Zentral für die Entwicklung war die Frage, welche Rolle Domänenwissen in Lernprozessen spielt. Unter Rückgriff auf Alexander (1992, p. 34) ist Domänenwissen als spezifischer, auf ein abgrenzbares Inhaltsgebiet bezogener Wissensbereich eines Individuums definiert und umfasst als solcher deklarative, prozedurale und konditionale Wissensbestände und Fähigkeiten sowohl auf verdeckter („tacit knowledge“) als auch auf expliziter Ebene („explicit knowledge“). Unter einer systemischen Perspektive kann Domänenwissen als fachbezogenes Beispiel für Vorwissen verstanden werden:

„[...] Domain knowledge is a specialized instance of an individual's prior knowledge. It is that segment of an individual's existing conceptual knowledge that is related to a specific "studied" area“ (Alexander, 1992, p. 35).

Glaser, Lesgold und Lajoie (1987) weisen darauf hin, dass sich Domänen erheblich voneinander unterscheiden und folglich auch das Domänenwissen verschiedene Strukturen und Eigenschaften aufweist. Der Erwerb von Domänenwissen kann dadurch geprägt sein, wie stark eine Domäne ausdifferenziert ist, in welchem Maße Domäneninhalte regelgebunden erlernt werden und sich an akademische Disziplinen orientieren.

Der Unterschied zwischen Domänen und Domänenwissen wurde insbesondere im Rahmen von Studien der Experten-Novizen-Forschung deutlich (vgl. u. a. Chi, 1985; Chi, Glaser & Farr, 1988; Spada & Wichmann, 1996). Hierbei ist folgendes Ergebnis von besonderem Interesse: Wenn kognitive Leistungen in wenig strukturierten Domänen oder unzureichend definierten Anforderungssituationen erbracht werden müssen, zeigen sich die Ergebnisse der Experten-Novizen-Forschung weit weniger stabil als für gewöhnlich angenommen (vgl. Pate, Alexander & Kulikowich, 1989, zit. n. Alexander, 1992). Studien der Experten-Novizen-Forschung bestätigen für sehr gut ausdifferenzierte Domänen und für klar definierte domänenbezogene Aufgaben, dass Experten über eine umfangreichere Wissensbasis und einen höheren Anteil richtiger Verknüpfungen von Wissen verfügen. Für den Bereich des prozeduralen Wissens unterscheiden sich Experten von Novizen in der Art der Problemrepräsentation. Experten greifen beim Abbilden der Problemsituation auf weniger Oberbegriffe und abstrakte Merkmale zurück als Novizen, da sie über effiziente, bereichsspezifische Problemlöseschemata verfügen, die den Suchaufwand reduzieren und eine sichere Lösung ohne kognitive Überlastung erlauben. Experten können ihre Lösungsstrategien modifizieren und an den jeweiligen Problembereich anpassen. Statt einer Suche können wissensbasierte Entscheidungen im Problemlöseprozess getroffen werden (vgl. Spada & Wichmann, 1996). Jedoch sind diese Unterschiede zwischen Experten und Novizen dann nicht mehr so eindeutig, wenn die zu bearbeitenden Aufgaben keine eindeutige Lösung verlangen oder wenn die zu zeigende Leistung Elemente eines spezifischen, gesellschaftlich geteilten Allgemeinwissens enthält (Hofstadter, 1985; Voss & Post, 1988; Alexander, Schallert & Hare, 1991).

Interessanterweise existieren zu diesen Befunden Beispiele aus dem kaufmännischen Bereich. Voss und Kollegen zeigen beispielsweise, wie an Alltagskontexte gebundene kaufmännische Aufgaben von Personen gelöst werden, die formale Ausbildungsprogramme im Ökonomiebereich absolvierten und von jenen, die keine gezielte Ausbildung erhielten. In beiden Gruppen sind Personen enthalten, die das College besucht haben und Personen, die über kaufmännische berufliche Vorerfahrungen verfügen. Es waren Aufgaben aus unterschiedlichen Bereichen der Ökonomie (Preisänderungen, Haushaltsdefizit, Zinsänderungen) zu be-

arbeiten, die ein offenes Antwortformat aufwiesen. Folgendes Beispiel sei gegeben (Voss, Blais, Means, Greene & Ahwesh, 1986, p. 299):

Assume that because of OPEC, the price of gasoline is expected to increase sharply in the next year. What effect, if any, do you think this would have on the price of automobiles? Why?

Das Ergebnis der Studie ist, dass während eine allgemeine Ausbildung am College die Lösungen positiv beeinflusst, weder kaufmännisch berufliche Vorerfahrungen noch eine formale Ausbildung im ökonomischen Bereich die Lösungsqualitäten erhöhen. Die Autoren schlussfolgern, dass Instruktion in der kaufmännischen Domäne nicht notwendigerweise dazu führt, dass kaufmännische Anforderungssituationen, die auf Alltagsbezüge im Sinne des Literacy-Konzepts abstellen, besser bearbeitet werden können. Sie stellen insbesondere allgemeine, auf den Kontext der Domäne bezogene Fähigkeiten heraus, die eine Bearbeitung der Aufgaben und weitere Lernprozesse in der Domäne unterstützen: (1) die Qualität der Problemrepräsentation, (2) der Gebrauch von Kriterienkatalogen, Gegenargumenten und Metastatements sowie (3) das Aufdecken von Unklarheiten und Widersprüchen (Voss et al., 1986, p. 269). Für die Entwicklung von kaufmännischen Vorwissenstests lassen sich auf Basis dieser Befunde zwei Überlegungen ableiten: Einerseits ist ein Test, der auf das Literacy-Konzept abstellt, geeignet, funktionale Aspekte der Bildung in einer spezifischen Domäne zu erfassen und damit die Heterogenität der Lernenden in der beruflichen Bildung aufzufangen. Dies gelingt, da anstelle domänenspezifischer Aufgabenstellungen, domänenverbundene Anforderungssituationen verarbeitet werden, die text-, bild- und/oder zahlssprachliche Informationen in Alltagsbezüge einbetten. Diese Informationen sind im Sinne eines fachlich argumentativen Wissens in Alltagsbezügen zu nutzen, um eigene Ziele erreichen und das eigene Wissen und Handlungspotential ausbauen zu können (vgl. OECD, 1995; Deutsches PISA-Konsortium, 2001). Andererseits wird mit dem Ausbildungsgang am Fachgymnasium Wirtschaft eine formale Qualifikation angestrebt, die u. a. auf ein kompetentes Handeln in der kaufmännischen Domäne abzielt. Dies wird durch domänenspezifische Lehr-Lernprozesse in den berufsbezogenen Fächern erreicht. Um domänenverbundene Vorwissenleistung mit domänenspezifischer Performanz in den Lernprozessen des Bildungsganges zu verknüpfen, ist curriculare Validität zu sichern.

5.1.2 Literacy-Konzeption und curriculare Validität des kaufmännischen Vorwissenstests

In den PISA-Studien wird das Konzept der „Literacy“ genutzt, um deutlich zu machen, dass Kenntnisse und Fähigkeiten, wie sie traditionell in den Curricula der allgemein bildenden Schulfächer definiert werden, im Rahmen der internationalen Vergleichsstudie nicht im Vordergrund stehen. Stattdessen wird über den Be-

griff der Literacy/Grundbildung die funktionale Anwendung von Kenntnissen in unterschiedlichen Kontexten eingeführt (Deutsches PISA-Konsortium, 2001).

In der vorliegenden Studie zur Messung kaufmännischer Kompetenz in vollzeitschulischen Ausbildungsgängen wird Literacy als Konzept gewählt, das auf die kulturelle Teilhabe und Entwicklungszustände in wirtschaftsbezogenen Kontexten abstellt. In Anlehnung an verschiedene Studien wie den International Adult Literacy Survey (IALS; OECD, 1995) sowie an die vom National Council on Economic Education herausgegebenen Standards und den Test ökonomischer Grundbildung (als deutsche Adaption vgl. Beck, 2001) werden text- und bildsprachliche Kenntnisse sowie das Verständnis von quantitativen Werten und Verhältnissen in wirtschaftlichen Alltagskontexten erfasst. Hierbei erfolgt die Entwicklung der Testitems in Abstimmung mit den niedersächsischen Rahmenrichtlinien des Fachgymnasiums Wirtschaft. Ein zentrales Ergebnis der Curriculum- und Domänenanalyse ist, dass der Zugang zu dem berufsbezogenen Fach „Betriebswirtschaft mit Rechnungswesen/Controlling“ des Fachgymnasiums Wirtschaft zum einen sprachlich-argumentativ und zum anderen mathematisch-analytisch erfolgt (vgl. Kapitel 3.2.3). Diese verschiedenen Zugangswege haben zentralen Einfluss auf die Methoden der Instruktion im Fach und folglich auf die Art der Wissensrepräsentation. Um die verschiedenen Zugänge zu der kaufmännischen Domäne darzustellen, werden bei der Re-Analyse des Vorwissenstests über die Ansätze der „economic literacy“ und „economic numeracy“ zwei Bereiche abgegrenzt, mit denen konzeptuales Wissen in sprachlich und prozedurales Wissen in mathematisch orientierten Anforderungssituationen erfasst werden kann.

5.1.2.1 Economic literacy und economic numeracy

Die nachfolgenden konzeptionellen Ausführungen zum Literacy/Numeracy-Konzept basieren vor allem auf den Arbeiten des Quantitative Literacy Design Teams des National Council on Education and the Disciplines (NCED, 2001) und den Vorgaben der OECD/PISA-Studien (u. a. OECD 1995; 2005). Ausgangspunkt der Überlegungen sind Veröffentlichungen, die auf die enge Beziehung zwischen alltagsbezogener Grundbildung und schulischer und beruflicher Performanz hinweisen (Murnane & Levy, 1996; Forman & Stehen, 1999). Hierbei wird dem Konzept der Literacy eine entscheidende Bedeutung zugesprochen. Für den quantitativ/mathematischen Bereich liegen sehr elaborierte Beschreibungen vor, an denen sich andere Domänen orientiert haben. Den Konzepten der verschiedenen Domänen ist gemeinsam, dass Literacy von der Fachwissenschaft abstrahiert und auf Situationen des Alltags heruntergebrochen wird. Für die quantitative Literacy (auch numeracy oder mathematical literacy) bedeutet dies, dass zwar Aspekte der Mathematik oder Statistik enthalten sind – diese jedoch ausschließlich bezogen auf die funktionale Anwendung.

„Quantitative literacy is more a habit of mind, an approach to problems that employs and enhances both statistics and mathematics [...] quantita-

tive literacy involves mathematics acting in the world. Typical numeracy challenges involve real data and uncertain procedures but require primarily elementary mathematics“ (NCED, 2001, pp. 5/6).

Im Rahmen des National Adult Literacy Survey (NALS) wird quantitative literacy beispielsweise wie folgt definiert:

„The knowledge and skills required to apply arithmetic operations, either alone or sequentially, using numbers embedded in printed material (e.g., balancing a checkbook, completing an order form)“ (NCES, 1993, p. 23).

Der International Life Skills Survey (ILSS) wählt hingegen eine breitere Definition der quantitative literacy:

„An aggregate of skills, knowledge, beliefs, dispositions, habits of mind, communication capabilities, and problem solving skills that people need in order to engage effectively in quantitative situations arising in life and work“ (ILSS, 2000, p. 12).

In den Dokumentationen zu den PISA-Studien findet sich folgende Definition zu mathematical literacy:

„An individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individuals' life as a constructive, concerned and reflective citizen“ (OECD, 2005, p. 14).

Die Definitionen verdeutlichen am Beispiel der Mathematik, dass das Literacy-Konzept unterschiedlich ausgestaltet wird. Es stellt ebenso auf grundlegende Fähigkeiten („arithmetic operations“) wie auf übergeordnete Strategien („well-founded judgments“) ab. Als Gemeinsamkeiten sind festzustellen, dass das Literacy-Konzept stets von einer Anforderungsbewältigung in spezifischen Kontexten ausgeht und dass die gestellten Anforderungssituationen einen direkten Bezug zu Alltagserfahrungen aufweisen.

Der Test zur Erfassung von Vorwissensleistungen am Fachgymnasium Wirtschaft greift das Literacy/Numeracy-Konzept auf. Der zugrunde gelegte Kontext erfasst verschiedene kaufmännische Inhaltsbereiche. So werden beispielsweise Vorstellungen zum betriebswirtschaftlichen Gewinnkonzept, zum Anleger- und Gläubigerschutz, zu marktwirtschaftlichen Unternehmensstrategien und zu den Anspruchs- und Interessengruppen einer Unternehmung erfasst. Der ökonomische Alltagsbezug wird dadurch realisiert, dass alle Anforderungssituationen an einen Zeitungsartikel aus dem Wirtschaftsressort der Frankfurter Allgemeinen Zeitung gebunden sind (Abbildung 5.1.1). Der Zeitungsartikel stellt eine realen Kontext zur Verfügung, über den ökonomische Fragestellungen bearbeitet werden können. Mit diesem Zeitungsartikel liegt darüber hinaus eine Quelle realer

Zahlen und Fakten vor, die von den Fachgymnasiasten erfasst und interpretiert werden müssen.

Abbildung 5.1.1 Realer Kontext des Vorwissenstests

VW korrigiert Gewinnziel nach unten

Mehr als 500 Millionen Euro unter Plan / Neues Oberklassemodell auf Eis gelegt

rit. HAMBURG, 20. Juli. Der Volkswagen-Konzern wird sein Ergebnisziel auch in diesem Jahr verfehlen. Nach Informationen dieser Zeitung rechnet Europas größter Automobilhersteller für 2004 nur noch mit einem operativen Ergebnis vor Sonderinflüssen von 1,5 bis 2 Milliarden Euro. Ursprünglich wollte Volkswagen das Vorjahresergebnis von 2,5 Milliarden Euro übertreffen. Die revidierte Prognose soll am Freitag im Halbjahresbericht veröffentlicht werden, heißt es. Ein VW-Sprecher wollte diese Informationen nicht kommentieren. Den Informationen zufolge ist im ersten Halbjahr – trotz der stabilen Erträge aus der Finanzdienstleistungssparte – mit einem kräftigen Gewinneinbruch zu rechnen. Als Konsequenz aus der abermals verschlechterten Ertragslage hat der Vorstand die Pläne für den Bau des „C1“ vorerst auf Eis gelegt. Mit dem Modell, das 2007 in Serie gehen sollte, wollte VW die Lücke zwischen Passat und Phaeton schließen.

Trotz der Gewinnrevision wird VW am Freitag voraussichtlich keine Verschärfung des im Frühjahr eingeleiteten Restrukturierungskurses ankündigen. Damals wurde das Sparziel für 2004 und 2005 auf insgesamt 4 Milliarden Euro verdoppelt. Die jetzige Zurückhaltung hängt womöglich damit zusammen, daß man vor den im September startenden Tarifverhandlungen keine Drohkulisse errichten will. Gleichwohl steigt der Druck auf die Gewerkschaften, bei den Personalkosten kompromißbereiter zu sein. VW will die Personalkosten bis 2011 um 30 Prozent senken.

Schon in den ersten drei Monaten des Jahres war das operative Konzernergebnis

um 46 Prozent auf 329 Millionen Euro eingebrochen. Seither hat sich das von Konjunktur- und Konsumschwäche geprägte Marktfeld kaum verbessert. Der Verband der Automobilindustrie erwartet allenfalls noch einen stagnierenden Inlandsabsatz. Auf die europaweit zögerliche Nachfrage reagieren alle Hersteller mit zum Teil kräftigen Preisnachlässen; das hinterläßt deutliche Spuren in den Bilanzen von Ford und General Motors (mitsamt der Tochtergesellschaft Opel). Auch die französischen Hersteller Renault und PSA Peugeot Citroën sowie die italienische Fiat bekommen den Preisdruck zu spüren. Unter den Massenherstellern schwimmt Toyota indes gegen den Strom. Dank der hohen Produktivität eilt der größte japanische Autobauer

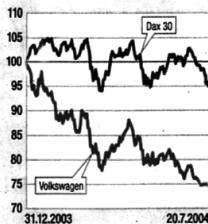
bei Absatz und Ergebnis von Rekord zu Rekord.

VW hat die Kaufanreize im Verlauf dieses Jahres stetig erhöht. Das gilt vor allem für den neuen Golf. Nach der kostenlosen Zugabe einer Klimaanlage gibt es seit Juni zusätzlich eine Händlerprämie von 928 Euro für die Inzahlungnahme eines Gebrauchtwagens. Dies stabilisiert den Absatz, verengt aber die Gewinnmargen – und zwar nicht nur beim Golf. Denn um die Differenzierung der Konzernmodelle untereinander aufrechtzuerhalten, müssen auch die Preise der übrigen Modelle wie VW Polo, Skoda Octavia oder Seat Ibiza (direkt oder indirekt) gesenkt werden.

Zudem gerät VW auf wichtigen Auslandsmärkten unter Druck. In Amerika macht den Wolfsbergen die Dollar-Schwäche zu schaffen: Bei Durchschnittskursen zwischen 1,20 und 1,25 Dollar je Euro dürfte das operative Ergebnis im Gesamtjahr um rund 1,2 Milliarden Euro belastet werden. Auch toben in Amerika, wo der VW-Absatz unter dem Modellwechsel bei Jetta und Passat leidet, noch härtere Rabattschlachten als in Europa. All dies bescherte allein im ersten Quartal in Nordamerika einen Verlust von 235 Millionen Euro. Der chinesische Markt, auf dem VW in der Vergangenheit blendend verdiente, hat stark an Dynamik verloren. Dort versuchen die Konkurrenten – allen voran General Motors – VW mit Kampfpreisen die Marktführerschaft zu entreißen. Die Wolfsberger müssen wohl mit einem geringeren Ergebnisbeitrag aus China rechnen.

Volkswagen im Vergleich

Kurs 31. Dezember 2003=Index 100



Quelle: Thomson Financial Datastream (20.7.10 Uhr)

Toyota erhöht Absatzziel, Seite 16.
Ford profitiert von seiner Finanzsparte, Seite 16.

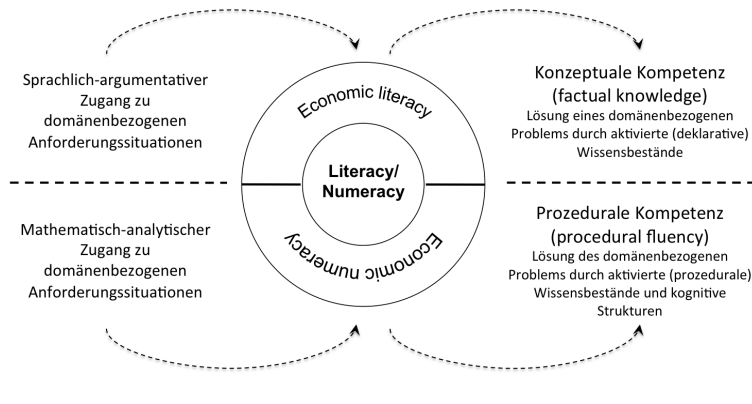
Quelle: FAZ 21.7.2004

Die entwickelten Testitems sind dabei zwei Skalen zugeordnet, die unterschiedliche domänenverbundene kognitive Fähigkeitstrukturen repräsentieren: economic literacy und economic numeracy (vgl. Abbildung 5.1.2). Die Skala der economic literacy wird über Items beschrieben, die vorrangig konzeptuale Wissensbestände erfassen und für deren Beantwortung die Lernenden deklarative Fähigkeiten einbringen müssen. Grundlage für die Beantwortung ist der Fließtext des Artikels, mit dessen Hilfe Anforderungssituationen konstruiert werden, die durch die Lernenden sprachlich-argumentativ erschlossen werden müssen.

Die Skala der economic numeracy erfasst grundlegende mathematische Kenntnisse und Fertigkeiten vor dem Hintergrund konkreter, im Artikel über Graphiken und Zahlen präsentierter Unternehmenswerte. Die Bezeichnung „numeracy“ wurde gewählt, da die Skala auf Basis von unternehmensinternen Realwerten Aufgaben vorgibt, zu deren Lösung prozedurale Fähigkeiten notwendig

sind. Items der Skala der economic numeracy beziehen sich vorrangig auf das Aufstellen und die Lösung von Rechenoperationen mit unternehmensinternen Bezügen. Damit liegen domänenverbundene Anforderungssituationen vor, deren Zugang mathematisch-analytisch erfolgt. Für die Re-Analyse des Vorwissenstests stellen diese zwei Skalen die zu überprüfenden Kompetenzstrukturen dar.⁹

Abbildung 5.1.2 Literacy-Skalen und kognitive Fähigkeitsstrukturen



Die Auswahl der Items orientierte sich dabei konsequent an den Inhalten des Lerngebiets 1: „Das Unternehmen als komplexes wirtschaftliches und soziales System“, der niedersächsischen Rahmenrichtlinien des Fachgymnasiums Wirtschaft für das Fach Betriebswirtschaft mit Rechnungswesen/Controlling, um (1) curriculare Validität zu realisieren und um (2) die Ergebnisse des Vorwissenstests inhaltlich auf nachfolgende Lernprozesse beziehen zu können.

5.1.2.2 Curriculare Validität

Zentrale Zielsetzung des Fachs Betriebswirtschaft mit Rechnungswesen/Controlling ist das Erschließen der ökonomischen Welt durch die Verwendung von weitgehend normierten Konzepten, Instrumenten und Regeln, um ein angemessenes Agieren in komplexen und dynamischen ökonomischen Anforderungssituationen zu gewährleisten. Mit dieser Zielsetzung ist inhaltlich-formal eine starke Parallele zu dem Konzept der Literacy/Numeracy gegeben. Es geht darum, ökologische, technologische, soziale und politische Zusammenhänge komplex abzubilden, in die die sich wandelnden Arbeits-, Organisations- und Produktionsprozesse im Unternehmensbereich eingebettet sind. Insgesamt sollen

⁹ Der Konstruktion des Vorwissenstests lag kein Kompetenzstrukturmodell zugrunde. Die Annahme zweier verschiedener Kompetenzstrukturen – economic literacy und economic numeracy – ist das Ergebnis der Bearbeitung des DFG-Projekts „Systemisches Verstehen von Geschäftsprozessen als kaufmännische Basiskompetenz“.

die Fähigkeit zu Aktualisierung, Generalisierung und Transfer ökonomischen Wissens sowie der Erwerb informationstechnischer und methodisch-sozialer Kompetenzen mit dem Ziel erreicht werden, zukünftige berufliche, private und gesellschaftlich-öffentliche Lebenssituationen zu gestalten. Das Fachgymnasium Wirtschaft vermittelt den Lernenden eine erweiterte und vertiefte ökonomische Bildung, die dazu befähigt, den Anforderungen von Studiengängen jeglicher Fachrichtung an einer Universität gerecht zu werden (Niedersächsisches Kultusministerium, 2001, S. 3f.; vgl. auch KMK, 2000b).

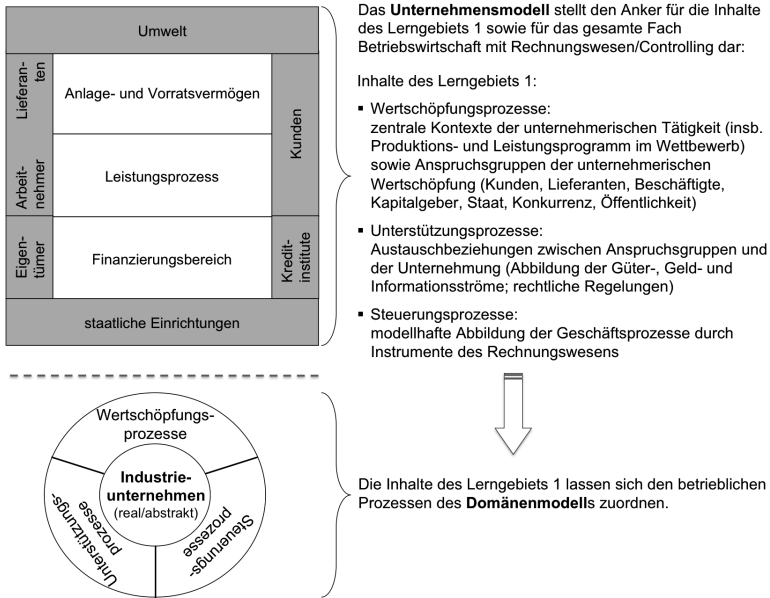
Das Lerngebiet 1: „Das Unternehmen als komplexes wirtschaftliches und soziales System“, stellt eine Einführung der in Kapitel 3.2.3 im Rahmen des Domänenmodells beschriebenen betrieblichen Unternehmensprozesse über 80 Unterrichtsstunden dar.¹⁰ Mit den Grundlagen im Hinblick auf betriebliche Wertschöpfungs-, Steuerungs- und Unterstützungsprozesse wird die Basis für ein angemessenes Verständnis der Unternehmung in wirtschaftlichen Zusammenhängen gelegt. Das Lerngebiet 1 soll den Lernenden einen Überblick vermitteln und sie befähigen, die vielfältigen Details und Fakten eines real existierenden Unternehmens sinnvoll einzuordnen und zu strukturieren. Zweck und Ziele eines Unternehmens werden vor dem Hintergrund unterschiedlicher Anspruchs- und Interessengruppen diskutiert; damit wird für die Lernenden deutlich, dass Unternehmen lebendige und gestaltbare ökonomische und soziale Systeme darstellen, die in ihren Marktbeziehungen zu analysieren sind.

Die Vermittlung der Lerninhalte erfolgt im Lerngebiet 1 mit Hilfe des Modellunternehmens Kettenfabrik A & S GmbH (Siemon, 2003). Im Modellunternehmen wird die betriebliche Realität durch Videosequenzen, durch Abbildungen der Geschäfts- und Produktionsprozesse, durch Produktlisten etc. didaktisch aufbereitet. Dabei wird auch das Zahlenwerk des externen und internen Rechnungswesens in angemessener Weise widerspiegelt (als ausgewählte Literatur vgl. Achtenhagen, 1998; 2001; 2002; 2003; Achtenhagen, Preiß & Weber, 2004; Getsch & Preiß 2003a; 2003b).

Auf Basis des Modellunternehmens werden im Lerngebiet 1 vielfältige ökonomische und soziale Eigenschaften der Unternehmen, komplexe Austauschbeziehungen zwischen vom Wertschöpfungsprozess betroffenen Anspruchsgruppen und dem Unternehmen (→ betriebliche Wertschöpfungsprozesse), rechtliche Regelungen (→ betriebliche Unterstützungsprozesse) sowie Regelungen der Finanzbuchführung und des Controllings (→ betriebliche Steuerungsprozesse) in ihren Grundlagen vermittelt. Der Unterrichtseinstieg erfolgt dabei über eine komplexe Ziel- und Inhaltsstruktur, die über ein grundlegendes Unternehmensmodell (vgl. Abbildung 5.1.3) strukturiert wird. Das Unternehmensmodell steht als Systematisierungshilfe für den gesamten Unterricht im Fach Betriebswirtschaft mit Rechnungswesen/Controlling zur Verfügung und stellt das curriculare Validierungsmodell für den Vorwissenstest dar.

¹⁰ Ein detaillierter Überblick über die Zielsetzungen und die Rahmenbedingungen des Fachs Betriebswirtschaft mit Rechnungswesen/Controlling findet sich bei Winther (2006).

Abbildung 5.1.3 Unternehmensmodell als curriculares Validierungsmodell



Entsprechend des Unternehmensmodells werden domänenverbundene Aufgaben zur Erfassung der Vorwissensleistung konstruiert, die den Inhalten des Lerngebiets 1 zugeordnet werden können und die sich auf die im Domänenmodell beschriebenen Wertschöpfungs-, Steuerungs- und Unterstützungsprozesse beziehen lassen. So werden Fragen der betrieblichen Wertschöpfung beispielsweise über das Gewinnkonzept sowie über grundlegende Annahmen zu Lieferfristen bearbeitet. Hierzu enthält der Vorwissenstest folgende Testitems:¹¹

Welche Gründe für den Gewinneinbruch bei der Volkswagen AG erkennen Sie in dem Zeitungsartikel? (Item 1)

Nennen Sie bitte weitere Gründe, die grundsätzlich den Gewinn einer Unternehmung reduzieren können. (Item 3)

Das Volkswagenwerk in Wolfsburg erhält vom Land Thüringen den Auftrag, möglichst kurzfristig 2.800 VW Passat zu liefern. Die Pkws sollen den Fuhr-

¹¹ Eine weiterführende Darstellung der Testitems des Vorwissenstests findet sich bei Winther (2006).

park der Polizei an einigen zentralen Standorten erweitern. Warum ist es für die Volkswagen AG wichtig, kurzfristig liefern zu können? (Item 16)

Diese Testitems, die der Skala der economic literacy zugeordnet sind, sind von den Lernenden sprachlich-argumentativ zu bearbeiten und sprechen insbesondere konzeptuale Fähigkeitsstrukturen an. Ein Beispielitem der Skala der economic numeracy des Bereichs der betrieblichen Wertschöpfung hat folgende Form:

Aus dem Zeitungsartikel geht hervor, dass das operative Konzernergebnis der Volkswagen AG in den ersten drei Monaten des Jahres 2004 im Vergleich zum Ergebnis der ersten drei Monate des Vorjahres um 46 Prozent auf 329 Millionen Euro eingebrochen ist. Berechnen Sie das operative Ergebnis der Volkswagen AG für die ersten drei Monate des Vorjahres. (Item 22)

Dieses mathematisch-analytische Beispielitem zeigt, dass zu einer Lösung des Items keine spezifischen betriebswirtschaftlichen Fähigkeiten notwendig sind. Es werden jedoch generelle Rechenoperationen (Dreisatz) gefordert, die von den Lernenden eingebracht werden müssen, um Unternehmenszahlen aus vorliegenden realen Daten zu ermitteln. Hierfür werden prozedurale Fähigkeitsstrukturen benötigt.

Insgesamt enthält der Vorwissenstest 27 Testitems. Die Mehrzahl der Testitems (21) ist entsprechend des Einführungscharakters des Lerngebiets 1: „Das Unternehmen als komplexes wirtschaftliches und soziales System“, der Skala der economic literacy zugeordnet; die Skala der economic numeracy wird aus sechs Items gebildet.

Die curriculare Validität des konstruierten Tests wird insgesamt als sehr gut beurteilt; 20 der 27 Testitems (74%) werden am Beispiel des Modellunternehmens Kettenfabrik A & S GmbH in den Lernprozessen des Lerngebiets 1 als Übungsaufgaben in leicht veränderter Form genutzt. Hierzu ist anzumerken, dass der Vorwissenstest unter der Prämisse konstruiert wurde, die gezeigte Vorwissensleistung als kognitive Disposition zur Beschreibung von Lernprozessen zu nutzen. Zur Erreichung dieses Ziels wurden im Verlauf des Lerngebiets 1 Daten zu Lernleistungen, motivationalen und volitionalen Ressourcen sowie zum Lernstrategieeinsatz im Längsschnitt erhoben. Insgesamt zeigte sich in einem Kausalmodell lernrelevanter Dispositionen, dass die Vorwissensleistung kein signifikanter Prädiktor der Lernleistung ist ($\beta = ,140$; $p > ,05$). Die Auswertungen des Tests basierten auf einem Ingesamtttestscore für alle Items des Tests. Zu diesem Analysezeitpunkt wurde noch keine Differenzierung zwischen den Skalen der economic literacy und economic numeracy vorgenommen. Bezogen auf die im Test erfassten unterschiedlichen Fähigkeitsstrukturen lässt sich zusammenfassen, dass die erbrachten Leistungen sehr moderat waren: So zeigt die Auswer-

tung des Items 1 beispielsweise für die konzeptuale Kompetenzstruktur, dass die Schüler die im Zeitungsartikel genannten Einflussfaktoren der unternehmerischen Gewinnkategorie kaum identifizieren konnten. Zentrale Kenngrößen für die Schüler sind das Kaufverhalten der Kunden sowie die Währungskursschwankungen auf den Auslandsmärkten. Auch die Leistungen im prozeduralen Kompetenzbereich sind schwächer ausgefallen als erwartet. Im Hinblick auf das Item 22 konnten nur 30 Prozent der Lernenden das operative Konzernergebnis ermitteln (im Detail vgl. Winther, 2006).

Die nachfolgend dargestellten Ergebnisse des Vorwissenstests verfolgen andere Auswertungs- und Analyseschwerpunkte. Im Mittelpunkt der Auswertung stehen die Kalibrierung der Testitems sowie die Prüfung eines zweidimensionalen Kompetenzstrukturmodells, das zwischen konzeptualler und prozeduraler domänenverbundener Kompetenz differenziert, im Vergleich zu einem Modell, das von einer eindimensionalen domänenverbundenen Kompetenzstruktur ausgeht. Die Ergebnisanalyse ist der Beschreibung der Anforderungssituationen gewidmet und hier insbesondere der Frage, ob sich allgemeine Konstruktionskriterien für den Einsatz eines Tests in einer für die berufliche Bildung typischen heterogenen Lernerschaft ableiten lassen.

5.1.3 Re-Analyse des kaufmännischen Vorwissenstests auf Basis der IRT

Für die Re-Analyse des kaufmännischen Vorwissenstests auf Basis der Item Response-Theorie wurden die erhobenen Leistungsdaten der einzelnen Testitems neu kodiert. Die Rohwerte der Items wurden dabei in dichotome Antwortmuster überführt, um die Daten auf Basis des 1-parametrischen logistischen Modells zu prüfen. Für die Wahl des Rasch-Modells sprechen dabei folgende Punkte:

- Ein zentrales Ziel der Re-Analyse bezieht sich auf die Kalibrierung der Items. Da beim Rasch-Modell für jedes Testitem jeweils nur ein Itemschwierigkeitsparameter geschätzt werden muss, lässt sich die Verteilung der Items auf der Logit-Skala intuitiv beschreiben.
- Die Vorteile des Rasch-Modells im Hinblick auf die relativ geringe Anzahl der zu schätzenden Modellparameter erweist sich darüber hinaus auch für die Modellprüfung als vorteilhaft. Hierbei ist zu beachten, dass der Vorwissenstest weder auf Basis einer Construct Map als Mittel der Konstruktrepräsentation noch anhand von schwierigkeitsklassifizierenden Kriterien entwickelt wurde. Es liegen folglich keine theoretischen Annahmen über die Itemschwierigkeiten vor. In Folge dessen sind die Antwortmuster der Probanden zu analysieren, bevor ihre Testwerte dazu genutzt werden können, die Ausprägung der Fähigkeit zu schätzen – ein Vorgehen, das auf Basis dichotomer Daten vereinfacht ist.

Nachfolgend werden zentrale Kennwerte und Ergebnisse der Testanalyse auf Basis der klassischen Testtheorie präsentiert, bevor auf die Re-Analyse des Tests mittels IRT eingegangen wird. Der Vorwissenstest wird dabei als Einpunktmesung behandelt. Die Stichprobe ist wie folgt charakterisiert: Für den Vorwissenstest liegen von insgesamt 340 Probanden Daten vor. Hiervon sind 162 weiblich

und 178 männlich. Das Alter der Probanden liegt zwischen 17 und 30 Jahren, wobei 85 Prozent der Lernenden das für das Wirtschaftsgymnasium erwartete Alter von 17 bis 19 Jahren aufweisen. 32 Prozent der an der Studie teilnehmenden Lernenden sind vom allgemein bildenden Gymnasium auf das Fachgymnasium Wirtschaft gewechselt; 55 Prozent waren Realschüler. 29 Prozent der Lernenden haben eine kaufmännische Vorbildung; Insbesondere die höhere Handelsschule ist hier ein bedeutender Zubringer; 77 Prozent der Probanden besuchten diese Schulform vor dem Fachgymnasium Wirtschaft. Abgeschlossene Berufsausbildungen im kaufmännisch-verwaltenden Bereich liegen hingegen nicht vor. Der Anteil der Lernenden mit Migrationshintergrund beträgt 22 Prozent. Insbesondere die deutliche Streuung hinsichtlich Alter und Vorbildung der Lernenden spiegelt die Heterogenität der Lernenden in beruflichen Ausbildungsgängen empirisch wider. Es ist zu prüfen, ob der Vorwissenstest die heterogene Struktur der Lernenden über das Literacy/Numeracy-Konzept einerseits sowie über die curriculare Anbindung andererseits angemessen berücksichtigen kann.

5.1.3.1 Ergebnisse auf Basis der klassischen Testtheorie

Für eine detaillierte Beschreibung der Auswertung des Vorwissenstests unter Rückgriff auf die klassische Testtheorie sei auf Winther (2006) verwiesen. Nachfolgend werden zum einen ausgewählte Befunde auf Basis des Ingesamttestscores und zum anderen die Weiterentwicklungen im Hinblick auf die domänenverbundenen Skalen economic literacy und economic numeracy dargestellt.

Befunde auf Basis des Ingesamttestscores

Der Vorwissenstest kann deskriptiv wie folgt beschrieben werden: Die Antworten für die Re-Analyse des Tests wurden dichotom kodiert, so dass max. 27 richtige Antworten vorliegen konnten. Im Mittel konnten die Probanden 12,41 Aufgaben lösen; die Standardabweichung beträgt 4,855. Der Standardfehler des Mittelwerts ist mit 0,261 für einen Leistungstest vergleichsweise gering. Die Verteilung des Antwortverhaltens weist die Form einer Normalverteilung (Schiefe = -0,02) mit leicht flacherem Verlauf auf (Kurtosis = -0,732).

Eine der zentralen Fragen der Re-Analyse befasste sich damit, ob individuelle Merkmale der Lernenden wie Geschlecht, Alter, kaufmännische Vorbildung oder Migrationshintergrund zu Unterschieden in der Vorwissensleistung führen. Interessant war vor allem, inwieweit die kaufmännische Vorbildung zu höheren Lösungsquoten der Testitems beiträgt. In Anlehnung an die Studien von Voss et al. (Voss, Blais, Means, Greene & Ahwesh, 1986) konnte erwartet werden, dass sich Vorwissensunterschiede stärker über die allgemeine Vorbildung als über eine konkrete kaufmännisch berufliche Vorbildung erklären lassen. Das Ergebnis zeigt jedoch, dass sich auf die Zubringerschulen generell keine Vorwissensunterschiede der Lernenden zurückführen lassen: Gymnasiasten erzielten vergleichbare Leistungen wie Realschüler ($t = 0,219$; $df = 292$; $p = ,834$) und Lernende mit aus der Schulbiographie ableitbarer kaufmännischer Vorbildung unterschieden

sich hinsichtlich ihres erreichten Ingesamtttestscores nicht von Lernenden ohne formale kaufmännische Vorbildung ($t = 1,142$; $df = 338$; $p = 0,254$; vgl. Abbildung 5.1.4).

Abbildung 5.1.4 T-Test zum Vergleich der Vorwissensleistungen in Abhängigkeit zur Vorbildung

Vorbildung	Mean	SD	T-Wert	p-Wert
Allgemeine schulische Vorbildung				
Realschulabsolventen (n=186)	12,61	4,830	0,210	0,834
Gymnasiasten (n=108)	12,49	4,751		
Kaufmännische berufliche Vorbildung				
Kaufmännische Vorbildung (n=100)	12,83	4,748	1,142	0,254
Keine Vorbildung (n=240)	12,17	4,934		

Anders als bei der Studie von Voss und Kollegen stellt sich über die allgemeine Vorbildung der Lernenden kein übergeordneter Leistungsfaktor heraus, der beispielsweise Lernende mit gymnasialer Bildung im Vergleich zu Absolventen der Realschulen dazu befähigt, an Alltagskontexte gebundene ökonomische Aufgabenstellungen über heuristische Methoden und Argumentationsstrategien besser zu lösen. Der Unterschied zwischen den Ergebnissen des Vorwissenstests und der Studie von Voss und Kollegen kann einerseits darin liegen, dass Ausbildungen am College zu stärkeren allgemeinen kognitiven Dispositionen führen, als dies bei Schulformen des Sekundarbereichs möglich ist. Andererseits zeigen am Fachgymnasium Wirtschaft mit Hilfe zweier Subskalen des kognitiven Fähigkeitstests für 4. bis 12. Klassen (KFT; Heller & Perleth, 2000), durchgeführte Intelligenztests, dass sich Lernende, die vom Gymnasium auf das Fachgymnasium wechselten, nicht von Realschulabsolventen unterscheiden ($t < 1,002$; $df = 257$; $p > 0,1$).

Wohl aber zeigt ein Vergleich der am Fachgymnasium erhobenen Daten mit den bei Heller und Perleth (2000, S. 83) veröffentlichten Angaben zu zentralen Tendenzen von Gymnasiasten deutliche Unterschiede (Abbildung 5.1.5). Für die Erhebungen am Fachgymnasium Wirtschaft wurden die Subskala Wortanalogien (V3-Form A) zur Erfassung verbaler Fähigkeiten und die Subskala Faltaufgaben (N3-Form A) zur Erfassung figural-räumlicher Fähigkeiten eingesetzt. Beide kognitive Dimensionen werden für die Bearbeitung sprachlich-argumentativer und mathematisch-analytischer Aufgaben des Vorwissenstests als bedeutsam angesehen. Die Abbildung 5.1.5 verdeutlicht, dass Lernende des Fachgymnasiums Wirtschaft in beiden Subtests geringere Mittelwerte aufweisen. Ein Vergleich der Quartile belegt insbesondere hinsichtlich des Subtests Wortanalogien diese Unterschiede deutlich: Am Fachgymnasium Wirtschaft liegen 50 Prozent der Schüler

unterhalb sieben richtiger Antworten; bei der Vergleichsstichprobe am Gymnasium beträgt der Median 11.

Abbildung 5.1.5 Leistungen in den KFT-Subskalen (Winther, 2006, S. 202)

	Mean	SD	Schiefe	PR25	PR50	PR75
N = 99; Klasse 11; Gymnasium, Testform A (Heller & Perleth, 2000)						
Wortanalogien	10,52	3,10	- 0,18	8	11	13
Faltaufgaben	11,54	2,98	- 0,77	10	12	14
N = 259; Klasse 11; Fachgymnasium Wirtschaft, Testform A						
Wortanalogien	7,57	2,52	0,08	6	7	9
Faltaufgaben	10,71	3,15	- 0,69	9	11	13

Die Unterschiede zwischen den Mittelwerten sind statistisch signifikant; die Schüler des Fachgymnasiums liegen hinsichtlich des Subtests Wortanalogien ca. eine Standardabweichung unterhalb der Gymnasiasten; hinsichtlich des Subtests Faltaufgaben beträgt der Unterschied etwas weniger als ein halbe Standardabweichung. Diese Befunde bestätigen, dass am Fachgymnasium eine im Mittel leistungsschwächere Schülerschaft im Vergleich zu den Gymnasien ausgebildet wird (vgl. insbesondere TOSCA; Köller, Watermann, Trautwein & Lüdtke, 2004).

Im Hinblick auf die individuellen Merkmale Geschlecht und Migrationshintergrund sind die Ergebnisse mit denen anderer kognitiver Leistungsstudien vergleichbar:

Abbildung 5.1.6 T-Test zum Vergleich der Vorwissensleistungen in Abhängigkeit von Geschlecht und Migrationshintergrund

	Mean	SD	T-Wert	p-Wert
Geschlecht				
männlich (n=178)	13,83	4,466	6,069	0,000
weiblich (n=162)	10,76	4,846		
Migrationshintergrund				
Muttersprachler (n=265)	12,60	4,891	1,679	0,096
Fremdsprachler (n=75)	11,53	4,794		

Lernende ohne Migrationshintergrund weisen in der Tendenz bessere Vorwissensleistungen auf ($t = 1,679$; $df = 338$; $p = 0,096$; $d = 0,22$) und männliche Pro-

banden erzielten einen höheren Ingesamtestscore als weibliche ($t = 6,069$; $df = 337$; $p = 0,000$; $d = 0,66$; Abbildung 5.1.6).

Insgesamt macht die Auswertung des Vorwissenstests über den Ingesamtestscore deutlich, dass der Test für verschiedene Subgruppen valide ist und sich keine unerwarteten Floor- oder Ceiling-Effekte für einzelne Subgruppen einstellen. Insbesondere die vergleichbaren Vorwissensleistungen von Probanden mit und ohne kaufmännischer Vorbildung sind zu diskutieren; sie lassen sich jedoch vor dem Hintergrund des Literacy/Numeracy-Konzepts plausibel erklären und replizieren im Ergebnis Befunde der Experten-Novizen-Forschung in fachübergreifenden Domänen, die weniger strukturiert und regelbasiert als akademische Fachbereiche sind.

Die Fokussierung auf das Literacy/Numeracy-Konzept vor dem Hintergrund der Re-Analyse des Vorwissenstests hat zum Ziel zu prüfen, ob mit dem Vorwissenstest unterschiedliche kognitive Fähigkeitsstrukturen – konzeptuale und prozedurale Kompetenz – erfasst werden konnten und wie sich die Vorwissensleistungen in diesen Kompetenzbereichen auf domänenspezifische Leistungen in den Lernprozessen am Fachgymnasium Wirtschaft beziehen lassen.

Befunde auf Basis zweier domänenverbundener Kompetenzbereiche

Als Ergebnis der inhaltlichen Analyse des Vorwissenstests vor dem Hintergrund des Literacy/Numeracy-Konzepts wurde der Test in zwei verschiedene Skalen zerlegt: die Skalen der economic literacy und der economic numeracy. Kognitionspsychologisch werden über diese Skalen zwei unterschiedliche Kompetenzbereiche erfasst: konzeptuale Kompetenz auf Basis überwiegend deklarativer Wissensbestände und prozedurale Kompetenz als Ergebnis der Aktivierung prozeduraler kognitiver Fähigkeiten und Strukturen. Die Skalenkennwerte vor dem Hintergrund der klassischen Testtheorie bestätigen diese theoretische Annahme (Abbildung 5.1.7).

Abbildung 5.1.7 Skalen- und Itemkennwerte des kaufmännischen Vorwissenstests

	Itemanzahl	Interne Konsistenz (Cronbachs Alpha)	Mittlere Item- trennschärfe	Mittlere Item- schwierigkeit
Economic literacy	21	0,771	0,364	0,478
Economic numeracy	6	0,785	0,384	0,406

Beide Skalen weisen eine annehmbare Reliabilität auf; dabei kann die interne Konsistenz der Skala der economic numeracy mit einem Cronbachs Alpha von 0,785 bei nur sechs Items der Skala als sehr gut bewertet werden. Die Werte für die mittlere Itemschwierigkeit belegen, dass die Items der beiden Skalen verschiedene Schwierigkeitsbereiche abdecken. Kein Item wies eine extreme Schwierigkeit im Hinblick auf das Minimum und Maximum der Antwortvorgaben auf. Der Trennschärfekoeffizient wurde auf Basis der Korrelation eines Items mit

dem Gesamtskalenwert ermittelt (Korrelationskoeffizient nach Pearson). Diese sind für beide Skalen vergleichsweise gering ($< 0,40$), so dass angenommen werden muss, dass die Skalen nur begrenzt zwischen den Personen differenzieren können. Aufgrund der geringen Itemanzahl der Skala economic numeracy und des gewählten Trennschärfekoeffizienten ist zusätzlich davon auszugehen, dass für die Items dieser Skala eine Überschätzung des Itemtrennschärfekoeffizienten vorliegen kann. Allerdings weist kein Item einen Trennschärfekoeffizienten kleiner 0,20 oder eine negative Trennschärfe auf, so dass zunächst alle Items im Test enthalten bleiben. Die Skalen korrelieren mit 0,326 eher moderat, was zusätzlich für eine zweidimensionale Fähigkeitsstruktur spricht. Insgesamt ergab die Itemanalyse die erwarteten Skalen- und Itemkennwerte. Dies lässt darauf schließen, dass die in den zwei Skalen angenommenen Itemmengen für die Stichprobe der Lernenden am Fachgymnasium Wirtschaft jeweils eine einfaktorielle Struktur aufweisen und voneinander unterschieden werden können.

Während die Vorwissensleistung auf Basis des Ingesamtestestscores kein Prädiktor der domänenspezifischen Lernleistung ist ($\beta = ,140$; $p > ,05$), lassen sich auf Basis der zwei verschiedenen Vorwissensskalen Lernleistungen im Lerngebiet 1 am Fachgymnasium Wirtschaft differenzierter erklären. Dabei ist die Relevanz des domänenverbundenen Vorwissens für die Erklärung domänenspezifischer Lernleistung nicht nur empirisch gegeben. Vielmehr kommt dem Vorwissen ein theoretisch deskriptiver Wert bei der Analyse von Wissenserwerbsprozessen zu. Wissenserwerb wird an dieser Stelle als Aufbauprozess definiert (im Sinne des gemäßigten Konstruktivismus; vgl. u. a. Mandl, Gruber & Renkl, 1995), womit eine dynamische Sicht eingenommen werden kann. Eine solche Perspektive wurde bereits von Aebli (1978; 1981) vertreten, der Wissen als die Basis des Wahrnehmens und des Handelns sieht. In der Tradition von Aebli lässt sich der Wissenserwerb auf die Qualität der Instruktion, auf den Schema-Aufbau sowie auf die Mechanismen des Verstehens und Behaltens beziehen; eine inhaltsbezogene Sicht ist diesem Fokus immanent. Die Wissenspsychologie nimmt an, dass Wissen im Gedächtnis repräsentiert ist und dass repräsentiertes Wissen eine organisierte Struktur aufweist. Der Prozess des Wissenserwerbs wird primär durch zwei Prozessarten gesteuert: (1) Der komplexe Prozess der Aktivierung von Vorwissen macht – ausgehend von begrifflichen Knoten in semantischen Netzwerken – aufgrund einer Erregungsausbreitung zahlreiche weitere vorhandene Wissensstrukturen unmittelbar zugänglich. (2) Der Prozess der Elaboration verknüpft neue Wissensbestände mit dem Vorwissen, wobei die Reduzierung der Zieldifferenz (aktuell gegebenes und angestrebtes Wissensniveau) verfolgt wird. Es entstehen neue Relationen zwischen den Netzwerkteilen: Das Vorwissen wird erweitert; es wird gelernt.

In diesem Zusammenhang kann angenommen werden, dass unterschiedliche Fähigkeitsstrukturen in domänenverbundenen Kompetenzbereichen das Lernen im domänenspezifischen Kompetenzbereich beeinflussen. Dabei gilt, dass die allgemein anerkannte Unterscheidung zwischen deklarativem und prozeduralem Wissen (vgl. u. a. Gagné, 1977) aus der Perspektive des Wissensaufbaus eher

zweitrangig ist. Die Unterscheidung gewinnt erst unter Bezugnahme auf domänenspezifische Inhalte an Bedeutung. Hier kann die Frage gestellt werden, in welchem Umfang das Lernen domänenspezifischer Inhalte von unterschiedlichen Fähigkeitsstrukturen abhängig ist. Im Zuge einer ersten Überprüfung dieser Annahmen wurde ermittelt, welche Zusammenhänge sich zwischen dem domänenspezifischen Wissen als Ergebnis von Lernprozessen am Fachgymnasium Wirtschaft sowie dem domänenverbundenen Vorwissen ergeben. Hierzu wurden domänenspezifische Lernleistungen in den domänenspezifischen betriebswirtschaftlichen Inhaltsbereichen „Betriebswirtschaftliche Theorie“ und „Rechnungswesen/Controlling“ erhoben und in Beziehung zu den domänenverbundenen Vorwissensleistungen der Bereiche „economic literacy“ und „economic numeracy“ gesetzt. Unterstellt ist diesem wechselseitigen System von Domänenspezifität und Domänenverbundenheit die These, dass Lernende domänenspezifische Ergebnisse durch die Anwendung von domänenverbundenem Grund- und Regelwissen erzielen. Verarbeitet wird in diesen Annahmen die curriculare und fachdidaktische Frage des Zugriffs auf die Struktur eines Lernbereichs. Hierbei wird explizit davon ausgegangen, dass in beruflichen Domänen unterschiedliche fachliche Zugriffe unumgänglich und als authentisch für die Festlegung der inneren Konsistenz anzusehen sind. Zentrale Ergebnisse sind, dass domänenspezifische Leistungen von domänenverbundenen Leistungen abhängen und dass die Stärke der Abhängigkeit zentral von den Lerninhalten beeinflusst wird. In Abbildung 5.1.8 sind die Ergebnisse dargestellt:

Abbildung 5.1.8 Korrelative Zusammenhänge zwischen domänenverbundenen und domänen-spezifischen Leistungsbereichen (n = 340; vgl. Winther & Achtenhagen, 2008c)

Vergleich zwischen	domänenspezifische Leistungsbereiche	
	Betriebswirtschaftliche Theorie	Rechnungswesen/Controlling
Einfache Korrelation zwischen domänenverbundener Vorwissensleistung und domänenspezifischer Lernleistung	,189***	,177***
Partialkorrelation zwischen domänenverbundener Vorwissensleistung und domänenspezifischer Lernleistung bei Auspartialisierung der Leistung im Bereich „economic literacy“	,004	,154**
Partialkorrelation zwischen domänenverbundener Vorwissensleistung und domänenspezifischer Lernleistung bei Auspartialisierung der Leistung im Bereich „economic numeracy“	,198***	,066

Anmerkung: Korrelation nach Pearson; ***p < 0,01; **p < 0,05

Die Daten belegen für den Betriebswirtschaftslehreunterricht im Fachgymnasium Wirtschaft den Einfluss domänenverbundener Vorwissensleistungen. Betrachtet man die einfachen Korrelationen für den Wissenszuwachs nach 40 Un-

terrichtsstunden, so sind signifikante Zusammenhänge für die Entwicklung der domänenspezifischen Leistungen feststellbar; das trifft sowohl für die eher betriebswirtschaftlichen Leistungen als auch für die Leistungen im Bereich Rechnungswesen/Controlling zu ($r_{\text{BWL}} = 0,189^{***}$; $r_{\text{REWE}} = 0,177^{***}$). Werden die domänenverbundenen Vorwissensleistungen jedoch herauspartialisiert, so ist anhand der signifikanten Abnahme der Korrelationskoeffizienten zu erkennen, dass sich die domänenspezifischen Leistungen in hohem Maße auf die domänenverbundenen Kompetenzbereiche zurückführen lassen. Diese Befunde belegen die Annahme, dass Grund- und Regelwissen in den beschriebenen domänenverbundenen Leistungsbereichen als Grundlage für die Bewältigung domänenspezifischer Anforderungen vorausgesetzt wird. Für den Unterricht im Fach Betriebswirtschaft mit Rechnungswesen/Controlling sind die relevanten domänenverbundenen Kompetenzbereiche über die Skalen economic literacy und economic numeracy beschrieben, mit denen konzeptuale und prozedurale Fähigkeitsstrukturen erfasst werden können. Auf Grundlage der Ergebnisse der klassischen Analyse wird im Folgenden die angenommene zweidimensionale Kompetenzstruktur mit Hilfe probabilistischer Verfahren geprüft.

5.1.3.2 Ergebnisse auf Basis der probabilistischen Testtheorie

Die Re-Analyse des Vorwissenstests konzentriert sich auf drei zentrale Fragestellungen: (1) Wird mit dem Vorwissenstest eine latente Fähigkeitsstruktur erfasst, oder sind über die Skalen economic literacy und economic numeracy verschiedene Fähigkeiten angesprochen? – Dies ist die Frage der Dimensionalität von Kompetenz. (2) Liegt mit dem kaufmännischen Vorwissenstest ein Instrument vor, das Items unterschiedlicher Schwierigkeitsgrade aufweist? – Bei dieser Frage steht die Itemkalibrierung im Mittelpunkt. Im Gegensatz zu der Studie im Bereich der Dualen Berufsausbildung (vgl. Kapitel 5.2) erfolgte die Konstruktion der Items des Vorwissenstests nicht auf Basis eines heuristischen Modells und in Anlehnung an eine Construct Map (Wilson, 2005), so dass keine begründeten Vorstellungen darüber vorliegen, wie komplex einzelne Items sind und wie diese Items eine gemeinsame Skala bilden. Die Auswertungen im Rahmen der klassischen Testtheorie lieferten zu dieser Frage erste Hinweise. (3) Gelingt es, mit dem im Vorwissenstest verankerten Literacy/Numeracy-Konzept die Heterogenität der Probanden insbesondere im Hinblick auf ihre berufsbezogene kaufmännische Vorbildung derart zu berücksichtigen, dass keine systematischen Vor- oder Nachteile für einzelne Subgruppen bestehen? – Zur Beantwortung dieser Frage werden Items dahingehend untersucht, ob bei gleicher Personenfähigkeit einzelne Subgruppen eine höhere oder geringere Lösungswahrscheinlichkeit aufweisen.

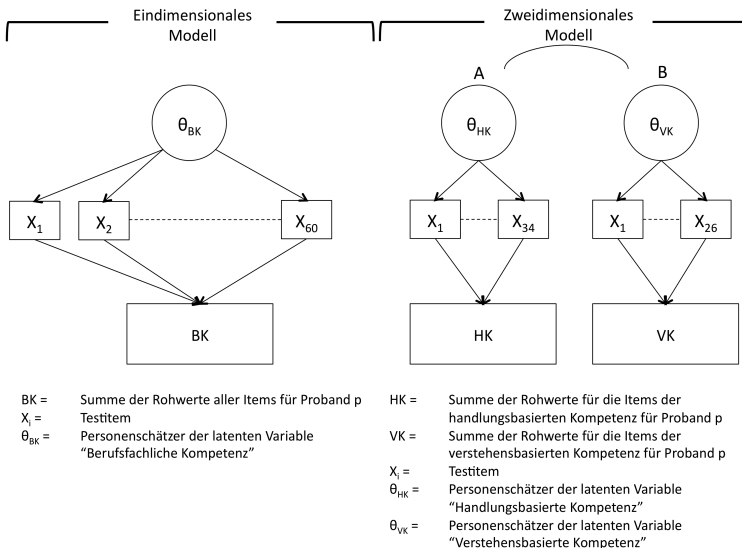
Die Dimensionalität kaufmännischer Kompetenz

Der Test zur Erfassung kaufmännischer Vorwissensleistungen unterscheidet zwischen zwei Fähigkeitsdimensionen: domänenverbundene konzeptuale Kompe-

tenz und prozedurale Kompetenz. Der Test ist dabei so konstruiert, dass ein Item jeweils nur eine der beiden möglichen latenten Fähigkeitsstrukturen messen kann. Der Fall, dass ein Item genau einer Kompetenzdimension zugeordnet wird, wird als between-item-Multidimensionalität beschrieben (Adams, Wilson & Wang, 1997; Wang, Wilson & Adams, 1997).

Um zu prüfen, ob das theoretisch formulierte Modell, das auf Basis des Literacy/Numeracy-Konzepts zwischen zwei Fähigkeitsstrukturen differenziert, für die erhobenen Daten passend ist, werden das eindimensionale und das zweidimensionale Modell miteinander verglichen. Abbildung 5.1.9 stellt die formalen Unterschiede dar.

Abbildung 5.1.9 Vergleich zwischen ein- und zweidimensionalem Modell



Für das eindimensionale Modell gilt, dass kaufmännische Vorwissenleistung als latente Variable höherer Ordnung betrachtet wird. Demzufolge ist der Summenscore aller 27 Items des Vorwissenstests die suffiziente Statistik für den Schätzer der übergeordneten Personenfähigkeit „Kaufmännische Kompetenz“. Für das zweidimensionale Modell gilt hingegen, dass kaufmännische Kompetenz über zwei sich von einander unterscheidenden Fähigkeitsstrukturen gemessen wird. Hierbei erfolgt die Schätzung der latenten Fähigkeiten über beide Skalen des Literacy/Numeracy-Konzepts simultan. Für den Test zur Messung kaufmännischer Kompetenz auf Basis des Vorwissenstests sind 21 Items der konzeptuellen

Fähigkeit und sechs Items eindeutig der prozeduralen Fähigkeit zugeordnet. Die simultane Schätzung der Item- und Personenparameter ermöglicht es, Messwerte für beide Skalen unabhängig von einander zu gewinnen. Zudem kann die Korrelation zwischen beiden Skalen als Effektstärke der Dimensionalität interpretiert werden (Embretson & Reise, 2001, p. 75).

Um Modelle direkt miteinander vergleichen zu können, sind IRT-Modelle zu wählen, die sich hierarchisch aufeinander beziehen lassen. Für das eindimensionale Modell wird auf das Rasch-Modell (Rasch, 1960) und für das zweidimensionale Modell auf das Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM; Adams, Wilson & Wang, 1997) zurückgegriffen. Beide Modelle lassen sich mit Hilfe der Software *ConQuest* berechnen (Wu, Adams, Wilson & Haldane, 2007). Das MRCML-Modell ist ein generalisiertes IRT-Modell der Rasch-Familie und kann folglich auch das 1-parametrische logistische Modell repräsentieren. Zur Bestimmung der Signifikanz wird ein Likelihood-Ratio-Test auf Basis der Chi-Square-Statistik durchgeführt. Die Fit-Statistik für die Kompetenzdimensionen ist in Abbildung 5.1.10 dargestellt.

Abbildung 5.1.10 Fit-Statistik für das ein- und zweidimensionale Modell

	Eindimen- sionales Modell	Zweidimen- sionales Modell	Differenz
Anzahl der Iterationen	21	21	0
Deviance (-2Log-Likelihood)	6.143,84	5.853,57	290,27
Anzahl der geschätzten Parameter	29	32	3

In das eindimensionale Modell sind alle 27 Items zur Messung der Vorwissensleistungen eingegangen; zur Modellierung der Dimensionen im zweidimensionalen Modell wurden die 21 Items der economic literacy-Skala der Dimension I und die sechs Items der economic numeracy-Skala der Dimension II zugeordnet. Beide Modelle wurden in 21 Iterationen geschätzt. Die Güte des Modellfits kann relativ zu der Veränderung im Deviance-Wert verglichen werden. Die Differenz zwischen den Deviance-Werten der beiden Modelle ist als Chi-Square mit drei Freiheitsgraden verteilt. Die Freiheitsgrade ergeben sich aus der Differenz der geschätzten Parameter.

Während im eindimensionalen Modell 27 Itemparameter sowie ein Populationsmittelwert und ein Varianzschätzer die Parameteranzahl ergeben, werden im zweidimensionalen Modell neben den Itemparametern zwei Populationsmittelwerte für jede latente Dimension, zwei Varianzschätzer und ein Kovarianzschätzer ermittelt. Die geschätzte Differenz zwischen den Deviance-Werten von 290,27 ist signifikant ($df = 3$; $p < ,001$), so dass gefolgert werden kann, dass das eindimensionale Modell die Daten weniger gut fitet als das zweidimensionale Modell. Folglich werden über die Skalen der economic literacy und der economic numeracy unterschiedliche latente Dimensionen gemessen. Zur Beurteilung der Effekt-

stärke wird die Korrelation zwischen den latenten Dimensionen herangezogen. Je niedriger die latenten Dimensionen korrelieren, desto höher ist die praktische Signifikanz eines mehrdimensionalen Modells. Konzeptuale und prozedurale Kompetenz korrelieren mit $r = 0,209$ (Kovarianz = 0,528). Diese geringe Korrelation spricht ebenfalls deutlich für eine zweidimensionale Lösung.

Zur Veranschaulichung der Berechnung mit Hilfe der Software *ConQuest* ist der einfachste Fall – das eindimensionale Rasch-Modell – in Abbildung 5.1.11 dargestellt (für die notwendigen Eingaben zur Berechnung eines zweidimensionalen Modells vgl. Abbildung 5.2.18 in Kapitel 5.2.4.2). Für die Berechnung von psychometrischen Modellen in *ConQuest* ist ein Ausgangsdatensatz im ASCII-Format bereitzustellen. Für das eindimensionale Rasch-Modell am Beispiel des kaufmännischen Vorwissenstests besteht dieser Datensatz aus 33 Spalten. Die ersten fünf Spalten enthalten die Kennung der Probanden (ID) und in den folgenden 27 Spalten ist das Antwortverhalten (RESPONSES) erfasst. Die Antworten der Lernenden sind mit 0 und 1 dichotom skaliert (CODES). Zur Berechnung der Modelle ist im Input-Bereich „Define model“ das ausgewählte Modell zu klassifizieren. Die Zeile „Model item“ ist das entsprechende Eingabekommando für das 1PL- oder Rasch-Modell. In *ConQuest* stehen verschiedene Outputs zur Verfügung. Im gewählten Beispiel basiert die Ausgabe der psychometrischen Analyse (.shw) auf latenten Schätzern. Zusätzlich werden für die Schätzung der Personenfähigkeiten die Ausgabe des WLE und des EAP/Plausible Value angefordert. Über das Kommando „Itanal“ sind Kennwerte der klassischen Testtheorie abrufbar.

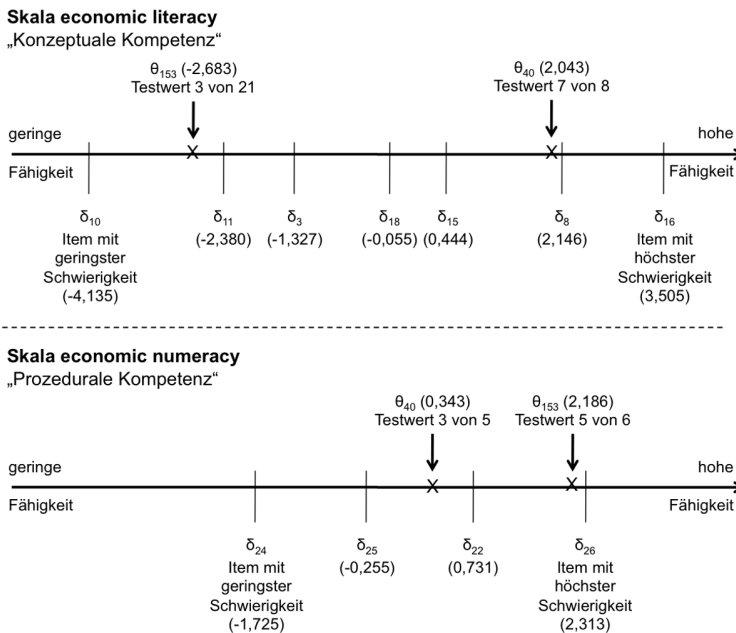
Abbildung 5.1.11 *ConQuest-Input zur Berechnung des eindimensionalen Rasch-Modells*

<pre>/* Title and data definition */ Datafile Vorwissen_dichotom.dat; Format id 1-5 responses 6-33;</pre>	Ausgangsdatensatz; 27 Antworten in den Spalten 6 bis 33
<pre>/* Valid codes and scoring */ Codes 0,1; set warnings=no,update=yes; set constraints=none;</pre>	Dichotome Daten; 1 für Erfolg / 0 für Fehler
<pre>/* Define model */ Model item;</pre>	1PL oder Rasch-Modell (Schätzung von Itemschwierigkeiten)
<pre>/* Estimate model parameters */ Estimate;</pre>	
<pre>/* Define output files */ Show ! estimates=latent >> Vorwissen_ALL.shw; Show cases! estimates=wle >> Vorwissen_ALL.wle; Show cases! estimates=eap >> Vorwissen_ALL.eap; Export parameters >> Vorwissen_ALL.prm; Itanal >> Vorwissen_ALL.itn;</pre>	Ausgabe der psychometrischen Analyse; Ausgabe der Schätzer der Personenfähigkeit (WLE und EAP); Ausgabe klassischer Kennwerte
<pre>quit;</pre>	

Itemkalibrierung und Parameterschätzung

Es kann davon ausgegangen werden, dass Skalen dann erfolgreich modelliert worden sind, wenn die einzelnen Testitems, die diese Skalen repräsentieren, auf dem Kontinuum des Personenmerkmals ausreichend streuen. Die Streuung bezieht sich in diesem Zusammenhang auf unterschiedliche Schwierigkeitsgrade, die die Items der Skalen aufweisen. Die Kalibrierung der Items erfolgt dabei auf der so genannten Logit-Skala. Diese Skala stellt die Relation zwischen Personenfähigkeit und Itemschwierigkeit dar. In Abbildung 5.1.12 sind für die Skalen economic literacy und economic numeracy einzelne Items mit ihren Itemschwierigkeiten sowie einzelne Testpersonen mit ihren Personenfähigkeiten auf dem Kontinuum der zwei Kompetenzbereiche abgebildet.

Abbildung 5.1.12 Itemkalibrierung in den Dimensionen kaufmännischer Kompetenz



Die Abbildung verdeutlicht, dass die Skala der economic literacy mit ihren 21 Testitems ein breiteres Schwierigkeitsspektrum umfasst als die Skala der economic numeracy, die mit sechs Testitems repräsentiert ist. Das leichteste Item (Item 10) der economic literacy-Skala weist einen Schwierigkeitsparameter von $-4,135$ Logits auf; das anspruchsvollste Item (Item 16) von $3,505$. Die übrigen 19 Items dieser Skala sind gleichmäßig zwischen den Skalenendpunkten – mit einer leich-

ten Häufung im mittleren Schwierigkeitsbereich – verteilt. Die prozedurale Kompetenz wird über Items im Bereich von -1,725 bis 2,313 Logits erfasst. Insgesamt wird die Skala der economic numeracy über zwei leichte Items ($\delta_{24} = -1,725$; $\delta_{23} = -1,631$), zwei Items im mittleren Schwierigkeitsbereich ($\delta_{25} = -0,255$; $\delta_{23} = 0,731$) sowie über zwei schwierige Items ($\delta_{27} = 1,475$; $\delta_{26} = 2,313$) repräsentiert. Um auf den Kontinua der zwei verschiedenen Personenmerkmale eine spezifische Person zu verorten, ist diese Person mit den Items, über die das jeweilige Personenmerkmal repräsentiert ist, zu testen. Das Antwortverhalten ausgedrückt als kognitive Fähigkeit der Person (θ_v) markiert die Lage der Person auf dem Kontinuum des zu untersuchenden Personenmerkmals.

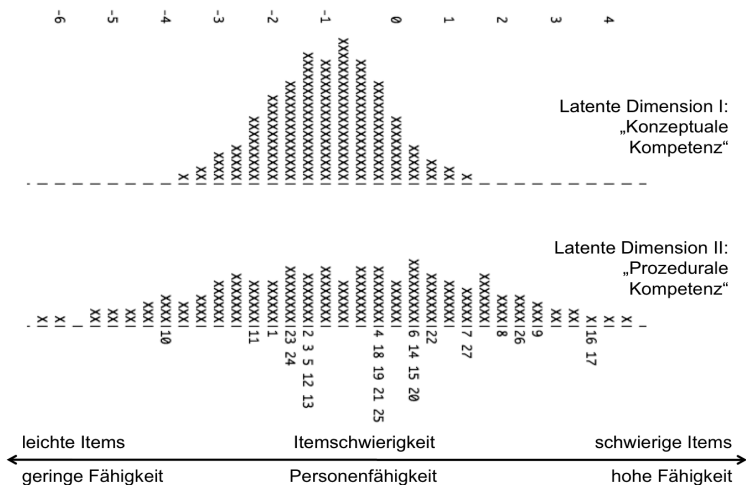
In Abbildung 5.1.12 sind zwei Beispiele gewählt, die zeigen, dass sich die über zwei Skalen erfassten Fähigkeitsstrukturen deutlich voneinander unterscheiden – eine hoch ausgeprägte Fähigkeit im konzeptualen Kompetenzbereich geht folglich nicht zwangsläufig mit einer hohen prozeduralen Kompetenz einher. Die geringe Korrelation von 0,209 zwischen den latenten Dimensionen ist ein Indiz für diesen Befund. Proband 153 verfügt im Bereich der konzeptualen Kompetenz über eine sehr geringe Personenfähigkeit (WLE) von $\theta_{153} = -2,683$. Dieser Fähigkeitsparameter wird auf der Basis geschätzt, dass drei von 21 Items der economic literacy-Skala richtig gelöst wurden. Alle Probanden, die ebenfalls drei Testitems erfolgreich bearbeiten haben, verfügen, unabhängig von den Schwierigkeitsparametern der gelösten Items, über den gleichen Schätzer der Personenfähigkeit. Um also zu erfahren, ob das Antwortmuster der Probanden den Annahmen des Modells entspricht – somit keine oder nur wenige unerwartete Antwortmuster vorliegen –, sind die Antworten der Probanden einzeln zu analysieren. Am Beispiel des Probanden 153 wird dabei deutlich, dass drei der vier leichtesten Items gelöst werden konnten. Im Bereich der prozeduralen Kompetenz verfügt der Proband 153 über eine vergleichsweise hohe Fähigkeitsausprägung. Werden fünf von sechs Testitems der Skala der economic numeracy gelöst, beträgt der Schätzer der Personenfähigkeit 2,186 Logits. Für den Probanden 153 gilt, dass entsprechend der Annahmen des Modells das schwierigste Item ($\delta_{26} = 2,313$) nicht gelöst wurde.

Die Schätzung der Personenfähigkeit ist auch dann möglich, wenn für einen Probanden nicht alle Informationen aus den zu bearbeitenden Skalen vorliegen. Proband 40 hat beispielsweise nur einen Teilbereich des Vorwissenstests – acht Items der economic literacy-Skala und fünf Items der economic numeracy-Skala – bearbeitet. Die Schätzung der Personenfähigkeit ist in diesem Fall möglich, da sich die Messwerte testunabhängig und linear verhalten. Für den Probanden 40 ergibt sich bei sieben von acht gelösten Items ein Personenschätzer von 2,043 Logits im Bereich der konzeptualen Kompetenz. Der Fähigkeitsschätzer von 0,343 Logits im prozeduralen Kompetenzbereich ergibt sich aus der erfolgreichen Bearbeitung von drei der möglichen fünf Testitems.

Für alle Testitems und alle Probanden lässt sich die Skalierung der Personenfähigkeiten und die Kalibrierung der Testitems für die zwei Skalen mit Hilfe einer Wright Map (Wilson, 2005; Abbildung 5.1.13) darstellen. Die Items sind entsprechend ihrer Schwierigkeitsparameter entlang der Logit-Skala angeordnet; die

Schätzer der Personenfähigkeit sind für die zwei latenten Dimensionen über ein Histogramm beschrieben, bei dem jedes X für 2,9 Fälle steht und den expected a posteriori-Schätzer (EAP/PV) repräsentiert. Die Verteilungen der Personenfähigkeiten auf den zwei Skalen unterscheiden sich deutlich. Der Populationsmittelwert der Verteilung der konzeptuellen Kompetenz beträgt -1,109 und die Standardabweichung 1,064 Logits. Der Standardfehler des Mittelwerts ist mit 0,057 sehr gering. Insgesamt zeigen diese Werte, dass im Bereich der konzeptuellen Kompetenz die unteren Leistungsbereiche dominieren. Über zwei Drittel der Lernenden weisen Fähigkeitsstrukturen auf, mit denen Items eines mittleren Schwierigkeitsgrades mit einer Wahrscheinlichkeit von weniger als 50 Prozent gelöst werden können.

Abbildung 5.1.13 Wright Map: Kaufmännische Kompetenz



Anmerkung: Jedes X repräsentiert 2,9 Fälle.

Die Verteilung der Personenfähigkeiten des prozeduralen Kompetenzbereichs ist mit einem Populationsmittelwert von -0,869 (0,127) und einer Standardabweichung von 2,373 beschrieben. Die relativ niedrigen Messfehler erlauben die Schlussfolgerung, dass es mit den gegebenen Skalen gelingt, die Personen zuverlässig auf verschiedenen Kompetenzdimensionen zu verorten. Ein Vergleich der Verteilungen zeigt zudem, dass die Skala der economic numeracy stärker als die Skala der economic literacy zwischen den Personen differenzieren kann. Dies wird durch die Reliabilitätskennwerte der latenten Dimensionen bestätigt. In Abbildung 5.1.14 sind die Reliabilitäten auf Basis des WLE und des EAP/PV dargestellt.

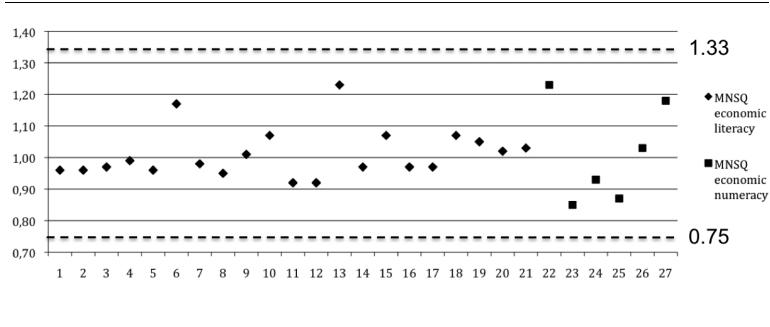
Abbildung 5.1.14 Reliabilitäten

Latente Dimension	EAP/PV-Reliabilität	WLE-Reliabilität
Konzeptuale Kompetenz	0,502	0,540
Prozedurale Kompetenz	0,585	0,604

Anmerkung: EAP/PV = expected a posteriori/plausible value; WLE = weighted likelihood estimator

Unter Rückgriff auf die geringe Itemanzahl ist die Reliabilität der economic numeracy-Skala mit 0,604 (basierend auf dem WLE) angemessen; die Reliabilität der economic literacy-Skala ist im Vergleich mit 0,540 sehr gering, wird die Anzahl der Items, die diese Skala repräsentieren, berücksichtigt.

Abbildung 5.1.15 Fit-Statistik für die Itemparameter



Im Hinblick auf die Qualität der Items ist zusammenfassend festzustellen, dass die gewichteten Abweichungsquadrate (weighted MNSQ; Wright & Masters, 1982) für alle Items innerhalb der zulässigen Grenzen von 0,75 bis 1,33 (Adams & Khoo, 1996; Bond & Fox, 2001; Chang, Lin & Lin, 2007; vgl. Abbildung 5.1.15) liegen und symmetrisch um den exakten Fit von 1,00 verteilt sind. Dabei variieren die Infit-Schätzer der economic literacy-Skala zwischen 0,92 und 1,23 und die Schätzer der economic numeracy-Skala zwischen 0,85 und 1,23. Am Beispiel der Skala der economic numeracy sind in Abbildung 5.1.16 Itemkennwerte zusammengestellt:

Abbildung 5.1.16 Itemkennwerte der Skala economic numeracy

Item	Item-schwierigkeit	weighted MNSQ	T-Wert	Konfidenzintervall	Lösungsquote in %	Item-diskrimination
Item 22	0,73	1,23	2,10	(0,85; 1,15)	30,26	0,22
Item 23	-1,63	0,85	-1,90	(0,84; 1,16)	60,81	0,57

Item	Item-schwierigkeit	weighted MNSQ	T-Wert	Konfidenzintervall	Lösungsquote in %	Item-diskrimination
Item 24	-1,73	0,93	-1,10	(0,84; 1,16)	61,96	0,58
Item 25	-0,25	0,87	-1,80	(0,85; 1,15)	42,94	0,62
Item 26	2,31	1,03	0,30	(0,79; 1,21)	13,83	0,36
Item 27	1,47	1,18	1,20	(0,69; 1,31)	20,37	0,33

Die Abbildung zeigt, dass Hinweise auf eine zu geringe Trennschärfe der einzelnen Items auf Basis der T-Statistik nur für Item 22 der economic numeracy-Skala vorliegen. Der T-Wert liegt mit 2,10 außerhalb des ermittelten Konfidenzintervalls; die Itemdiskrimination weist mit 0,22 jedoch einen noch annehmbaren Wert auf, so dass das Item in der Skala enthalten bleibt.

Differenzielle Itemanalyse zur Prüfung von Vorbildungseffekten

Ein zentrales Ergebnis der Auswertungen des kaufmännischen Vorwissenstests auf Basis der klassischen Testtheorie war, dass die Vorwissensleistung von Lernenden mit kaufmännischer, formal in beruflichen Vollzeitschulen erworbener Vorbildung sich nicht von der gezeigten Leistung der Lernenden ohne kaufmännische Vorbildung unterscheidet. Die Mittelwertdifferenz im Hinblick auf diese Fragestellung ist nicht signifikant ($t = 1,142$; $df = 338$; $p = 0,254$). Dieser Befund lässt sich unter Rückgriff auf die Studien von Voss und Kollegen (Voss, Blais, Means, Greene & Ahwesh, 1986) erklären, indem auf das Konzept der Literacy/Numeracy sowie auf den fachübergreifenden Charakter der kaufmännischen Domäne abstrahiert wird. Beide Erklärungsansätze gehen davon aus, dass ein Teil der im Test gezeigten Performanz auf einem allgemeinen, domänenverbundenen Grund- und Regelwissen basiert, das zudem Elemente eines spezifischen, gesellschaftlich geteilten Allgemeinwissens enthält (Hofstadter, 1985; Voss & Post, 1988; Alexander, Schallert & Hare, 1991).

Im Hinblick auf die curriculare Validität des Tests bietet dieser Befund jedoch Anlass für vertiefende Analysen. Die Curricula der kaufmännischen Bildungsgänge weisen insbesondere in den ersten Unterrichtssequenzen weite Überschneidungsbereiche auf. Hier werden zentrale Ideen kaufmännischen Handelns und Entscheidens, grundlegende kaufmännische Konzepte und Einführungen in das Rechnungswesen bearbeitet. Auf diese Inhalte stellt auch der Vorwissenstest ab. Eine vertretbare Annahme in diesem Zusammenhang ist, dass Lernende mit kaufmännischer Vorbildung Vorteile bei der Bearbeitung des Tests haben, da sie (1) mit den Inhalten des Tests vertraut sind und (2) die Sprache sowie die Arbeitsmethoden in der kaufmännischen Domäne erlernt haben. Unter Rückgriff auf die probabilistische Testtheorie am Beispiel des Rasch-Modells wird im Folgenden der Frage nachgegangen, ob sich systematische Unterschiede zwischen Lernenden mit und ohne kaufmännische Vorbildung in den zwei Kompetenzdimen-

sionen bestätigen lassen. Hierbei wird anhand differenzieller Itemanalysen (Differential Item Functioning Analysis (DIF); vgl. Moore, 1996) geprüft, ob sich bei einzelnen Items der Skalen economic literacy und economic numeracy Unterschiede zwischen den Subgruppen ergeben. Zentrales Merkmal der DIF-Analyse ist, dass keine generellen differenziellen Effekte zwischen Gruppen untersucht werden, sondern Unterschiede im Antwortverhalten von Personen mit gleicher Fähigkeit im Vergleich über die differenziell zu unterscheidenden Gruppen.

„If the responses to an item have different frequencies for different subgroups, that is evidence of differential impact of the item on those subgroups [...] Instead DIF studies focus on whether respondents at the same locations give similar responses across the different subgroups“ (Wilson, 2005, p. 165).

Bei der DIF-Analyse wird folglich geprüft, ob die Testitems „fair“ für alle Probanden sind, oder ob Items im Test enthalten sind, die eine Subgruppe systematisch bevor- oder benachteiligen. Empirisches Kriterium für die Existenz differenzieller Unterschiede ist die Differenz zwischen gruppenspezifischen Itemschwierigkeiten. Übertragen auf das Rasch-Modell ergibt sich folgende formale Struktur:

Rasch-Modell (Logarithmierter Wettquotient):

$$\ln \frac{P(X_{vi} = 1 | \theta_v, \delta_i)}{P(X_{vi} = 0 | \theta_v, \delta_i)} = \theta_v - \delta_i$$

Rasch-Modell für DIF:

$$\ln \frac{P(X_{vi} = 1 | \theta_v, \delta_i, \gamma)}{P(X_{vi} = 0 | \theta_v, \delta_i, \gamma)} = \theta_v - \delta_i - \gamma * X_i Z_v$$

Die Personenfähigkeit ist mit θ_v und die Itemschwierigkeit mit δ_i gekennzeichnet. Der DIF-Effekt beschreibt die Item*Gruppe-Interaktion (γ). X_i ist dabei der Indikator für das Item und Z_v der Indikator für die Persongruppen. Werden systematische Vorbildungseffekte auf Itemebene untersucht, ist Z_v ein dichotomer Gruppenindikator, bei dem die Subgruppe ohne kaufmännische Vorbildung (Referenzgruppe) mit 0 und die Subgruppe mit formaler kaufmännischer Vorbildung (Vergleichsgruppe) mit 1 kodiert ist. Zusammenfassend ergeben sich damit folgende Funktionsgleichungen für Items ohne DIF- und für Items mit systematisch angenommenem DIF-Effekt (vgl. Abbildung 5.1.17).

Abbildung 5.1.17 Funktionsgleichungen für die DIF-Analyse auf Basis des Rasch-Modells

Subgruppe	Z	Non-DIF-Items	DIF-Items
Referenzgruppe	0	$\theta_v - \delta_i$	$\theta_v - \delta_i$
Vergleichsgruppe	1	$\theta_v - \delta_i$	$\theta_v - \delta_i - \gamma^* X_i$

Für die Darstellung von DIF-Effekten wurden die Skalen economic literacy und economic numeracy separat analysiert. Im Hinblick auf die Schätzung der mittleren Fähigkeit ergeben sich für die Bereiche der konzeptualen und prozeduralen Kompetenz folgende Ergebnisse (vgl. Abbildung 5.1.18):

Abbildung 5.1.18 Mittlere konzeptuale und prozedurale Fähigkeit im Gruppenvergleich

	Fähigkeits-schätzer	Standard-fehler	Chi-Square (df)
Konzeptuale Kompetenz			
Kaufmännische Vorbildung	0,144	0,018	9,63 (1)
Keine kaufmännische Vorbildung	-0,144		
Prozedurale Kompetenz			
Kaufmännische Vorbildung	0,142	0,070	4,12 (1)
Keine kaufmännische Vorbildung	-0,142		

Für beide Kompetenzbereiche gilt, dass Lernende mit kaufmännischer Vorbildung eine signifikant höhere Leistung im Vorwissenstest erbracht haben. Im Bereich der konzeptualen Kompetenz beträgt der Unterschied 0,288 Logits und im Bereich der prozeduralen Kompetenz 0,284 Logits zugunsten der Vergleichsgruppe ($p < 0,05$). Insbesondere im Hinblick auf die Skala der economic literacy sind die differenziellen Effekte praktisch hoch bedeutsam; der Fähigkeitsschätzer ist achtmal so groß wie sein Standardfehler und der p-Wert des Chi-Squares ist kleiner 0,005. Dieses Ergebnis lässt darauf schließen, dass die Skala der economic literacy Items mit substantiellen DIF enthält. In Anlehnung an Paek (2002; zit. n. Wilson, 2005, p. 167) sind für die Beschreibung von Logit-Differenzen einzelner Items folgende Effektgrößen anzunehmen: Ein Differenzwert kleiner als 0,426 ist vernachlässigbar, Differenzen zwischen 0,426 und 0,638 sprechen für mittlere und Werte größer als 0,638 für hohe Effekte. Von den 21 Items der economic literacy-Skala weisen neun Items (42,86 %) mittlere oder hohe DIF-Effekte auf. In Abbildung 5.1.19 sind aus Sicht der Vergleichsgruppe die Interaktionen zwischen den Items und der Gruppenzugehörigkeit dargestellt. Ein positiver Interaktionseffekt bedeutet, dass sich die Itemschwierigkeit für die Vergleichsgruppe um den Schätzwert der Interaktion erhöht; ein negativer Schätzwert reduziert hingegen

die Itemschwierigkeit aus Perspektive der Lernenden mit kaufmännischer Vorbildung.

Abbildung 5.1.19 DIF-Analyse der Skala *economic literacy* (Item*Vorbildung)

Item	Schätzer	Standardfehler	Chi-Square (df)	p-Wert
Item 1	0,288	0,116	103,28 (20)	0,000
Item 5	0,311	0,110		
Item 7	0,388	0,146		
Item 8	0,452	0,174		
Item 13	0,466	0,163		
Item 15	0,783	0,192		
Item 16	-0,774	0,247		
Item 17	-1,472	0,247		
Item 21	-0,254	0,775		

Die Abbildung zeigt zusammenfassend, dass sechs der neun DIF-Items Lernende mit kaufmännischer Vorbildung systematisch benachteiligen (positiver Schätzwert der Interaktion), während bei drei Items die Itemschwierigkeit für die Referenzgruppe erhöht ist (negativer Schätzwert der Interaktion). Die differenziellen Unterschiede in der Vorwissensleistung lassen sich damit auf das unterschiedliche Antwortverhalten in den Subgruppen bezogen auf einzelne Items der Skala zurückführen. Diese Ergebnisse können erneut auf die Befunde der Studien von Voss und Kollegen (Voss, Blais, Means, Greene & Ahwesh, 1986) sowie auf die Annahmen des Literacy/Numeracy-Konzepts bezogen werden: Items, die allgemeine ökonomische Vorstellungen bezogen auf Alltagskontexte zum Inhalt haben, sind für Lernende mit kaufmännischer Vorbildung schwieriger als für Lernende ohne kaufmännische Vorbildung – vice versa für Items, die starke curriculare Bezüge aufweisen. Das Item 1 steht exemplarisch dafür, dass in Anlehnung an das Konzept der *economic literacy* insbesondere text- und bildsprachliche kaufmännische Aufgaben mit konkreter Alltagsverknüpfung bearbeitet werden.

Welche Gründe für den Gewinneinbruch bei der Volkswagen AG erkennen Sie in dem Zeitungsartikel? (Item 1)

Im Item 1 sollen die Lernenden aus dem Zeitungsartikel Gründe und Zusammenhänge herausfiltern, die den Gewinneinbruch bei Volkswagen erklären. Für dieses Item beträgt die Logit-Differenz der Itemschwierigkeit 0,576 und kennzeichnet damit einen mittleren DIF-Effekt. Item 16 ist hingegen ein Beispiel für ein stark am Curriculum ausgerichtetes Item. Die Lernenden sollen diskutieren,

warum eine fristgerechte Lieferung für das produzierende Unternehmen von entscheidender Bedeutung ist.

Das Volkswagenwerk in Wolfsburg erhält vom Land Thüringen den Auftrag, möglichst kurzfristig 2.800 VW Passat zu liefern. Die Pkw sollen den Fuhrpark der Polizei an einigen zentralen Standorten erweitern. Warum ist es für die Volkswagen AG wichtig, kurzfristig liefern zu können? (Item 16)

Das Item erfasst das konzeptuale Verständnis der Lernenden im Hinblick auf die Innen- und Außenbeziehungen eines Unternehmens im Rahmen der Auftragsabwicklung. Dieser Lerninhalt wird im kaufmännischen Eingangsunterricht intensiv vermittelt, so dass diese Aufgabenstellung Lernende mit kaufmännischer Vorbildung systematisch bevorteilt. Die Itemschwierigkeit unterscheidet sich zwischen den Subgruppen um 1,548 Logits; der DIF-Effekt für dieses Item ist groß.

Die Skala der economic numeracy enthält ein Item mit mittlerem und ein Item mit großem DIF (vgl. Abbildung 5.1.20), die inhaltlich in Bezug auf das Literacy/Numeracy-Konzept interpretiert werden können.

Abbildung 5.1.20 DIF-Analyse der Skala economic numeracy (Item*Vorbildung)

Item	Schätzer	Standardfehler	Chi-Square (df)	p-Wert
Item 26	0,295	0,135	11,27 (20)	0,046
Item 27	-0,335	0,166		

Item 26 zielt im Sinne des Literacy/Numeracy-Konzepts auf die Übersetzung einer bildsprachlichen Information in eine Rechenoperation. Die Aufgabe lautet wie folgt:

Sie sind Aktionär der Volkswagen AG. Der Wert Ihrer Volkswagen-Aktien betrug am 31.12.2003 17.500 Euro. Verwenden Sie für die folgenden Berechnungen bitte die Angaben aus der Grafik des Zeitungsartikels.

Berechnen Sie bitte Ihren Verlust im Vergleich zu einer Investition in den „DAX 30“. Wie hoch ist Ihr Verlust in Prozent? (Item 26)



Zur Lösung dieses Items ist kein domänenspezifisches Wissen notwendig. Die Graphik der Kursentwicklungen ist zu interpretieren. Hierfür sind die notwendigen Daten herauszusuchen und in eine Gleichung zu bringen. Für dieses Item beträgt die Differenz der Itemschwierigkeiten zwischen den Subgruppen 0,590 Logits. Um diesen Wert ist das Item für Lernende mit kaufmännischer Vorbildung schwieriger als für Lernende ohne kaufmännische Vorbildung. Das Item 27 weist hingegen für Lernende mit kaufmännischer Vorbildung im Vergleich zu Lernenden ohne kaufmännische Vorbildung einen um 0,670 Logits kleineren Schätzer der Itemschwierigkeit auf.

Das Volkswagenwerk in Wolfsburg erhält vom Land Thüringen den Auftrag, möglichst kurzfristig 2.800 VW Passat zu liefern. Die Pkw sollen den Fuhrpark der Polizei an einigen zentralen Standorten erweitern. Bevor die 2800 VW Passat an die Polizei des Landes Thüringen ausgeliefert werden können, müssen die Pkw mit den für Polizeiwagen typischen grünen Folienstreifen beklebt werden. Es gibt zwei Zulieferer dieser Folien, die zu den unten genannten fixen Bruttopreisen zu liefern bereit sind. Der Bedarf an Folie pro Wagen beträgt 7 m Folie am Stück. Der Zulieferer A liefert für 14,11 Euro pro laufenden Meter (14,11€/lfm) in 1.000 Meter Rollen. Der Zulieferer B liefert für 14,29 Euro pro laufendem Meter (14,29€/lfm) in 140 Meter Rollen. Ermitteln Sie für diesen Auftrag das preisgünstigere Angebot! (Item 27)

Auch zur Lösung dieses Items sind ausschließlich mathematische Grundkenntnisse und kein domänenspezifisches Wissen notwendig. Die Angebotsermittlung ist jedoch in den Curricula der kaufmännischen Bildungsgänge verankert, so dass die Lernenden mit den Berechnungen und der Art der Fragestellung vertraut sind – ein Vorteil, der die Itemschwierigkeit im Vergleich zur Referenzgruppe reduziert.

In der Assessmentpraxis wird angestrebt, Items zu konstruieren, die möglichst keinen DIF aufweisen, um eine „faire“ Messung zu gewährleisten. Im Bereich der beruflichen Bildung ist jedoch die Heterogenität der Schülerschaft von besonderer Bedeutung – in Dualen Ausbildungsgängen noch deutlicher als im vollzeitschulischen Bereich –, so dass Items gefunden werden müssen, die den unterschiedlichen Subgruppen gerecht werden. Vor allem im Hinblick auf die unterschiedlichen Bildungsbiographien und beruflichen Vorerfahrungen stellt sich für Assessments im beruflichen Bereich die Herausforderung, Tests zu konstruieren, die für jeden Probanden zu validen Ergebnissen führen. Die Re-Analysen des kaufmännischen Vorwissenstests zeigen hierfür eine Möglichkeit auf, indem die Heterogenität der Lernenden zum Gegenstand der Messung gemacht wird. Für den beruflichen Bereich kann folglich eine Methode der Testentwicklung diejenige sein, im Rahmen derer Items für mögliche Subgruppen kombiniert werden. Zur Erfassung von Vorwissensleistungen scheint auf Basis der Stichprobe der

Fachgymnasiasten eine Kombination aus curricular validen Items und Items, die sich am Literacy/Numeracy-Konzept orientieren, zielführend zu sein. Mit einem solchen Item-Mix liegen sowohl Herausforderungen vor, die auch ohne spezifische Vorbildung lösbar sind, als auch solche, bei deren Lösung spezifische kaufmännische Konzepte und Arbeitstechniken vorteilhaft eingebracht werden können, jedoch nicht eingebracht werden müssen.

Insgesamt macht die Re-Analyse des kaufmännischen Vorwissenstests am Beispiel des Ausbildungsganges am Fachgymnasium Wirtschaft deutlich, dass (1) unterschiedliche Fähigkeitsstrukturen den Erwerb von domänenspezifischen Wissen sowie die domänenspezifische Performanz beeinflussen, dass (2) kaufmännische Kompetenz zumindest über zwei Kompetenzdimensionen beschrieben werden kann, die unterschiedlich systematische Zugänge zur kaufmännischen Domäne aufzeigen, und dass (3) Instrumentenentwicklung auf einem heuristisch begründetem Modell basieren sollte, um zu einer optimalen Skalierung der Items zu gelangen.

5.2 Studie II: Kaufmännische Kompetenz in Ausbildungsgängen des Dualen Systems

Während es sich bei der Studie an den niedersächsischen Fachgymnasien Wirtschaft um die Re-Analyse eines Tests mittels Item Response-Theorie handelt, wurde die Erhebung im Rahmen des BMBF-Projekts „Konstruktvalidität von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz“ unter Rückgriff auf Assessmentmodelle (vgl. Kapitel 3.1) und die Leitidee des „Measuring-Prozesses“ entwickelt, durchgeführt und evaluiert. Measuring beschreibt dabei eine zentrale Einstellung des Testers zum Test, die sich (1) auf die theoretisch konzeptionelle Herleitung des zu messenden Konstrukts, (2) auf das theoretisch begründete Antwortverhalten der Probanden in einem gewählten Testformat unter Berücksichtigung der Kriterien Authentizität und Validität, (3) auf die Konstruktion des Testformats und der Items und (4) letztlich auf die Auswertung und Analyse der Testergebnisse bezieht, wobei das Vorgehen auf sämtlichen Prozessstufen durch ein zuvor festgelegtes Messmodell geleitet wird (u. a. Wilson, 2005; 2008). Diese Entwicklungsschritte wurden insbesondere vor dem Hintergrund gegangen, Testformate und Testmethoden zu prüfen, die sich für ein international vergleichendes Large-Scale Assessment als tragfähig erweisen. Das Ziel des Projekts war es, berufliche Handlungskompetenz in ihren unterschiedlichen Facetten zu erfassen und dabei möglichst realitätsbezogene Arbeitsanforderungen sowie ein als realistisch geltendes Testsetting als Grundlage der Testung zu konstruieren. Es geht folglich um die Frage, welche Möglichkeiten einer tätigkeitsbezogenen Erfassung beruflicher Handlungskompetenz gegeben sind. Generell lassen sich in diesem Zusammenhang zwei Methoden unterscheiden, die vielfältige Ausprägungen aufweisen können (vgl. u. a. Sonntag & Schaper, 1997; Höft & Funke, 2001; Rauner, Haasler, Heinemann & Grollmann, 2009):

- Die Fremd- und/oder Selbsteinschätzung des Arbeitsverhaltens durch fragebogenbasierte bzw. raterbasierte Verfahren sowie
- simulationsbasierte Verfahren, in denen sich durch Wissens- und Performanzmessungen das Arbeitsverhalten beobachten und diagnostizieren lässt.

Die Feasibility Study zur Durchführung eines international vergleichenden Large-Scale Assessments in der beruflichen Bildung („Berufsbildungs-PISA“ oder VET-LSA) setzt auf simulationsbasierte Verfahren und folgt damit den aktuellen Befunden und Empfehlungen großer internationaler Vergleichsstudien. Verfahren der Fremd- und Selbsteinschätzungen können für simulationsbasierte Erhebungen insbesondere zur Validierung eine sinnvolle Ergänzung sein. Im Hinblick auf die simulationsbasierten Verfahren werden vor allem in der politischen Diskussion Unterschiede zwischen Arbeitsproben – also der Messungen des Arbeitsverhaltens am Arbeitsplatz – und betrieblichen Simulationen – also der Messung des Arbeitsverhaltens am Computer – betont. Bei dieser Unterscheidung, die vor allem auf berufspädagogischen Analysen in der Fertigung beruht, wird jedoch außer Acht gelassen, dass es sich bei beiden Verfahren um standardisierte Aufgaben zur Beobachtung und Diagnose des Arbeitsverhaltens handelt und damit beide Verfahren folglich Simulationen einer rekonstruierten Realität darstellen. Dies bedeutet, dass die Verfahren mehr oder weniger authentisch und mehr oder weniger komplex ausgestaltet sein können. Der Authentizitätsgehalt sowie der Komplexitätsgrad sind abhängig von der konstruierten Anforderungssituation und nicht von dem Verfahren selbst. Dies gilt sowohl für Erhebungen im gewerblich-technischen Bereich (vgl. Nickolaus, Gschwendtner & Abele, 2009) als auch für Erhebungen im kaufmännischen Bereich (vgl. Winther & Achtenhagen, 2008b; Achtenhagen & Winther, 2009; Seeber, 2008).

Die nachfolgend skizzierte Studie bezieht sich auf die Entwicklung und Auswertung beruflich authentischer Anforderungssituationen im Ausbildungsberuf Industriekaufmann/Industriekauffrau. In dem simulationsbasierten Verfahren tätigkeitsspezifischer Kompetenzmessung werden mit Hilfe der betrieblichen Unternehmenssimulation ALUSIM Arbeits- und Geschäftsprozesse von Industrieunternehmen visualisiert. Über die Testplattform ALUSIM lassen sich die von den Auszubildenden zu bearbeitenden authentischen betrieblichen Aufgaben einschließlich der für die Bearbeitung notwendigen Handlungs- und Kommunikationsabläufe darstellen. Über die Computersimulation lässt sich sicherstellen, dass (1) konkrete Arbeitsschritte und -prozesse beobachtet werden können und (2) Arbeitsprodukte als standardisierte Ergebnisse der Arbeitsprozesse für die Analyse der Kompetenzen der Auszubildenden zur Verfügung stehen.

5.2.1 *Measuring-Prozessschritt I: Definition und Umfang berufsfachlicher Kompetenz*

Ausgangspunkt des Projekts waren Diskussionen, die sich auf die Machbarkeitsstudie zum Berufsbildungs-PISA bezogen (vgl. Baethge, Achtenhagen, Arends, Babic, Baethge-Kinsky & Weber, 2006). Obwohl dort die Begründungen für die Operationalisierung des Konzepts beruflicher Handlungskompetenz detailliert vorge-

stellt sind und zugleich gegenüber den Ansätzen eines EQF abgegrenzt werden, wurde vielfach die Befürchtung geäußert, dass in einem solchen VET-LSA die typische Leistung des deutschen Systems einer Dualen Berufsausbildung nicht hinreichend zur Geltung käme: die Herausforderung einer beruflichen Handlungskompetenz in den verschiedenen beruflichen Fachgebieten. Die Herausforderung für das Projekt lag folglich darin, über die Wahl der Testinhalte und Testformate sicherzustellen, dass das Spezifikum der deutschen Dualen Ausbildung: nämlich das Befähigen zu selbstständigen und selbstverantworteten Handlungen am Arbeitsplatz, kurz: berufliche Handlungskompetenz, zentral zur Geltung kommt. Das Konzept der beruflichen Handlungskompetenz (vgl. Kapitel 2.4.1) wurde mit den Dimensionen Selbst-, Sach-, Methoden- und Sozialkompetenz als Grundstruktur eines beruflichen Kompetenzgefüges in der Machbarkeitsstudie aufgegriffen. Für die Messung beruflicher Handlungskompetenz wurde jedoch vorgeschlagen, den Bereich der Selbstkompetenz getrennt mit Hilfe bereits existierender Messinstrumente zu erheben und auch den Bereich der Sozialkompetenz gesondert zu erfassen. Für den Bereich der Sachkompetenz wurde in der Machbarkeitsstudie als übergreifende Kennzeichnung „Umgang mit Konzepten“ gewählt. Zur Unterscheidung ist der Bereich der Methodenkompetenz über die Kennzeichnung „Umgang mit technischer Ausstattung“ beschrieben (Baethge, Achtenhagen, Arends, Babic, Baethge-Kinsky & Weber, 2006, S. 45ff.). Mit dieser Unterscheidung liegt die in der bildungspolitischen Diskussion häufig vorgenommene Trennung von Wissen einerseits und Fertigkeiten/Fähigkeiten andererseits nahe. Auch der EQF macht von den Dimensionsüberschriften her eine solche Unterscheidung. Aus messtheoretischer Perspektive ist in diesem Zusammenhang eine besondere Schwierigkeit jedoch darin zu sehen, dass die Vierteilung des übergreifenden Modells beruflicher Handlungskompetenz in seiner Idealtypik plausibel und auch sorgfältig pädagogisch-anthropologisch sowie didaktisch begründet ist, es zugleich aber äußerst schwierig, wenn nicht unmöglich scheint, die einzelnen Bereiche bei dem gegenwärtigen Forschungsstand angemessen zu operationalisieren.

Vor diesem Hintergrund wurde das Konzept der beruflichen Handlungskompetenz im Rahmen des Projekts inhaltlich reduziert, indem (1) die Vierteilung der Dimensionen aufgelöst und (2) zwischen kognitiven Verstehens- und Handlungskomponenten beruflicher Aktivität differenziert wird. Zur Unterscheidung zwischen den Konzepten wird für das Projekt die Notation *„berufsfachliche Kompetenz“* gewählt.

5.2.1.1 Adaptation des Kompetenzstrukturmodells

Berufsfachliche Kompetenz wird als primär tätigkeitsbezogene Kompetenz in beruflichen Anforderungssituationen definiert, die sich über Verstehens- und Handlungskomponenten eines typischen Arbeitsprozesses beschreiben lässt. Auf Basis eines spezifischen Arbeitsprozesses ist es damit auch möglich, Aspekte sozialer Kompetenz – in dem Umfang, wie sie zur Bewältigung des Arbeitsprozesses not-

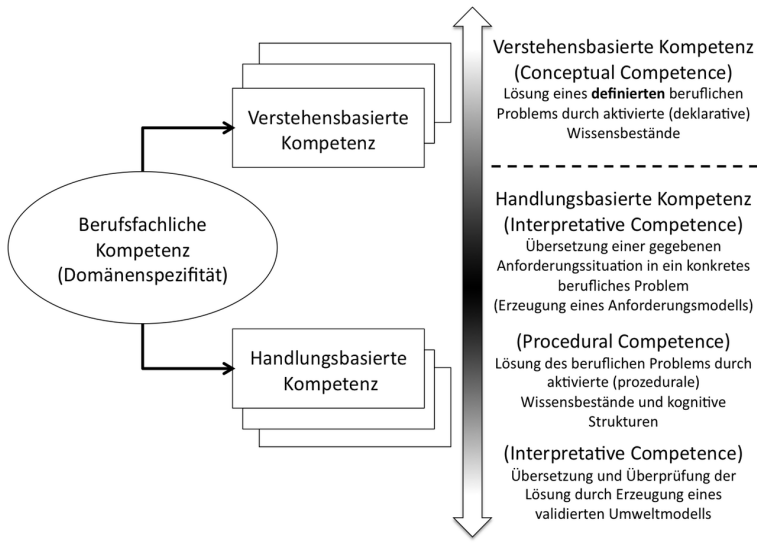
wendig sind – mitzuerfassen, ohne eine empirische Trennbarkeit der Dimensionen voraussetzen zu müssen.

Unter Rückgriff auf das Kompetenzstrukturmodell für die kaufmännische Bildung (Kapitel 2.4.2; vgl. Winther & Achtenhagen, 2008b; 2009) bezieht sich eine zentrale inhaltliche Reduzierung darauf, dass berufsfachliche Kompetenz als domänenspezifische Kompetenz über Geschäftsvorfälle erfasst wird und per definitionem die domänenverbundenen Kompetenzbereiche unberücksichtigt bleiben. Mit diesen Einschränkungen ist berufsfachliche Kompetenz als *Subkategorie der kaufmännischen Kompetenz*, die über das Kompetenzstrukturmodell definiert ist, zu interpretieren. Kaufmännische Kompetenz bezieht sich auf das Wechselspiel von domänenverbundenen und domänenspezifischen Kompetenzbereichen, auf dessen Basis Auszubildende über die Fähigkeit verfügen, in berufsrealen Situationen unternehmerische Entscheidungen treffen und diese validieren zu können, um damit das eigene Wissens- und Handlungspotential vor dem Hintergrund der Entwicklung individueller beruflicher Regulationsfähigkeit auszubauen (vgl. Winther & Achtenhagen, 2008b).

Das Kompetenzstrukturmodell geht dabei im Kern von einer differenzierten Struktur kaufmännischer Kompetenz aus, die sich sowohl auf die betrieblichen und schulischen Inhalte als auch auf die verschiedenen kognitiven Zugriffe, mit denen typische berufliche Inhalte bewältigt werden, beziehen lässt. Damit wird dem Umstand Rechnung getragen, dass Kompetenz in einer Domäne (1) inhaltsbezogen und (2) keine eindimensionale kognitive Disposition ist und es folglich im Hinblick auf Kompetenzerwerb und -entwicklung nicht genügen wird, nur auf eine spezifische kognitive Struktur zu fokussieren (Kilpatrick, Swafford & Findel, 2001; Brand, Hofmeister & Tramm, 2005). Aus dieser Perspektive werden im Modell drei verschiedene Kompetenzbereiche im Sinne einer Handlungsabfolge betrachtet: die Bearbeitung einer definierten konkreten Lern- und Arbeitsanforderungen durch aktivierte deklarative Wissensbestände (conceptual competence) vor dem Hintergrund eines ganzheitlichen und funktionalen Verständnisses zentraler domänenbezogener Ideen, die Selektion und Ausführung domänenbezogener Handlungen und Prozeduren (procedural competence) sowie die kognitive Erfassung der Anforderungssituation und die angemessene Interpretation und Validation der gefundenen Lösung (interpretative competence). Die Kompetenzstruktur bezogen auf die drei inhaltlichen Handlungsbereiche economic literacy, economic numeracy und Geschäftsvorfall ist damit prozessual und nicht hierarchisch. Diese Modellannahme erlaubt es, die einzelnen inhaltlichen Kompetenzbereiche separiert zu analysieren, ohne die Struktur des Modells zu verletzen. Für das Projekt lassen sich die Kompetenzstrukturen folglich ausschließlich auf domänenspezifische Geschäftsvorfälle beziehen (vgl. Abbildung 5.2.1). Wie in Kapitel 2.4.2 erläutert, wird unter einem Geschäftsvorfall eine über konkrete Arbeitsprozesse definierbare Anforderungssituation verstanden, die sich in unternehmensspezifischen Geschäftsprozessen verorten lässt. Diese Annahme entspricht der im Projekt geforderten tätigkeitsbezogenen Kompetenzmessung, da über Geschäftsvorfälle der Aktionsraum der Auszubildenden auf vollständige Ar-

beits- und Geschäftsprozesse in ihrem ganzheitlich zeitlichen und organisationalem Umfang bezogen ist.

Abbildung 5.2.1 Kompetenzstrukturmodell: Berufsfachliche Kompetenz



Die Fokussierung auf berufsfachliche Kompetenz verbunden mit der Gestaltung der Geschäftsvorfälle bedingen zudem eine Änderung der Kompetenzstrukturannahmen: Die drei Kompetenzstrukturen werden den Verstehens- und Handlungskomponenten der berufsfachlichen Kompetenz zugeordnet (vgl. Abbildung 5.2.1). Dieser Schritt ist notwendig, da jeder Geschäftsvorfall mit einer handlungsorientierten Bearbeitung der Anforderungssituation beginnt und mit einer verstehensbasierten Bearbeitung abschließt. Für die handlungsorientierte Bearbeitung stehen prozessuale Simulationsaufgaben zur Verfügung. Jeder Geschäftsvorfall ist zunächst durch die Auszubildenden in ein Anforderungsmodell zu übersetzen (interpretative competence), um anschließend vorrangig prozedural gelöst zu werden (procedural competence). Das in den Simulationsaufgaben erzeugte Arbeitsprodukt umfasst dabei Aspekte der Lösungsvalidierung (interpretative competence). Diese prozessuale Kombination aus „procedural competence“ und „interpretative competence“ wird als *handlungsbasierte Kompetenz* definiert. Hierunter wird verstanden, dass domänenspezifische Handlungen und Prozeduren flexibel, fehlerfrei, effizient und angemessen im Hinblick auf die Anforderungssituation selektiert, ausgeführt und bewertet werden können (vgl. hierzu auch die Ausführungen zu „procedural fluency“; u. a. Hayes, 2006).

Verstehensbasierte Kompetenz bezieht sich hingegen nicht auf prozessuale Strukturannahmen, da Interpretation, Lösung und Validierung des Geschäftsvor-

falls aus der handlungsorientierten Bearbeitung bereits vorliegen. Über konzeptuale Anwendungsaufgaben werden in diesem Kompetenzbereich die Fähigkeiten der Auszubildenden erfasst, eine zuvor bereits definierte berufliche Problemstellung durch aktivierte deklarative Wissensbestände zu lösen. Verstehensbasierte Kompetenz beschreibt folglich das ganzheitliche und funktionale Verständnis zentraler domänenspezifischer Ideen – und umfasst damit mehr als das isolierte Wissen von Fakten und Methoden. Für die Testentwicklung und die Testanalyse haben die Veränderungen im Kompetenzstrukturmodell der berufsfachlichen Kompetenz im Vergleich zum Idealtypus folgende Auswirkungen:

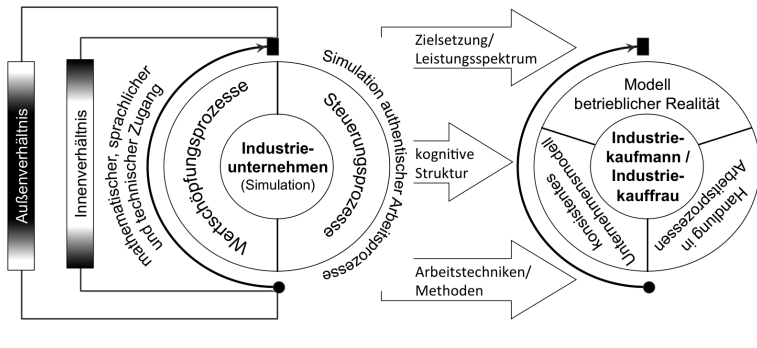
- Ausgangspunkt der Testentwicklung sind komplexe, an der Arbeitswirklichkeit orientierte domänenspezifische Geschäftsvorfälle.
- Der Anforderungsgrad der Geschäftsvorfälle ist so zu wählen, dass zu ihrer Bearbeitung domänenspezifische Fachkenntnisse vorauszusetzen sind.
- Die Geschäftsvorfälle sind so zu konstruieren, dass sich mit ihnen zwei unterschiedliche Fähigkeitsstrukturen abbilden lassen: verstehensbasierte und handlungsbasierte Kompetenz.
- Die Abbildung verstehensbasierter Kompetenz erfordert Testitems, die vorrangig konzeptuale bzw. deklarative Wissensbestände ansprechen.
- Die Abbildung handlungsbasierter Kompetenz erfordert Testitems, die den prozessualen Charakter betrieblicher Aktivität ansprechen. Hierzu sind Testitems gefordert, die vorrangig auf prozedurale Wissensbestände abstellen und in denen die Arbeitsprodukte einen Validierungsprozess durchlaufen müssen.
- Die Modellprüfung erfolgt auf Basis eines mehrdimensionalen Rasch-Modells.

5.2.1.2 Adaptation des Domänenmodells

Das BMBF-Projekt „Konstruktvalidität von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz“ ist eine von drei Vorstudien vor dem Hintergrund der Entwicklungen eines internationalen Vergleichs in der europäischen Berufsbildung. Neben der Studie im kaufmännischen Bereich wurden von Nickolaus und Kollegen (Nickolaus, Gschwendner & Abele, 2009) Verfahren der Kompetenzmessung im KFZ-Bereich untersucht und unter der Leitung von Baethge und Arends (2009) eine Studie durchgeführt, in der geprüft wurde, ob in den für eine Vergleichsstudie zunächst ausgewählten Berufen ein internationaler Vergleich möglich ist. Alle drei Studien kommen übereinstimmend zu der Einschätzung, dass sowohl aus formaler als auch aus methodischer Perspektive ein Large-Scale Assessment durchführbar ist. Diesen Vorstudien sind intensive internationale Gespräche vorausgegangen, so dass sich die Inhalte der Messung berufsfachlicher Kompetenz in den deutschen Vorstudien bereits auf Themen und Aufgabenstellungen beziehen, bei denen in der internationalen Diskussion ein Konsens erzielt werden konnte. Für den hier dargestellten kaufmännischen Bereich ergeben sich aus dieser Diskussion Konsequenzen im Hinblick auf das in

Kapitel 3.2.3 vorgestellte Domänenmodell. Das auf die Fragestellungen des Projekts adaptierte Domänenmodell hat die in Abbildung 5.2.2 dargestellte Form:

Abbildung 5.2.2 Domänenmodell: Inhalte berufsfachlicher Kompetenz



Das adaptierte Domänenmodell dient als inhaltlicher Rahmen für die Messung berufsfachlicher Kompetenz bei Auszubildenden im dualen Ausbildungsgang Industriekaufmann/Industriekauffrau. Aus inhaltlicher Perspektive weist das adaptierte Modell eine Reduzierung der internen Unternehmensprozesse auf; aus organisationaler Perspektive führt die zentrale Ausrichtung auf simulationsbasierte Verfahren der tätigkeitsbezogenen Kompetenzmessung zu Veränderungen im Vergleich zum Originalmodell (Abbildung 3.10 in Kapitel 3.2.3). Es zeichnet sich ab, dass Inhalte, die sich auf betriebliche Unterstützungsprozesse beziehen lassen, im internationalen Vergleich sehr unterschiedlich ausgelegt werden. Dies betrifft insbesondere rechtliche Fragestellungen sowie einzelne Bereiche des Rechnungswesen und des Controlling. Aus diesem Grund konzentrieren sich die Messinhalte im Rahmen des Projekts auf betriebliche Wertschöpfungs- und Steuerungsprozesse, wobei weite Aspekte des Rechnungswesen unberücksichtigt bleiben. Die Auswahl der Inhalte im Detail orientiert sich (1) an dem Rahmenlehrplan zur Ausbildung von Industriekaufleuten und (2) an Beobachtungen des Arbeitsplatzes (vgl. hierzu Kapitel 5.2.2).

Die Adaptation auf organisationaler Ebene ist dadurch geleitet, dass das Testformat der Studie vorgegeben war: Die Auszubildenden sollten mit authentischen Simulationsaufgaben getestet werden. Diese sollten so konstruiert sein, dass es mit ihnen möglich ist, berufliche Handlungen vor dem Hintergrund spezifischer Arbeits- und Geschäftsprozesse zu erfassen. Die Erfahrungen mit multidimensionalen Modellunternehmen zeigen, dass Authentizität in einem hohen Maße davon abhängt, wie umfangreich das Modell eines Unternehmens konstruiert ist (vgl. hierzu die Modellierungsschritte des Modellunternehmens Kettenfabrik A & S GmbH; Siemon, 2003; Getsch & Preiß, 2003a; 2003b). Ein weiteres zentrales Ergebnis der Arbeiten zur Entwicklung komplexer Lehr-Lern-Arrangements ist,

dass Anforderungssituationen und Assessments umso anschlussfähiger und fachlich präziser formuliert werden können, je detailgetreuer und konsistenter der Basisdatensatz des Modells konstruiert ist. Die zentrale Herausforderung bestand folglich darin, ein Industrieunternehmen als Modell möglichst komplex und parallel zu den zu testenden Inhalten zu entwickeln. Im Hinblick auf das Domänenmodell ergeben sich aus der Simulation eines Industrieunternehmens folgende Konsequenzen:

- Der Stimulus der Wissensrepräsentation wird durch ein *Modell betrieblicher Realität* gesetzt. Hierfür ist betriebliche Realität in miniature nachzubilden. Um dies leisten zu können, ist es notwendig, die Simulation anhand eines real existierenden Paten durchzuführen. Für das Projekt konnte die Novelis Deutschland GmbH, Werk Göttingen gewonnen werden. Die Wahl des Unternehmens war durch drei Kriterien bestimmt: Es sollte (1) ein Unternehmen gefunden werden, dessen Produktion eine sehr geringe Stückliste aufweist und dessen Produktionsprozesse transparent und technisch leicht beschreibbar sind. Das Werk Göttingen konzentriert sich ausschließlich auf Aluminium-Walzerzeugnisse. Hierzu werden Aluminiumplatten gewalzt und teilweise lackiert und bedruckt – Prozesse, die technisch für Industriekaufleute verschiedener Branchen nachvollziehbar sind. Die Endprodukte des Unternehmens sollten (2) einen hohen Bekanntheitsgrad besitzen, um über das Produkt eine stärkere Identifikation der Auszubildenden mit dem Simulationsunternehmen zu erreichen und um die Auszubildenden dabei zu unterstützen, ein mentales Modell des Simulationsunternehmens zu entwickeln. Das Werk Göttingen walzt und lackiert beispielsweise Aluminiumbleche für die Nivea Creme-Dosen und bedruckt Kartuschen und Dosen für die Wella-Haarpflegeserie – Produkte die (international) bekannt sind und deren Nutzen von allen Auszubildenden nachvollzogen werden kann. Das Unternehmen sollte darüber hinaus (3) international agieren. Novelis ist im Besitz der indischen Hindalco Industries Limited, hat Standorte in elf Ländern auf vier Kontinenten und ist mit einem Umsatz von USD 11,2 Milliarden im Jahr 2008 und mit mehr als 12.500 Mitarbeitern der weltgrößte Hersteller für Aluminium-Walzerzeugnisse. Durch die enge Ausrichtung der Simulation an einem real existierenden Unternehmen ist es möglich, authentische Relationen beispielsweise im Hinblick auf Kapitalbedarf, Gewinn, Produktionszeiten bis hin zu Verpackungsgrößen zu erzeugen.
- Wissenserwerbsprozesse in betrieblichen Umgebungen sind vorrangig durch *Handlungen in Arbeitsprozessen* geprägt. Für die Testentwicklung bedeutet dies, dass an der betrieblichen Realität orientierte Assessments, die die Abbildung von Handlungsfähigkeiten in spezifischen Anforderungssituationen zum Ziel haben, sich an dem Prinzip der Minimierung von Fehlern orientieren und eher auf prozedurale Fähigkeiten abstellen (vgl. Kapitel 3.2.3). Im Kompetenzstrukturmodell wird die hierzu korrespondierende kognitive Fähigkeitsstruktur als handlungsbasierte Kompetenz umschrie-

ben. Zur Überprüfung des ökonomischen Verständnisses werden neben handlungsausführenden zusätzlich auch handlungsbeschreibende Anforderungssituationen konstruiert, um die deklarative kognitive Struktur der Auszubildenden in Anlehnung an den Rahmenlehrplan zu testen. Diese Tests korrespondieren mit der im Kompetenzstrukturmodell definierten verstehensbasierten Kompetenz.

- Die äußere Struktur des Assessments ist durch ein *konsistentes Unternehmensmodell* abgebildet. Das Unternehmensmodell stellt den inhaltlichen Rahmen der Anforderungssituationen dar und gibt die Arbeitsmittel und Methoden der Anforderungsbewältigung vor. Aus messtheoretischer Perspektive werden durch die Simulation authentischer Anforderungssituationen die Prozesse und Produkte der Anforderungsbewältigung standardisiert, was einen Vergleich der Leistungsfähigkeiten erst zulässig macht. Durch das Simulationsunternehmen wird sichergestellt, dass alle Auszubildenden unabhängig von der Branche, Größe oder Wirtschaftskraft des Ausbildungsbetriebes über die gleichen betrieblichen Voraussetzungen sowie Arbeits- und Hilfsmittel verfügen und dass die über die Simulation verfügbaren Lösungshinweise standardisiert für alle Auszubildenden zugänglich sind – Maßnahmen, die die Durchführungsobjektivität des Assessments positiv unterstützen.

Die Bestimmung von Definition und Umfang berufsfachlicher Kompetenz auf Basis des Kompetenzstrukturmodells und des Domänenmodells soll den Anspruch sichern, dass Testziel, Testinhalt und zu testende kognitive Fähigkeitsstruktur korrespondieren. Zur Umsetzung des Testinhalts wurden über domänenspezifische Geschäftsvorfälle (1) konzeptuale Anwendungsaufgaben und (2) prozessuale Simulationsaufgaben entworfen, die in eine computerbasierte Testumgebung – ALUSIM – eingebunden sind. Nachfolgend wird dargestellt, wie das Simulationsunternehmen ALUSIM konstruiert wurde, welche Funktionen und Inhalte eingebunden sind und wie die Authentizität der Anforderungssituationen zur Messung berufsfachlicher Kompetenz sichergestellt wurde.

5.2.2 Measuring-Prozessschritt II:

Ausbildungsreale Authentizität und tätigkeitsspezifische Validität

Der zweite Schritt des Measuring-Prozesses befasst sich mit Annahmen darüber, welche Auswirkungen ein gewähltes Testformat auf das Antwortverhalten der Auszubildenden im Test hat und inwieweit über das gezeigte Antwortverhalten Rückschlüsse auf die Fähigkeitsstrukturen und Leistungsfähigkeiten gezogen werden können. Mit der Wahl eines Simulationsunternehmens als Testformat wird implizit vorausgesetzt, dass es durch Simulation authentischer Arbeitsprozesse besser gelingt, Arbeitshandlungen und Arbeitsverhalten abzubilden, als dies mit Testformaten, die auf Papier und Bleistift basieren, möglich wäre (Kriterium der Authentizität). Wesentliche Argumente hierfür sind, dass durch den Einsatz von Simulationen eine verbale Beschreibung der Bearbeitungsschritte einer Anforderungssituation entfällt und dass der „Umgang mit Konzepten“ (Sachkom-

petenz) sowie der „Umgang mit technischer Ausstattung“ (Methodenkompetenz) simultan in der für reale Arbeitsprozesse typischen Komplexität gemessen werden können.

Die zentrale Herausforderung des Projekts war es, ein Testformat zu entwickeln, das den Implikationen des Kompetenzstrukturmodells gerecht wird und mit denen die Inhalte des Domänenmodells aufgegriffen werden: Bei der Auswahl der Testinhalte zur Erhebung des Konstrukts „Berufsfachliche Kompetenz“ wurde davon ausgegangen, dass ein systemisches Verstehen von unternehmensinternen Prozessen und ein entsprechendes Handeln in diesen Prozessen wesentliche Qualifikationsanforderungen für den Bereich der Aus- und Weiterbildung von Industriekaufleuten und damit eine zentrale Bedingung für die Ausformung beruflicher Handlungskompetenz darstellen. Für die angemessene Erfassung berufsfachlicher Kompetenz ist es erforderlich, ein Testformat zu entwerfen, das theoretisch begründet und fachdidaktisch entsprechend ausgestaltet ist; denn nur so ist es möglich, ein systemisches Verstehen von und ein entsprechendes Handeln in Geschäftsprozessen als eine Basiskompetenz von Industriekaufleuten über entsprechende Teilkompetenzen zu definieren und in Graduierungsschritten auszubilden. Hierzu ist es notwendig, Testinhalte, Testitems und empirisches Testmodell parallel zu entwerfen und wechselseitig aufeinander abzustimmen. Das Ziel ist es, berufliche Anforderungssituationen so zu entwickeln, dass es möglich wird, die zu bearbeitenden Aufgaben auf die zugrunde liegenden Arbeits- und Lernprozesse zu beziehen und zugleich Auskunft über die kognitiven Verarbeitungsschritte während des Lösungsprozesses zu gewinnen (Kriterium der tätigkeitsspezifischen Validität). Anders formuliert: Die Konstruktionsschritte erlauben die Modellierung von Aufgaben, die betriebliche Arbeitssituationen abbilden und zugleich als Psychometrik Aussagen über die kognitiven Anforderungen bei ihrer Bearbeitung erlauben. Damit ist die traditionelle Trennung aufgehoben, nach der zunächst versucht wurde, angemessene Testaufgaben zu konstruieren, diese dann den Probanden vorzulegen und erst anhand der Auswertung Rückschlüsse auf die Fähigkeiten der Probanden zu ziehen.

5.2.2.1 Kriterium der Authentizität

In der Machbarkeitsstudie zum Berufsbildungs-PISA (vgl. Baethge, Achtenhagen, Arends, Babic, Baethge-Kinsky & Weber, 2006) wurde herausgestellt, dass es notwendig ist, für die Erfassung von berufsfachlicher Kompetenz tätigkeitsbezogene komplexe Aufgaben einzusetzen. Vor diesem Hintergrund galt es sicherzustellen, dass die Testformate, mit denen die Aufgaben präsentiert werden, als „authentisch“ gelten können (vgl. hierzu Achtenhagen, 2001; Achtenhagen & Weber, 2003). Authentizität im Hinblick auf die Messung berufsfachlicher Kompetenz beinhaltet den Anspruch einer ausbildungsrealen Messung. Die folgenden Ausführungen verdeutlichen, wie versucht wurde, eine größtmögliche Nähe zur beruflichen Ausbildungsrealität zu wahren. Im Vorgriff kann bereits hier darauf hingewiesen werden, dass berufliche Handlungskompetenz immer im Hinblick auf eine berufstypische Aufgabenbewältigung definiert ist und entsprechend

auch in den Zwischen- und Abschlussprüfungen zu ermitteln versucht wird. Ausbildungsrealität bedeutet, dass der ausschließliche Rückgriff und die damit verbundene Einschränkung auf arbeitsplatz- und betriebspezifische Anforderungen die übergeordneten Ziele der Berufsbildung nur unzureichend berücksichtigen würde: nicht nur im Sinne einer Bereitstellung von ausreichenden Humanressourcen für die Gesellschaft, sondern auch im Sinne einer breit angelegten Persönlichkeitsentwicklung.

Für die Formulierung von Testinhalten sowie für die Entwicklung des Testformats wurden aus diesem Grund Urteile von Fachleuten eingeholt, die sich sowohl auf den Erfolg am Arbeitsplatz – als Erbringung der Arbeitsleistung – als auch auf den Erfolg der Ausbildungsbemühungen zur Ermöglichung des Erfolgs am Arbeitsplatz bezogen (schulische Initiative). In der neueren Lerntheorie, vor allem im Bereich des „Situating Learning“ (vgl. Resnick, 1987; Greeno, 1998; Byrnes, 2008), wird darauf verwiesen, dass es darauf ankomme, Lernprozesse zu gestalten, „wie sie im Leben zu finden sind“. Dabei wird dem Lernen in der Arbeitswelt per se Authentizität zugesprochen (vgl. Achtenhagen & Weber, 2003, S. 187). Eine entscheidende Einsicht, die aus der Kritik an der „Situating Learning“-Bewegung resultiert, ist jedoch, dass „Authentizität“ nicht per se existiert, sondern jeweils inszeniert werden muss. Für die Formulierung von Testinhalten und die Entwicklung des Testformats bedeutet dies, dass, wenn berufsfachliche Kompetenz gemessen werden soll, die Anforderungssituationen, in denen sie zu zeigen ist, modelliert werden müssen. Dies bedeutet, dass Authentizität gekoppelt ist an berufstypische Situationen bzw. Handlungen, in denen eine berufstypische Leistung gezeigt wird. Authentizität wird folglich in Anforderungssituationen umgesetzt, indem die Spezifität der Situation betont wird: Charakterisierungen von Kunden oder Lieferanten, Darstellung von Preisen sowie Lieferkonditionen etc. als Details einer Inszenierung.

Doyle (2000) hat verschiedene Ansätze zur Bestimmung und Realisierung von „Authentizität“ einer tiefgreifenden Kritik unterzogen und schlägt vor, drei Dimensionen von Authentizität zu unterscheiden:

- Lernerbezogene Authentizität: Hier geht es um das Anknüpfen an das Vorwissen und die spontanen Interessen der Lernenden.
- Fach- und inhaltsbezogene Authentizität: Hier handelt es sich um reale Daten, Instrumente und Operationen einer Fachdisziplin, wie sie im Curriculum repräsentiert sind; ermöglicht werden soll ein forschendes Lernen mit realen Gegenständen.
- Situationsbezogene Authentizität: Hier ist die Partizipation in realen Situationen einer „Community of Practice“ angesprochen, d. h. es geht um ein Lernen, mit dem eine Enkulturation in eine Expertenkultur angestrebt wird.

Die Diskussion um unterschiedliche Authentizitätsdimensionen zeigt, dass berufliche Handlungskompetenz nicht allein aus der Erfahrung am Arbeitsplatz entstehen kann und dass auch im Dualen System die „Realität“ der Berufserfahrung nicht von allein gegeben ist, sondern schrittweise vermittelt werden muss. „Realität“

tät“ wird folglich erst über Übersetzungsleistungen erschlossen (vgl. die Diskussionen zur Repräsentation von Wissen und Erfahrungen; Markmann, 1999; Mislevy, 2007).

Für den theoretisch angenommenen Zusammenhang zwischen Testformat, Testinhalt und Antwortverhalten lässt sich schlussfolgern, dass je umfassender lerner-, fach- und situationsbezogene Authentizität inszeniert wird, desto präziser kann die Ausbildungsrealität abgebildet werden. Damit steigt die Wahrscheinlichkeit, Anforderungssituationen modelliert zu haben, in denen die Auszubildenden berufsfachliche Kompetenz präsentieren können.

5.2.2.2 Kriterium der tätigkeitsspezifischen Validität

Ein weiteres zentrales Kriterium für die Entwicklung von Testformat und Testinhalt ist die Validität. Für dieses Projekt und im Hinblick auf die Durchführung eines international vergleichenden Large-Scale Assessments in der beruflichen Bildung war es nicht das Ziel, curriculare Validität herzustellen. Gerade im internationalen Vergleich der Bildungssysteme wird deutlich, dass die inhaltlichen und organisatorischen Strukturen der Ausbildungsgänge, die auf eine vergleichbare Berufsgruppe zielen, sich voneinander unterscheiden. Als Folge ergeben sich verschiedene Zuschnitte für Berufe, für die allerdings im internationalen Vergleich Tätigkeitscluster identifiziert werden können, die als tätigkeitsbezogenes Validitätskriterium nutzbar sind. In der KOMET-Studie von Rauner und Kollegen wird in diesem Zusammenhang beispielsweise von beruflicher Validität gesprochen (Berufliche Kompetenzen und berufliche Identität von Auszubildenden in Elektroberufen: Eine „Large-Scale“-Untersuchung an berufsbildenden Schulen in Hessen und Bremen (KOMET); Rauner, Haasler, Heinemann & Grollmann, 2009). Allerdings zeigt die Feasibility Study VET-LSA für den Vergleich unterschiedlicher nationaler Bildungsgänge, dass sich eine Übereinstimmung nicht für Berufe, wohl aber für berufliche Tätigkeiten erzielen lässt (Baethge & Arends, 2009; für den kaufmännischen Bereich vgl. Breuer, Hillen & Winther, 2009).

Für das Projekt wurden tätigkeitsspezifische Validität und Authentizität durch folgende, miteinander kombinierte Maßnahmen gesichert:

- Es wurden Beobachtungen an den Arbeitsplätzen für Industriekaufleute durchgeführt. Hierzu wurden zum einen strukturierte Interviews eingesetzt. Zum anderen wurde den Interviewpartner ein Fragebogen vorgelegt, mit dem allgemeine berufliche Fähigkeiten („abilities“) bewertet werden sollten. Dieser Fragebogen wurde auf Basis der Empfehlungen des Occupational Information Network (O*Net) konstruiert (McCloy, Waugh & Medsker, 1996; Peterson, Mumford, Borman, Jeanneret & Fleishman, 1998; U.S. Department of Labor Employment and Training Administration, 2006).
- Das Testformat wird als Simulation eines realen Industrieunternehmens umgesetzt. Die Simulation beinhaltet authentisch modellierte Arbeitsmittel (Betriebssoftware, Datenbanken, Kataloge etc.) und berücksichtigt Arbeitsprozesse im Hinblick auf ihre strukturelle und zeitliche Ausgestaltung.

In dem als Paten für die Simulation fungierenden Industrieunternehmen werden ausgebildete Industriekaufleute in verschiedenen Abteilungen eingesetzt – im Einkauf und Vertrieb, in der Personalverwaltung und der Kreditorenbuchhaltung sowie im Rahmen des internen Führungskräfteprogramms. Vor dem Hintergrund des Domänenmodells sind insbesondere Tätigkeitsbeschreibungen aus den Bereichen der betrieblichen Wertschöpfung und Steuerung relevant; die Beobachtungen am Arbeitsplatz beziehen sich folglich schwerpunktmäßig auf Beschaffungs- und Vertriebsvorgänge sowie auf mit der Wertschöpfung verbundene Steuerungsfragen.

Insgesamt wurden elf Industriekaufleute interviewt. Das Interview bestand aus drei thematischen Bereichen: (1) Allgemeine Fragen zur Ausbildung und dem Karriereverlauf; (2) detaillierte Fragen zur Beschreibung des konkreten Arbeitsplatzes und (3) kritische Reflexion und Bearbeitung einer simulierten Anforderungssituation, die den Arbeitsplatz und die Aufgaben des Mitarbeiters möglichst authentisch abbilden soll. Von den Befragten haben alle in dem Unternehmen ihre Ausbildung zum Industriekaufmann/zur Industriekauffrau absolviert; drei von ihnen im Rahmen eines Berufsakademiestudiums (Abschluss: Industriekaufmann (BA)). Es wurden vier Beschäftigte befragt, die im Einkauf des Unternehmens tätig sind, fünf, die Tätigkeiten im Vertriebsbereich wahrnehmen, und zwei, die derzeit im internen Führungskräfteprogramm auf Aufgaben im mittleren Management vorbereitet werden.

Als gemeinsames Ergebnis ist für die Bereiche Einkauf und Vertrieb eine sehr stark ausgeprägte Serviceorientierung festzustellen. Zwei Drittel der Arbeitszeit werden darauf verwandt, Kundenanfragen zu bearbeiten, Änderungen für Aufträge zu kalkulieren und die Erfüllung von Aufträgen zu avisieren und zu prüfen. Für den Bereich des mittleren Managements sind hingegen Fragen der Produktionseffizienz, der Kostenreduktion und Preisgestaltung sowie Aspekte des Lean Management und der Lean Production bedeutsam. Die Beschreibung des konkreten Arbeitsplatzes und der Arbeitsaufgaben orientierte sich an drei von den Beschäftigten als für ihren Arbeitsalltag typisch angesehenen Tätigkeiten. Für jede dieser Tätigkeiten war im Detail festzuhalten,

- wie oft die Tätigkeit ausgeführt wird,
- welcher Zeitaufwand zur Ausführung der Tätigkeit notwendig ist,
- in welche Arbeitsschritte sich die Tätigkeit untergliedern lässt,
- welche Arbeitsmittel eingesetzt werden,
- wie die Tätigkeit bzw. die Ergebnisse der Tätigkeit protokolliert werden und
- welche Abteilungen von der Tätigkeit direkt/und indirekt betroffen sind.

Darüber hinaus wurden Belege gesammelt, mit deren Hilfe eine konkrete Tätigkeit möglichst exakt nachgebildet werden kann. Hierzu wurden u. a. folgende Fragen gestellt:

- Wie sehen die Arbeitsmittel konkret aus und über welche zentralen Funktionen verfügen die Arbeitsmittel?

- Wie wird innerhalb des Unternehmens kommuniziert und wie erfolgt die Ansprache der Kunden und Lieferanten des Unternehmens?
- Welche Form haben Eingangs- und Ausgangsdokumente?
- Welche Datenbanken gibt es im Unternehmen und wie erfolgt der Zugriff auf diese?

Mit Hilfe dieser Informationen wird es möglich, einen Arbeitsplatz und konkrete Arbeitsaufgaben im Hinblick auf den Beschaffungs- und Vertriebsbereich authentisch zu modellieren. Der Bereich der betrieblichen Wertschöpfung lässt sich damit valide erfassen. Die Darstellung betrieblicher Steuerungsprozesse ist in diesem Zusammenhang etwas komplizierter: In den verschiedenen Einkaufs- und Vertriebsabteilungen des Unternehmens erfolgt zwar regelmäßig ein Reporting der Zahlen der jeweiligen Abteilung als Vorlage für die Geschäftsführung, jedoch werden das Entscheidungshandeln und die Verantwortung für Steuerungsprozesse weitgehend abgegeben. Zusätzlich gilt, dass die auf Ebene des mittleren Managements relevanten Fragen zu einem großen Teil zu umfassend sind, als dass sie von Auszubildenden bearbeitet werden könnten. Auf Basis der Interviews mit denjenigen Industriekaufleuten, die das interne Führungskräfteprogramm des Unternehmens absolvieren, erfolgte eine Einigung auf den Bereich der Arbeitsvorbereitung, und hierbei insbesondere die Maschinenstundensatzkalkulation, und damit auf Fragen der mittelfristigen Preisgestaltung.

Zum Abschluss des Interviews wurden die Befragten mit beruflichen Anforderungssituationen konfrontiert, von denen sie einschätzen sollten, inwieweit diese für ihre Tätigkeit relevant wären. Von besonderer Bedeutung dabei war die Bewertung des Realitätsgehalts. Insgesamt zeigte sich, dass die bereits vor den Interviews ausgearbeiteten beruflichen Anforderungssituationen die Tätigkeiten der Beschäftigten sehr genau abgebildet haben und mit ihrer Formulierung und Intention als für den Arbeitsalltag hoch realistisch bewertet wurden. Die Interviews wurden in diesem Zusammenhang dazu genutzt, Detailfragen der Präsentation der Anforderungssituation zu besprechen. Es wurde festgelegt, welche Informationen und Hilfen Auszubildende nach einer Einarbeitungszeit zur Bearbeitung einer solchen Anforderungssituation erhalten, welche Form diese Informationen haben und nach welchen Arbeitsschritten in der Regel eine Kontrolle der Arbeit erfolgt. Die spezifischen, die Tätigkeit von Industriekaufleuten betreffenden Fragen wurden durch einen auf allgemeine berufliche Fähigkeiten abzielenden Fragebogen ergänzt. Dieser Fragebogen wurde in Anlehnung an das Rating-Verfahren für 812 Berufe der U.S.-amerikanischen Wirtschaft konstruiert (vgl. Donsbach, Tsacoumis, Sager & Updegraff, 2003). Hintergrund dieses Ratings ist eine durch das U.S. Department of Labor entwickelte Datenbasis, das Occupational Information Network (O*NET; vgl. Peterson, Mumford, Borman, Jeanneret & Fleishman, 1998). Diese Datenbasis wird in regelmäßigen Zyklen aktualisiert. Hierzu werden die Berufe durch Experten im Hinblick auf verschiedene zur Ausübung des Berufs notwendige Fähigkeiten geratet. Insgesamt umfasst der Ratingkatalog 52 verschiedene Fähigkeitskategorien. Unter beruflichen Fähigkeiten

werden in diesem Zusammenhang beständige Eigenschaften von Arbeitnehmern verstanden:

Abilities are [...] „relatively enduring attributes of an individual's capability for performing a particular range of different tasks“ (Fleishman, Co-stanza & Marshall-Mies, 1999, p. 175).

Jede dieser Fähigkeiten wird im Hinblick auf das notwendige Fähigkeitsniveau (level), auf die Wichtigkeit der Fähigkeit (importance) innerhalb eines Berufes und auf die Häufigkeit (frequency), mit der die Fähigkeit zur Berufsausübung eingesetzt wird, bewertet. Hinzukommt die Einschätzung, ob die Fähigkeit als Einstellungsvoraussetzung angesehen wird oder ob sich die Fähigkeit bei der Berufsausübung erwerben lässt.

Für die Befragung der Industriekaufleute wurden die 52 Fähigkeiten auf 20 Kategorien reduziert. Erfragt wurden Leseverständnis, Aktives Zuhören, Schreiben, Sprechen, Mathematik, Kritisches Denken, Aktives Lernen, Überprüfen, Soziales Wahrnehmungsvermögen, Organisationsfähigkeit, Beeinflussen, Gesprächsführung, Instruieren, Dienstleistungsorientierung, Komplexe Problemlösefähigkeit, Analysefähigkeit, Auswahl der Arbeitsgeräte, Systemanalysefähigkeit, Beurteilen und Entscheiden sowie Zeitmanagement. In Abbildung 5.2.3 ist für die Fähigkeitskategorie „Beurteilen und Entscheiden“ die Struktur des Fragebogens exemplarisch dargestellt.

Die Bewertung allgemeiner beruflicher Fähigkeiten basiert jeweils auf einer Definition der zu bewertenden Fähigkeit. Die Einschätzung des Fähigkeitslevels erfolgt über eine siebenstufige Skala, bei der eine geringe Ausprägung, die mittlere Ausprägung und eine höhere Ausprägung zur Veranschaulichung der Skala berufsbezogen charakterisiert sind. Die Wichtigkeit der Fähigkeit wird über eine fünfstufige Skala und die Häufigkeit über eine siebenstufige Skala eingeschätzt. Die Einschätzung der Häufigkeit ist insbesondere vor dem Hintergrund eines umfassenden Tätigkeits- und Fähigkeitsprofils von Bedeutung. Sie gibt Auskunft darüber, ob die zu bewertende Fähigkeit alltäglich für routinemäßig anfallende Aufgaben benötigt wird oder aber für besondere, in der Regel periodisch oder betriebsbedingt auftretende Arbeitsaufträge Verwendung findet.

Die Auswertung der Fragebogen orientiert sich an dem Verfahren, das auch für die Ratings des O*NET-Katalogs eingesetzt wird. Bei der Auswertung ist zu berücksichtigen, dass die Anzahl der Befragten sehr klein ist ($N = 11$) und dass die berufsspezifischen Tätigkeitsprofile – Einkauf, Vertrieb, mittleres Management – sich erheblich voneinander unterscheiden. Für das Kriterium einer tätigkeitsbezogenen Validität sind diese Unterschiede von besonderer Bedeutung, da solche Fähigkeiten identifiziert werden können, die stark vom Arbeitsplatz eines Industriekaufmanns/einer Industriekauffrau abhängen und von daher eher als tätigkeitsspezifische denn als allgemeine Fähigkeiten charakterisiert werden sollten. Das Ziel der Erhebung ist, solche Fähigkeiten zu identifizieren, die allgemein für die Beschäftigung von Industriekaufleuten von Bedeutung sind.

Abbildung 5.2.3 Fragebogen: Allgemeine berufliche Fähigkeiten

19.

Beurteilen und Entscheiden

Fähigkeit, unter Abwägung der Kosten und des Nutzens die bestmögliche Entscheidung zu treffen

Level

Welches Level dieser Fähigkeit benötigen Sie, um Ihre derzeitige Tätigkeit ausführen zu können?

hoch

7

6

5

4

3

2

1

NR

Entscheiden, ob ein neues Produkt in das Sortiment aufgenommen werden soll

Entscheiden, ob ein Zusatzauftrag angenommen werden soll

Entscheiden, ob eine Tätigkeit am Vormittag noch erledigt werden muss

nicht relevant, um die derzeitige Tätigkeit ausführen zu können

Wichtigkeit

Wie wichtig ist diese Fähigkeit, um Ihre derzeitige Tätigkeit ausführen zu können?

nicht wichtig

1

2

3

4

5

äußerst wichtig

Häufigkeit

Wie häufig ist es notwendig, diese Aktivität für Ihre derzeitige Tätigkeit auszuführen?

einmal pro Jahr oder weniger

1

2

3

4

5

6

7

stündlich oder häufiger

Einstellungsvoraussetzung

Ist dieses Level der Fähigkeit notwendig, um diese Tätigkeit beginnen zu können?

1

Ja, es ist notwendig, um diese Tätigkeit beginnen zu können

2

Nein, dieses Level kann während der Tätigkeit erreicht werden

Abbildung 5.2.4 stellt die Ergebnisse in der Zusammenfassung dar (Level und Bedeutung). Um die Übereinstimmung zwischen den Befragten zu klassifizieren,

wurde für jede Fähigkeit und für jede Skala die Standardabweichung (SD) der Ratings über die Personen und der Standardfehler für die einzelnen Ratings (SEM) berechnet. Für beide Werte gilt, dass je niedriger ihr Wert, desto höher die Übereinstimmung und vice versa. Zusätzlich ist der Median (MD) als Wert der zentralen Tendenz angeführt. Die Interrater Reliabilität wird durch den Intraclass-Korrelationskoeffizienten angegeben (two-way random; ICC [2,1]; Shrout & Fleiss, 1979). Ein Wert größer .80 spricht für eine starke Übereinstimmung zwischen den Ratern.

Abbildung 5.2.4 *Ergebnisse: Allgemeine berufliche Fähigkeiten*

Fähigkeit	Level			Bedeutung		
	MD	SD	SEM	MD	SD	SEM
Leseverständnis	5	0,900	0,340	4	0,378	0,143
Aktives Zuhören	6	0,951	0,360	4	0,756	0,286
Schreiben	6	0,756	0,286	4	1,000	0,378
Sprechen	5	0,900	0,340	5	0,488	0,184
Mathematik	6	1,345	0,508	4	0,826	0,309
Kritisches Denken	6	1,345	0,508	4	0,690	0,261
Aktives Lernen	5	1,069	0,404	4	1,113	0,421
Überprüfen	5	0,756	0,286	5	0,756	0,286
Wahrnehmungsvermögen	6	1,272	0,481	4	0,690	0,260
Organisationsfähigkeit	5	1,155	0,436	3	0,535	0,202
Beeinflussen	5	1,113	0,421	4	1,069	0,404
Gesprächsführung	5	1,618	0,612	3	1,397	0,528
Instruieren	5	0,951	0,360	4	1,069	0,404
Dienstleistungsorientierung	5	0,951	0,360	5	0,951	0,360
Problemlösefähigkeit	6	0,900	0,340	5	0,951	0,360
Analysefähigkeit,	6	1,134	0,429	3	1,254	0,474
Auswahl der Arbeitsgeräte	5	1,952	0,738	3	1,380	0,522
Systemanalysefähigkeit	6	1,113	0,421	3	1,113	0,421
Beurteilen und Entscheiden	6	0,951	0,360	5	0,756	0,286
Zeitmanagement	6	0,976	0,369	4	0,690	0,261

Folgende allgemeine berufliche Fähigkeiten lassen sich übereinstimmend als bedeutsam für verschiedene Tätigkeitsprofile von Industriekaufleuten identifizieren: Sprechen, Überprüfen, Dienstleistungsorientierung, Problemlösefähigkeit, Beurteilen und Entscheiden. Dieses Fähigkeiten weisen einen Median von 5, eine relativ geringe Standardabweichung im Bereich von 0,488 bis 0,951 sowie einen niedrigen Standardfehler der Messung im Bereich von 0,184 bis 0,360 auf. Für diese als bedeutsam identifizierten Fähigkeiten ergibt sich ein ICC [2,1] von $\geq 0,898$. Die bedeutsamen allgemeinen Fähigkeiten setzen durchgängig ein hohes Fähigkeitslevel voraus. Der Median variiert zwischen 5 und 6 auf einer siebenstufigen Likert-Skala. Die Übereinstimmung zwischen den Ratern ist auch hinsicht-

lich des Fähigkeitslevels sehr hoch. Der ICC [2,1] beträgt $\geq 0,853$; die Standardabweichung liegt im Bereich von 0,756 bis 0,951 und der Standardfehler ist nicht größer als 0,360.

In die Konstruktion der einzelnen Anforderungssituationen im Rahmen des computerbasierten Testformats ALUSIM fließen die allgemeinen beruflichen Fähigkeiten mit ein. Es wird davon ausgegangen, dass diese Fähigkeiten zentral zur Beschreibung beruflicher Aktivität beitragen und damit helfen, Anforderungssituationen authentisch zu gestalten. Fähigkeiten, wie z. B. Gesprächsführung oder Analysefähigkeit, scheinen hingegen stark von der Stellung bzw. vom spezifischen Arbeitsplatz des Befragten im Unternehmen abhängig zu sein und können vor diesem Hintergrund als arbeitsplatz- bzw. positionsspezifisch und nicht als allgemein generalisierbar klassifiziert werden.

5.2.2.3 Beschreibung des Testformats ALUSIM

Die Beziehung zwischen Kontext, Inhalt und Design eines Tests wird in modernen Assessmentmodellen auf der Basis kognitiver Theorien hergestellt (vgl. Kapitel 3.1; Embretson, 2002; Mislevy, 2007; Wilson, 2008). Diese Modelle veranschaulichen, dass die Entwicklung eines Tests Modellierungsschritte voraussetzt, die die Vorstellungen von Realität im Hinblick auf die Ideen und Ziele der Testung übersetzen (Kriterien der Authentizität und der Validität). Erst auf Basis dieser Übersetzungsleistung können die Inhalte identifiziert, ausformuliert und für die Testaufgaben in Abstimmung mit dem Testformat aufbereitet werden. Zusammenfassend sind hierfür vier Entwicklungsschritte notwendig:

- Die Analyse von Inhalten, die als domänentypisch gelten können,
- die Ausformulierung von Arbeits- und Denkstrukturen vor dem Hintergrund der identifizierten Inhalte auf Basis kognitiver Theorien,
- die Abstimmung der Arbeitsmittel und -methoden sowie
- die Wahl eines geeigneten psychometrischen Modells, mit dem die Besonderheiten der Aufgaben und die unterschiedlichen Fähigkeitsausprägungen der Lernenden/Auszubildenden abgebildet werden können.

Diese vier Entwicklungsschritte sind durch die theoretischen Vorarbeiten im Rahmen der Entwicklung von Kompetenzstrukturmodell sowie Domänenmodell weitestgehend gegangen worden. In diesen Modellierungen wird herausgearbeitet, dass spezifische berufliche Erfordernisse für das erfolgreiche Ausführen beruflicher Handlungen zum Ausgangspunkt der Kompetenzmessung zu machen sind. Die Handlungsprodukte stellen damit die Grundlage für die Beschreibung von berufsfachlicher Kompetenz dar. Im kaufmännisch-verwaltenden Bereich sind diese Produkte i. d. R. durch Bearbeitungs- und Entscheidungsprozesse erzeugte Daten bzw. Datenzusammenfassungen, die unterschiedlich komplex sind und verschiedene Formate aufweisen können. Beispiele hierfür sind Daten in Auftragsbestätigungen oder Bestellformularen bzw. auf Daten basierende Kundenkorrespondenz. Die Daten ihrerseits werden auf Basis betrieblicher Arbeits- und Geschäftsprozesse erzeugt, die sich über ein Zusammenspiel verschiedener Arbeitsmittel, unternehmensinterner ERP-Software und kaufmännischer Entschei-

dungsalgorithmen modellieren lassen. Für die Messung berufsfachlicher Kompetenz wurden diese internen Abläufe und Produkte simuliert, d. h. authentisch und valide für das Testformat modelliert. Die Modellierung dient dem Zweck, ausgewählte domänentypische Geschäftsvorfälle als Wertschöpfungs- bzw. Steuerungsprozesse abzubilden. Diese Geschäftsvorfälle sollten eine Struktur aufweisen, die es möglich macht, Items zu konstruieren, mit deren Hilfe berufsfachliche Kompetenz gemessen werden kann. Für das Testformat bedeutet dies, dass eine authentische Beschreibung und Erfassung der Handlungsorientierung und des Handlungsvollzuges in den Anforderungssituationen möglich werden. Das jeweilige Testformat muss daher

- reale Arbeitsprozesse (Verhandlungen, Arbeitsverteilung),
- reale Geschäftsprozesse (z. B. die konkrete Abwicklung einer Bestellung) sowie
- reale kaufmännische Entscheidungen (z. B. Ermittlung eines Liefertermins) abbilden können. Dies geschieht mit Hilfe der Unternehmenssimulation ALU-SIM. Das Testformat ALUSIM ist web-basiert und enthält eine allgemeine Einführung in die Unternehmensstruktur (einschließlich der Unternehmenshistorie) sowie allgemeine Zusammenstellungen zur geschäftlichen Lage des Unternehmens (textbasierte Analyse der Geschäftszahlen sowie Darstellungen der Bilanz, Erfolgsrechnung und Geldflussrechnung). Diese allgemeinen Informationen sind über die Startseite abrufbar (Abbildung 5.2.5). Die Startseite enthält darüber hinaus ein Video, in dem das simulierte Unternehmen vorgestellt wird.

Abbildung 5.2.5 Startseite der Simulation ALUSIM

ALUSIM GmbH | Goethestraße 33 | 34119 Kassel | service@aluslim.de

aluslim
Startseite

Die ALUSIM GmbH ist Experte für Aluminiumverpackungen. Getränkedosen, Kosmetikflaschen, Bauschaumkartuschen – die Produktionspalette ist umfangreich. Das Unternehmen mit Sitz in Kassel hat 35 Mitarbeitenden; zwei davon absolvieren ihre Ausbildung zum Industriekaufmann / zur Industriekauffrau.

Das Unternehmen
in Zahlen
Drei Testsituationen
Start

Impressum: Georg-August-Universität Göttingen, Professur für Wirtschaftspsychologie, Platz der Göttinger Seiden 5, 37073 Göttingen | Foto: 2005, NOVILLIS AG

Unternehmenszahlen ALUSIM GmbH.xls [Freigegeben]

Bilanz per 31. Dezember 2009

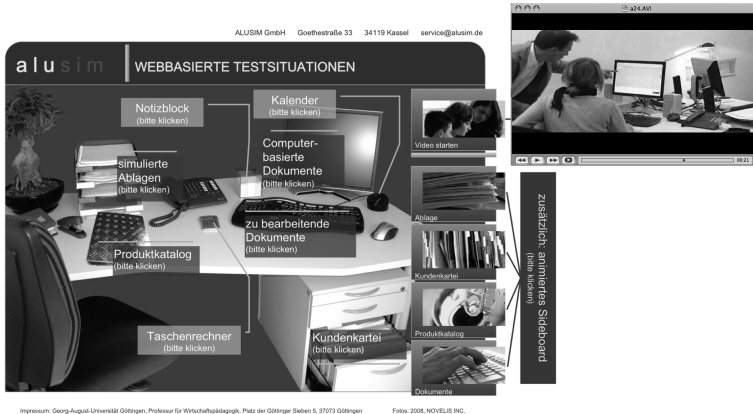
	Erklärungen	2009	%	2008	%
Rechnungen	(3)	1.204		1.094	
Immaterielle Leistungen	(6)	211		202	
Beteiligungen	(26)	4		22	
Umsatz-Forderungen	(7)	127		120	
Umsatz-Rückstellungen	(12)	51		52	
Abgabeleistungen		1.596	81	1.488	85
Vorräte	(36)	597		589	
Forderungen aus Lieferungen und Leistungen	(38)	401		409	
Umsatz-Forderungen	(116)	46		639	
Verkaufskosten	(7)	12		11	
Fremde Mittel	(116)	389		142	
Umsatzleistungen		1.001	49	1.400	47
Bilanz		3.050	100	3.120	100
Reinvermögen		423		434	
Kapitalrücklagen		129		129	
Darlehensrücklagen		281		490	
Gegenkonto		633	21	1.054	34

Bilanz | Erfolgsrechnung | Geldflussrechnung | Erklärung...

Über die Startseite gelangt der Nutzer direkt zu den drei Testsituationen. Jede der Testsituationen wird über einen animierten Arbeitsplatz (Schreibtisch mit

Zugriff auf sämtliche Ablagen und die simulierte Software) sowie ein animiertes Sideboard gesteuert (Abbildung 5.2.6).

Abbildung 5.2.6 User-Interface der Simulation ALUSIM



Das User-Interface ist so gestaltet, dass die Auszubildenden über eine Videosequenz in die Anforderungssituation eingeführt werden. Zur Bearbeitung der Anforderungssituation sind Hilfsmittel, wie Taschenrechner, Notizfunktion oder Kalender, in jeder Testsituation zugänglich. Im Sideboard werden hingegen in Abhängigkeit von der Aufgabenstellung jene Dokumente abgelegt, die für die Bearbeitung der Testsituation genutzt werden müssen. Diese Dokumente sind zusätzlich über den Arbeitsplatz abrufbar. Hierbei symbolisiert der Monitor computer-basierte Dokumente, Datenbanken und ERP-Software. Die Tastatur beschreibt die zu bearbeitenden Dokumente in Form von Arbeitsprodukten. Der Arbeitsplatz enthält darüber hinaus den Produktkatalog sowie die Kundenkartei des Unternehmens. In der Ablage findet der Nutzer zusätzliche Dokumente, in denen zeitlich vorangegangene Arbeitsprozessschritte dokumentiert sind oder die die Anforderungssituationen über das Video hinaus beschreiben.

In der Unternehmenssimulation ALUSIM sind reale Betriebsprozesse der Bereiche Wertschöpfung (Vertrieb, Einkauf) und Steuerung (Arbeitsvorbereitung) verarbeitet. Für diese Bereiche wurden typische Arbeitssituationen als Modelle zur Abbildung von domänenspezifischen Geschäftsvorfällen herausgegriffen, so z. B. für den Vertrieb der Eingang einer Bestellung per Fax. Eine Erweiterung dieser Bereiche sowie die Hinzunahme weiterer Inhalte ist aufgrund der umfangreichen Modellierung der simulierten Unternehmung jederzeit möglich. Das Testformat gibt für die drei betrieblichen Bereiche den organisationalen und strukturellen Rahmen vor.

5.2.3 Measuring-Prozessschritt III: Testitems berufsfachlicher Kompetenz

Am Beispiel des Testbereichs Vertrieb werden im Folgenden einzelne Testitems vorgestellt. Die Testbereiche sind so konstruiert, dass sie jeweils als eigenständige Testumgebung eingesetzt werden könnten. Insgesamt werden über das Testformat ALUSIM Aufgaben für ca. vier Zeitstunden zur Verfügung gestellt. Für ein international vergleichendes Large-Scale Assessment in der beruflichen Bildung ist dieser zeitliche Rahmen zu umfassend, so dass die einzelnen Testbereiche so ausgestaltet sind, dass jeder von ihnen die Skalen berufsfachlicher Kompetenz möglichst vollständig repräsentiert. Hierfür waren Items zu formulieren, die von unterschiedlicher Komplexität und so beschaffen sind, dass sie die zwei verschiedenen Fähigkeitsstrukturen – handlungsbasierte sowie verstehensbasierte Kompetenz – abbilden können. Während die Zuordnung der Items zu den Kompetenzdimensionen auf Basis der zugrunde liegenden kognitiven Struktur erfolgt, ist die Einschätzung der Komplexität eines Items von der Konstruktrepräsentation abhängig (vgl. Kapitel 3.3; u. a. Glaser, Lesgold & Lajoie, 1987). Nach den Prinzipien der funktionalen Modellierung, inhaltlichen Komplexität und kognitiven Taxonomie werden die Items entwickelt und entsprechend des angenommenen Antwortverhaltens der Testpersonen kodiert. Dieses Vorgehen stellt sicher, dass (1) Annahmen über kognitive Theorien in die Itemkonstruktion einfließen und (2) Items formuliert werden, die gut zwischen den zu testenden Personen trennen können, da sie unterschiedliche Fähigkeitsstufen ansprechen.

Der Testbereich Vertrieb umfasst zwei Bearbeitungssequenzen zunehmender Komplexität. Die Ausgangssituation ist wie folgt beschrieben:

Ausgangssituation Testsituation Vertrieb: Die ALUSIM GmbH ist ein kundenorientiertes Unternehmen, das unter anderem aufgrund des schwierigen Marktumfeldes versucht, durch eine halbjährliche Erstellung individueller Preislisten für die wichtigsten Kunden Preisvorteile direkt an den Kunden weiterzugeben. Dieses Vorgehen hat sich bisher bewährt, so dass die Kunden bei Preissenkungen teilweise die Bestellmengen erhöht haben. Auch die Kundenzufriedenheit konnte durch dieses Vorgehen verbessert werden; das ergeben vor kurzem durchgeführte Kundenbefragungen. Die individuellen Preislisten werden jeweils zum Oktober und April eines Jahres an die ausgewählten Kunden zusammen mit einer Übersicht der angebotenen Industrieprodukte versendet. Auf Basis dieser Preisvereinbarungen geht am Mittwoch, 1. 12. 2010, eine Bestellung des Kunden FACT Dosiertechnik GmbH per Fax ein.

Diese Ausgangssituation wird den Auszubildenden mit Hilfe einer einleitenden Videosequenz präsentiert. Ein sich anschließendes Video enthält die Aufforderung, die Kundenanfrage der FACT Dosiertechnik GmbH zu bearbeiten. Hierfür werden den Auszubildenden neben einer Computereingabemaske des simulier-

Abbildung 5.2.7 Eingabemaske Kundenauftrag/Arbeitsprodukt Sequenz I

Der gewählte Preis ist von zeitlichen Fristen im Rahmen der rechtlichen Kaufvertragsbestimmungen abhängig. Der Auszubildende muss alle notwendigen Informationen eigenständig sammeln und im Hinblick auf ihre Lösungsrelevanz bewerten.

Zur Erstellung der Auftragsbestätigung mit Hilfe des ERP-Systems müssen die Auszubildenden auf verschiedene Informationen zugreifen: die Kundenkartei, der Produktkatalog, die halbjährlich erstellten Preislisten. Eine besondere Schwierigkeit bei der Bearbeitung der ERP-Eingabemaske liegt darin, den für den Kundenauftrag gültigen Preis zu ermitteln. Die Auszubildenden müssen davon Kenntnis haben, dass der gewählte Preis des bestellten Produkts von zeitlichen Fristen im Rahmen der rechtlichen Kaufvertragsbestimmungen abhängig ist. Für den konkreten Fall bedeutet dies, dass nicht der Preis aus dem Produktkatalog, sondern der Preis aus der halbjährlich versandten kundenindividuellen Preisliste zugrunde zu legen ist. Für die Bearbeitung der entsprechenden Spalte der ERP-Eingabemaske („Preis pro Stück“) wird daher erwartet, dass im Vergleich zu den anderen Spalten weniger Auszubildende die korrekte Lösung finden. Der Anstieg an Komplexität und damit die Abnahme der Lösungswahrscheinlichkeit lässt sich unter Rückgriff auf die Prinzipien der Konstruktrepräsentation begründen: Die

Ermittlung des gültigen Preises setzt eine im Vergleich umfangreichere funktionale Modellierungsleistung voraus.

Abbildung 5.2.8 Auftragsbestätigung

01 / 12 2010 MI 13:10 FAX +49 651 988-34 ALUSIM IP		001/001
--	--	---------

FAX MESSAGE	
--------------------	--

	<table border="0" style="width: 100%;"> <tr> <td style="width: 50%;">Empfänger: FACT Dosiertechnik</td> <td style="width: 50%;">Fax: 0351 9885031</td> </tr> <tr> <td>Frau Margot Glüsing</td> <td>Datum: 01.12.2010</td> </tr> <tr> <td>Absender: Herr Heiner Kolbe</td> <td>Seitenanzahl: 1</td> </tr> <tr> <td>Betreff: Auftragsbestätigung</td> <td>Bemerkung:</td> </tr> </table>	Empfänger: FACT Dosiertechnik	Fax: 0351 9885031	Frau Margot Glüsing	Datum: 01.12.2010	Absender: Herr Heiner Kolbe	Seitenanzahl: 1	Betreff: Auftragsbestätigung	Bemerkung:
Empfänger: FACT Dosiertechnik	Fax: 0351 9885031								
Frau Margot Glüsing	Datum: 01.12.2010								
Absender: Herr Heiner Kolbe	Seitenanzahl: 1								
Betreff: Auftragsbestätigung	Bemerkung:								

ALUSIM GmbH	34119 Kassel
	Goethestraße 33
	Email: service@alusim.de

ALUSIM GmbH, Goethestraße 33, 34119 Kassel

FACT Dosiertechnik GmbH
 Frau Margot Glüsing
 Königsbrücker Str. 103
 01099 Dresden

Ihr Zeichen, Ihre Nachricht vom Glü, 01.12. 2010	Unser Zeichen, unsere Nachricht vom Ko	Telefon, Name 0561 988- 33, Heiner Kolbe	Datum 01.12.2010
---	---	---	---------------------

Auftragsbestätigung

Sehr geehrte Frau Glüsing,

wir bedanken uns für Ihre Bestellung und bestätigen Ihnen hiermit die Lieferung von 5000 Aluminiumkartuschen Artikel - Nr. 230330 (Volumen 330 ml, Durchmesser 46 mm, Höhe 218 mm) am 04. Dezember 2010.

Für die Zahlung gelten die für Sie bekannten Konditionen:
 Zahlung innerhalb von 10 Tagen unter Abzug von 3,0 % Skonto, ansonsten 30 Tage netto Kasse.

Mit freundlichen Grüßen

ALUSIM GmbH

Heiner Kolbe
 (Dieser Brief wurde maschinell erstellt und enthält deswegen keine Unterschrift.)

Ust.-ID Nr.: DE 126889977 Geschäftsführer: Dr. Konrad Kluge Handelsregister: Amtsgericht Kassel, HRB 38092	Konten: Deutsche Bank Kassel Konto-Nr. 0181890, BLZ 520 700 24 Sparkasse Kassel Konto-Nr. 003 020 99, BLZ 520 503 53
--	---

Der Lagerbestand betrug bei der Bearbeitung der Kundenanfrage 20.625 Stück des Produkts 230330 Aluminiumkartusche bei eingebuchten Aufträgen in Höhe von 18.540 Stück, so dass die vom System generierte und an den Kunden FACT Dosiertechnik GmbH versendete Auftragsbestätigung einen Liefereingang

für den 4. 12. 2010 avisiert. Diese Information stellt die Grundlage für die Bearbeitung der zweiten Sequenz des Testbereichs Vertrieb dar.

Abbildung 5.2.9 E-Mail aus der Fertigungssteuerung

Seite 1 von 1

Kolbe, Heiner

Von: „Theresa Heinrich“ [heinrich@alusim.de]
Gesendet: Mittwoch, 01. Dezember 2010, 17:16 Uhr
An: Kolbe, Heiner
Betreff: Auftragsnummer 800003AK330

Auftragsnummer 800003AK330 konnte vom Versand nicht ausgeführt werden; die geforderte Menge war nicht vorrätig.

Der Auftrag wurde daher von uns in die aktuelle Produktionsplanung eingebucht.

Fertigungssteuerung

ALUSIM GmbH, Werk Kassel
Goethestraße 33
D-34119 Kassel
Direct: +49 (0) 561 / 988-45
Fax: +49 (0) 561 / 988-46
E-mail: heinrich@alusim.de

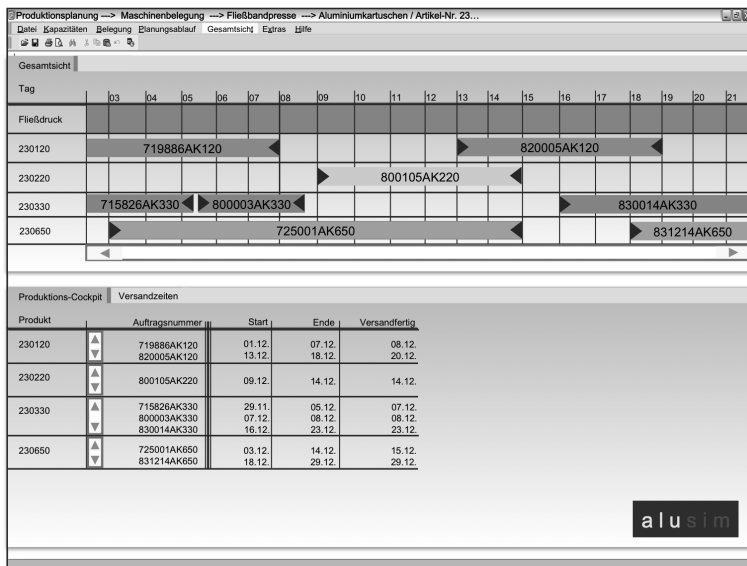
ALUSIM GmbH
Sitz: Kassel (HRB 38092)
Geschäftsführung: Dr. Konrad Kluge

In dieser Sequenz werden die Auszubildenden damit konfrontiert, dass am Nachmittag des 1. 12. 2010 eine Meldung aus der Fertigungssteuerung via E-Mail eingeht (Abbildung 5.2.9), in der mitgeteilt wird, dass der Kundenauftrag aufgrund fehlender Lagerbestände nicht zu dem in der Auftragsbestätigung angegebenen Termin erfüllt werden kann. Der Kundenauftrag wurde daher in die aktuelle Produktionsplanung aufgenommen und entsprechend terminlich fixiert. Die E-Mail aus der Fertigungssteuerung liegt als Ausdruck zusammen mit der Bestellung des Kunden und der Auftragsbestätigung in der Ablage des Auszubildenden.

In einer Videosequenz, die den zweiten Teil des Testbereichs Vertrieb einleitet, werden die Auszubildenden aufgefordert, dem Kunden FACT Dosiertechnik GmbH eine E-Mail zu schreiben, in der der neue Liefertermin mitgeteilt wird. Hierfür werden den Auszubildenden der eingebuchte Produktionsauftrag sowie die Versandzeiten in Tagen zu Verfügung gestellt (Abbildung 5.2.10). Die geschriebene E-Mail stellt für die Sequenz 2 das Leistungskriterium/Arbeitsprodukt dar. Die von den Auszubildenden geschriebene E-Mail wird anhand von drei Kriterien bewertet: (1) Kann der Auftrag zugeordnet werden, (2) ist der Liefertermin korrekt ermittelt und (3) werden in der E-Mail formale Aspekte der Höflichkeit eingehalten? Am Beispiel der E-Mail lässt sich verdeutlichen, wie auf der Ba-

sis komplexer Anforderungssituationen nicht nur fachliche Kenntnisse und Fertigkeiten geprüft werden können. Bei der Bearbeitung der zweiten Sequenz des Testbereichs Vertrieb werden darüber hinaus sowohl methodische als auch personale Kompetenzen vorausgesetzt: methodische Kompetenzen im Hinblick auf den Umgang mit betrieblichen ERP-Systemen einschließlich der Produktionsplanung und personale/soziale Fähigkeiten im Hinblick auf die angemessene Kommunikation mit einem Kunden.

Abbildung 5.2.10 Produktionsplanung



Die ex-ante-Einschätzung des Anforderungsniveaus dieser Testsequenz fokussiert insbesondere auf das Prinzip der kognitiven Taxonomierung. Während im Rahmen der ersten Sequenz insbesondere der situationsspezifische Abruf von Wissen im Mittelpunkt stand (Komplexitätsniveau 1), werden bei der Ermittlung eines korrekten Liefertermins und der entsprechenden Kundenkommunikation die schlussfolgernde Auslegung von Wissen im Hinblick auf die Plausibilität einer situationsspezifischen Lösung (Komplexitätsniveau 3) vorausgesetzt.

Die beschriebenen Sequenzen stehen repräsentativ für die Skala der handlungsbasierten Kompetenz. Die Items sind so konstruiert, dass überwiegend prozedurale Fähigkeitsstrukturen sowie interpretative Entscheidungen zur Lösung abgerufen werden müssen. Jeder Testbereich enthält jedoch auch Items, die vorrangig auf die Aktivierung deklarativer Wissensbestände abzielen und in denen ein grundlegendes Verständnis der betriebswirtschaftlichen Theorie abgefragt

wird. Hierfür wurden auf Basis der Inhalte der handlungsbasierten Testsequenzen Anwendungsaufgaben konstruiert, die sich auf die Simulation beziehen lassen und die im Rahmen der Simulation durchgeführten Handlungen und getroffenen Entscheidungen überprüfen. Für den Testbereich Vertrieb wurden insgesamt zwölf Items für die Skala der verstehensbasierten Kompetenz entwickelt (vgl. Abbildung 5.2.11).

Abbildung 5.2.11 Items der Skala der verstehensbasierten Kompetenz



Entsprechend der Ausgangssituation des Testbereichs Vertrieb beziehen sich die verstehensbasierten Anwendungsaufgaben auf rechtliche Aspekte des Kaufvertrags sowie auf praktische Auslegungen von Angebot und Annahme. Die Anwendungsaufgaben der Skala der verstehensbasierten Kompetenz dienen insbesondere dazu zu prüfen, in welchem Maße die Auszubildenden über grundlegende ökonomische Begriffs- und Konzeptstrukturen verfügen und, unter Rückgriff auf spezifische Beispiele, Systemkohärenzen herstellen können. Die Items sind in die web-basierte Computersimulation als Multiple Choice-Items bzw. als Kurzesaufgaben eingebunden.

Die Entwicklung der Items für die Testbereiche Einkauf und Arbeitsvorbereitung erfolgte analog. Zur Erfassung der berufsfachlichen Kompetenz liegen aus den drei Testbereichen insgesamt 60 Items vor. 34 Items dienen der Erfassung handlungsbasierter Kompetenz über Simulationen innerhalb betrieblicher Anforderungssituationen und 26 Items der Erfassung verstehensbasierter Kompetenz über Anwendungsaufgaben vor dem Hintergrund betrieblicher Anforderungssituationen. Die einzelnen Testbereiche sind wie folgt über Items der handlungsbasierten Simulationen und verstehensbasierten Anwendungsaufgaben repräsentiert (Abbildung 5.2.12):

Abbildung 5.2.12 Itemanzahl der Skalen berufsfachlicher Kompetenz

	Betriebliche Wertschöpfung		Betriebliche Steuerung	
	Testbereich Einkauf	Testbereich Vertrieb	Testbereich Arbeitsvorbereitung	
Itemanzahl der Skala „Handlungsbasierte Kompetenz“	18	12	4	34
Itemanzahl der Skala „Verstehensbasierte Kompetenz“	11	12	3	26
Insgesamt	29	24	7	60

Hinsichtlich der verschiedenen Testbereiche fällt auf, dass der Bereich der „Arbeitsvorbereitung“ in Relation zu den zwei anderen Testbereichen unterrepräsentiert ist. Dies liegt daran, dass dieser Testbereich ausschließlich den Umgang mit dem betriebswirtschaftlichen Konzept der Maschinenstundensatzberechnung umfasst – eine Ausweitung dieses auf die Steuerungsprozesse von Unternehmen abzielenden Testbereichs ist jedoch auf Grundlage der umfassend modellierten Datenstruktur jederzeit möglich. Das stärkere Gewicht der Simulation lag auf betrieblichen Wertschöpfungsprozessen. Hierfür waren insbesondere zwei Gründe ausschlaggebend: (1) Die Analyse der Ausbildungsplatzanforderungen zeigt, dass die Bereiche Vertrieb und Einkauf für einen internationalen Vergleich von besonderem Interesse wären, da mit ihnen die Internationalität der Geschäftsbeziehungen, Aspekte der Kunden- und Lieferantenbindungen und insgesamt ein Verständnis für betriebliche Geschäfts- und Arbeitsprozesse getestet werden können (Breuer, Hillen & Winther, 2009). (2) Die Analyse der deutschen Curricula hat deutlich gemacht, dass – will man weite Teile des Rechnungswesen von einem internationalen Vergleich ausschließen – der inhaltliche Fokus auf dem Bereich der betrieblichen Wertschöpfung liegt.

5.2.4 Measuring-Prozessschritt IV: Ergebnisse

Die nachfolgende Auswertung der Daten und Ergebnisanalyse bezieht sich auf ausgewählte Fragestellungen:

- Wie lassen sich die im Kompetenzstrukturmodell definierten Skalen beruflicher Handlungskompetenz über repräsentative Testitems beschreiben? – In diesem Zusammenhang ist zu klären, ob sich berufliche Handlungskompetenz, wie im Kompetenzstrukturmodell angenommen, über eine mehrdimensionale Kompetenzstruktur abbilden lässt, oder aber eine eindimensionale Lösung resultiert.
- Lässt sich im Hinblick auf die Personenfähigkeit ein Unterschied zwischen handlungsbasierten Simulationsaufgaben und verstehensbasierten Anwendungsaufgaben empirisch sichern? – Die Hypothese ist, dass handlungsbasierte Simulationsaufgaben durch ihren Authentizitätsgehalt und

ihre Nähe zur betrieblichen Praxis in einer Gesamtanalyse von Auszubildenden des Dualen Systems besser gelöst werden können als Aufgaben, die stärker von der betrieblichen Praxis abstrahieren und deren vorrangiger Sinn darin liegt zu testen, ob betriebswirtschaftliche Konzepte verstanden worden sind und ob diese vor dem Hintergrund simulierter Unternehmensprozesse von den Auszubildenden interpretiert und validiert werden können.

- Lässt sich für berufliche Handlungskompetenz ein Kompetenzniveau-Modell formulieren? – Dies erfordert Aussagen darüber, welche Itemeigenschaften charakteristisch für mögliche Kompetenzstufen sind und über welche Merkmale ein Item demnach als schwierig oder als leicht zu identifizieren ist. Im Zusammenhang mit den Ausführungen zur Konstruktrepräsentation (vgl. Kapitel 3.3) wurden hierfür Anforderungsparameter vorgestellt, mit denen (a) die Komplexität von Anforderungssituationen systematisch variiert werden kann und anhand derer (b) Testpersonen verschiedenen Kompetenzstufen zuzuordnen sind. Der Anspruchsgehalt einer beruflichen Anforderungssituation ist demnach abhängig von (1) der funktionalen Modellierung, (2) der inhaltlichen Komplexität und (3) von der Art der kognitiven Beanspruchung.

Zusätzlich sind die ermittelten Personenfähigkeiten in Hinblick auf verschiedene Zusatzinformationen zu analysieren, um Aussagen darüber treffen zu können, ob Faktoren wie das Geschlecht, die Leistungen in den Berufsschulfächern Deutsch und Englisch oder die Größe des Ausbildungsbetriebes Anteile der Personenfähigkeit erklären können. Die untersuchungsleitende Fragestellung hierzu lautet:

- Lassen sich Faktoren identifizieren, die Unterschiede zwischen den Kompetenzausprägungen der Auszubildenden aufzuklären vermögen? – Insbesondere im Hinblick auf diese Fragestellung sind die Grenzen der vorliegenden Studie sehr deutlich: Die über die Individualdaten erzeugten Gruppen von Auszubildenden weisen eine zu geringe Teilstichprobengröße auf, um mehrebenenanalytische Auswertungen durchführen zu können. Es können folglich nur Tendenzen auf Basis des vorliegenden Samples aufgezeigt werden.

Die vier skizzierten Forschungsfragen werden im Folgenden einzeln beantwortet, bevor eine zusammenfassende Bewertung gegeben wird.

5.2.4.1 Design der Studie

Die Anlage der Studie wird als one-group posttest-only-Design beschrieben (vgl. Cook & Campbell, 1979) und ist damit eher pre-experimentell als quasi-experimentell. Zur Erhöhung der Validität wurden spezifische Hintergrundinformationen erhoben. Neben allgemeinen Angaben zu Alter, Geschlecht und Bildungsabschluss wurden insbesondere Details des Ausbildungsbetriebes (Unternehmensgröße und Branche) erfragt.

Insgesamt wurden 264 Auszubildende (3. Ausbildungsjahr, sieben Berufsschulen, 61 Betriebe, drei Bundesländer) erfasst; 56,5 Prozent von ihnen sind weiblich. Das Alter der Auszubildenden variiert zwischen 18 und 34 Jahren, wobei 83,4 Prozent der Befragten in der Altersgruppe von 20 bis 23 Jahren liegen – dies ist typisch für eine Abschlussklasse im Rahmen einer dreijährigen kaufmännischen Berufsausbildung.

Abbildung 5.2.13 Beschreibung der Stichprobe (N = 264)

Information	Beschreibung	Ergebnis
Geschlecht	weiblich:	56,5 %
	männlich:	43,5 %
Alter	Streuung:	18 bis 34 Jahre
	Häufung:	20 bis 23 Jahre
		(Anteil = 83,4 %)
Höchster Schulabschluss	allgemeine Hochschulreife:	43,1 %
	Fachhochschulreife:	40,5 %
	Sekundarabschluss I:	10,7 %
	qualifizierter Hauptschulabschluss:	0,8 %
Betriebsgröße	bis 50 Beschäftigte:	6,1 %
	51 bis 500 Beschäftigte:	38,2 %
	501 bis 2500 Beschäftigte:	27,1 %
	mehr als 2500 Beschäftigte:	24,8 %
Branche (Auswahl)	Metallverarbeitung:	19,9 %
	Holzverarbeitung:	10,1 %
	Chemie:	10,8 %

Mit dem Sample konnte eine angemessene Verteilung der Auszubildenden im Hinblick auf die Branchen und die Betriebsgrößen realisiert werden. Nur sechs Prozent der Auszubildenden absolvieren ihre Ausbildung in Kleinstbetrieben (bis 50 Beschäftigte); die Mehrzahl lernt in Unternehmen mit bis zu 500 (38,2 Prozent) und 24,8 Prozent in Unternehmen mit mehr als 2.500 Beschäftigten. 19,9 Prozent der Auszubildenden kommen aus der Metallverarbeitung, 10,8 Prozent aus der Chemie und 10,1 Prozent aus der Holzverarbeitung. Darüber hinaus sind u. a. Auszubildende aus den Bereichen Energieversorgung, Bau sowie Automobilzulieferung im Sample enthalten. Die Stichprobe zeigt, dass der Ausbildungsberuf des Industriekaufmanns/der Industriekauffrau ein höherer kaufmännischer Ausbildungsgang ist – 43,1 Prozent der Auszubildenden verfügen über die allgemeine Hochschulreife und 40,5 Prozent über die Fachhochschulreife; nur zwei Probanden haben einen qualifizierten Hauptschulabschluss.

5.2.4.2 Modellfit: Dimensionen berufsfachlicher Kompetenz

Der Test zur Erfassung berufsfachlicher Kompetenz wurde auf Basis des Kompetenzstrukturmodells für die kaufmännische Bildung konstruiert (Winther & Achtenhagen, 2008a; 2009b). Das Kompetenzstrukturmodell beschreibt, welche ko-

gnitiven Prozesse zur Bearbeitung eines Inhaltsbereichs vorrangig zu aktivieren sind. Das Modell unterscheidet hier grundsätzlich zwischen zwei verschiedenen Kompetenzdimensionen: verstehensbasierte und handlungsbasierte Kompetenz. Die Kompetenzdimensionen werden durch verschiedene Testitems beschrieben. Die Items zur Messung handlungsbasierter Kompetenz sind tätigkeitsbezogene Simulationsaufgaben, die auf Handlungsfähigkeiten in spezifischen beruflichen Anforderungssituationen abstellen. Testitems, mit denen verstehensbasierte Kompetenz gemessen wird, werden hingegen als Anwendungsaufgaben vor dem Hintergrund spezifischer beruflicher Anforderungssituationen charakterisiert. Der Test zur Messung berufsfachlicher Kompetenz wurde so konstruiert, dass ein Item jeweils nur eine der beiden möglichen latenten Fähigkeitsstrukturen abbildet (between-item-Multidimensionalität; Adams, Wilson & Wang, 1997; Wang, Wilson & Adams, 1997). Bei der Modellprüfung gehen 34 Items zur Messung der handlungsbasierten Fähigkeit und 26 Items zur Messung der verstehensbasierten Fähigkeit ein. Da es sich bei den im Test zur Messung berufsfachlicher Kompetenz gewonnenen Daten um ordinale Daten handelt, werden Modelle benötigt, in denen die ordinale Datenstruktur auch abgebildet werden kann. Für das eindimensionale Modell wird auf das Partial Credit Model (PCM; Masters, 1982) und für das zweidimensionale Modell auf das Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM; Adams, Wilson & Wang, 1997) zurückgegriffen. Beide Modelle lassen sich mit Hilfe der Software *ConQuest* berechnen (Wu, Adams, Wilson & Haldane, 2007). Das MRCML-Modell ist ein generalisiertes IRT-Modell der Rasch-Familie und kann folglich auch das Partial Credit Model repräsentieren. Diese hierarchische Beziehung – das PCM ist ein Untermodell des MRCMLM – erlaubt es, die Güte der Modellpassung direkt zu vergleichen. Über einen Likelihood-Ratio-Test auf Basis der Chi-Square-Statistik wird die Signifikanz bestimmt. Die Fit-Statistik für die Kompetenzdimensionen ist in Abbildung 5.2.14 dargestellt.

Abbildung 5.2.14 Fit-Statistik für das ein- und zweidimensionale Modell

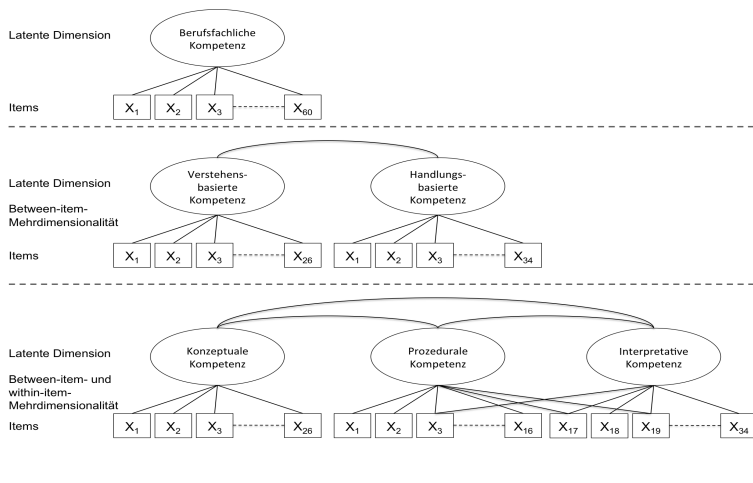
	Eindimen- sionales Modell	Zweidimen- sionales Modell	Differenz
Deviance (-2Log-Likelihood)	14.505,50	14.403,59	101,91
Anzahl der geschätzten Parameter	80	83	3

Alle 60 Items zur Messung berufsfachlicher Kompetenz sind in die Modellprüfung eingegangen; zur Modellierung der Dimensionen im zweidimensionalen Modell wurden die 34 Items der handlungsbasierten Kompetenz der Dimension I und die 26 Items der verstehensbasierten Kompetenz der Dimension II zugeordnet. Insgesamt sind 78 Item- und Itemstepparameter sowie die jeweiligen Populationschätzer der zwei Modelle zu schätzen. Die Güte des Modellfits wird über die Veränderung im Deviance-Wert bestimmt. Die geschätzte Differenz zwischen den Deviance-Werten von 101,91 ist signifikant ($df = 3$; $p < ,000$). Damit fittet das

zweidimensionale Modell die Daten besser als das eindimensionale Modell. Die handlungsbasierten Simulationsaufgaben und die verstehensbasierten Anwendungsaufgaben messen folglich unterschiedliche latente Dimensionen. Oder anders formuliert: Zur Bewältigung der betrieblichen Simulationen werden von den Probanden andere Fähigkeitsstrukturen eingesetzt als zur Bewältigung der Anwendungsaufgaben. Während die Simulationsaufgaben eher eine prozedurale Struktur aufweisen, werden in den Anwendungsaufgaben vorrangig deklarative Wissensbestände benötigt. Zusätzlich zum Signifikanzwert wird die Korrelation zwischen den latenten Dimensionen zur Beurteilung der Effektstärke herangezogen. Verstehensbasierte und handlungsbasierte Kompetenz korrelieren mit $r = 0,587$ (Kovarianz = 0,300). Diese Korrelation ist signifikant verschieden von eins ($p = ,000$) und spricht ebenfalls deutlich für eine zweidimensionale Lösung.

Unter Rückgriff auf das adaptierte Kompetenzstrukturmodell kann eine weitere mehrdimensionale Modellierung geprüft werden. Handlungsbasierte Kompetenz wird im Rahmen des Kompetenzstrukturmodells als prozessuale Kombination aus prozeduralen und interpretativen Fähigkeitsstrukturen definiert; verstehensbasierte Kompetenz wird über konzeptuale Fähigkeiten beschrieben. Die Items der Skala der handlungsbasierten Kompetenz sind dabei teilweise so konstruiert, dass zu ihrer Lösung beide Fähigkeitsstrukturen eingesetzt werden müssen. In dem Fall, dass ein einzelnes Item mehr als eine latente Dimension abbilden kann, wird von within-item-Mehrdimensionalität gesprochen.

Abbildung 5.2.15 Vergleich zwischen ein-, zwei- und dreidimensionaler Modellierung



Für die Definition berufsfachlicher Kompetenz unter Rückgriff auf das Kompetenzstrukturmodell kann folglich eine dreidimensionale Kompetenzstruktur angenommen werden, die zwischen konzeptueller, prozeduraler und interpretativer

Kompetenz differenziert. Hierbei gilt, dass die 26 Items der verstehensbasierten Anwendungsaufgaben der konzeptualen Kompetenz zugeordnet werden, die 34 Items der handlungsbasierten Simulationsaufgaben können sich hingegen teilweise auf zwei latente Dimensionen beziehen. In Abbildung 5.2.15 sind die Modellierungen der Dimensionen dargestellt. Die Abbildung zeigt für den dreidimensionalen Fall, dass beispielsweise für die Lösung des Items X_3 der handlungsbasierten Simulationsaufgaben sowohl prozedurale als auch interpretative Kompetenzen notwendig sind (within-item-Mehrdimensionalität).

Die Prüfung der Modellgüte erfolgt wie oben beschrieben: Das dreidimensionale Modell weist mit einem Unterschied von 112,85 in den Deviance-Werten einen signifikant besseren Modellfit als das eindimensionale Modell auf ($df = 7$; $p < ,001$). Der Vergleich zwischen dem zweidimensionalen und dreidimensionalen Modell wird zugunsten der dreidimensionalen Lösung signifikant (Chi-Square = 10,94; $df = 4$; $p < ,05$; vgl. Abbildung 5.2.16).

Abbildung 5.2.16 Fit-Statistik für das ein-, zwei- und dreidimensionale Modell

Modell	Deviance	Anzahl der geschätzten Parameter	Differenz Deviance	df
Eindimensionales Modell	14.505,50	80		
Zweidimensionales Modell	14.403,59	83	101,91	3
Dreidimensionales Modell	14.392,65	87	10,94	4

Allerdings zeigt die Prüfung der Effektstärke, dass prozedurale und interpretative Kompetenz mit $r = ,901$ sehr hoch korrelieren. Zudem wurde der Modellfit nach 1.000 Iterationen abgebrochen. Die zweidimensionale Lösung konnte hingegen in 29 Iterationen geschätzt werden. Ein Vergleich der Varianzschätzer zeigt darüber hinaus, dass die latente Dimension „interpretative Kompetenz“ eine sehr geringe Varianz aufweist (Abbildung 5.2.17).

Abbildung 5.2.17 Kovarianz, Korrelation und Varianz des dreidimensionalen Modells

Latente Dimension	Prozedurale Kompetenz	Interpretative Kompetenz	Konzeptuale Kompetenz	Varianz
Prozedurale Kompetenz	---	0,188	0,266	0,498
Interpretative Kompetenz	0,901	---	0,132	0,088
Konzeptuale Kompetenz	0,574	0,679	---	0,433

Anmerkung: Schätzer der Kovarianzen oberhalb und der Korrelationen unterhalb der Diagonalen

Die Gesamtwürdigung der Kennwerte legt nahe, dass das zweidimensionale Modell insgesamt die beste Modellgüte aufzeigt und aus diesem Grund für die Analyse der Ergebnisse zugrunde gelegt werden sollte. Zur Veranschaulichung der Berechnung des zweidimensionalen Modells sind in Abbildung 5.2.18 die notwendigen Eingaben in *ConQuest* dargestellt. Die Antworten der Auszubildenden sind mit 0, 1 und 2 ordinal skaliert. Im Input-Bereich „Valid codes und scoring“ erfolgt zum einen die Benennung gültiger Eingabewerte und zum anderen die Zuordnung der einzelnen Items zu den latenten Dimensionen. Die ersten drei score-Zeilen enthalten die Zuordnung der Testitems zu der latenten Dimension „Handlungsbasierte Kompetenz“. Die Items eins bis zwölf sind dabei Items des Testbereichs Vertrieb; die Items 25 bis 42 repräsentieren den Testbereich Einkauf und die Items 54 bis 57 entsprechend den Testbereich Arbeitsvorbereitung. Jedes dieser Items der handlungsbasierten Simulationsaufgaben ist ausschließlich der handlungsbasierten Kompetenzdimension zugeordnet. Diese between-item-Mehrdimensionalität wird in *ConQuest* dadurch ausgedrückt, dass die Klammer, die die zweite latente Dimension repräsentiert, unbesetzt bleibt. Entsprechend erfolgt die Zuordnung der verstehensbasierten Items zur Fähigkeitsdimension „Verstehensbasierte Kompetenz“ in den score-Zeilen vier bis sechs. Die Items 13 bis 24 des Testbereichs Vertrieb, die Items 43 bis 53 des Testbereichs Einkauf sowie die Items 58 bis 60 des Testbereichs Arbeitsvorbereitung sind ausschließlich der zweiten latenten Dimension zugewiesen; die Klammern, die die erste latente Dimension repräsentieren, bleiben leer.

Abbildung 5.2.18 *ConQuest-Input zur Berechnung des zweidimensionalen Modells*

```
/* Title and data definition */
Datafile ALUSIM_ALL.dat;
Format id 1-7 responses 8-67;
Labels << ALUSIM_ALL.lab;
```

```
/* Valid codes and scoring */
codes 0,1,2;
score (0,1,2) (0,1,2) ( ) ! items(1-12);
score (0,1,2) (0,1,2) ( ) ! items(25-42);
score (0,1,2) (0,1,2) ( ) ! items(54-57);
score (0,1,2) ( ) (0,1,2) ! items(13-24);
score (0,1,2) ( ) (0,1,2) ! items(43-53);
score (0,1,2) ( ) (0,1,2) ! items(58-60);
```

```
set warnings=no,update=yes;
```

```
set constraints=none;
```

```
/* Define model */
```

```
Model item + item*step;
```

```
/* Estimate model parameters */
```

```
Estimate;
```

```
/* Define output files */
```

```
Show ! estimates=latent, tables=1:2:3:4 >> ALUSIM_ALL_DIMENSION.shw;
```

```
show cases! estimates=wle >> ALUSIM_ALL_DIMENSION.wle;
```

```
show cases! estimates=eap >> ALUSIM_ALL_DIMENSION.eap;
```

```
export parameters >> ALUSIM_ALL_DIMENSION.prm;
```

```
Itanal >> ALUSIM_ALL_DIMENSION.itn;
```

```
quit;
```

Ordinale Daten;

Zweidimensionales Modell;

Between-item-Mehrdimensionalität

MRCML-Modell

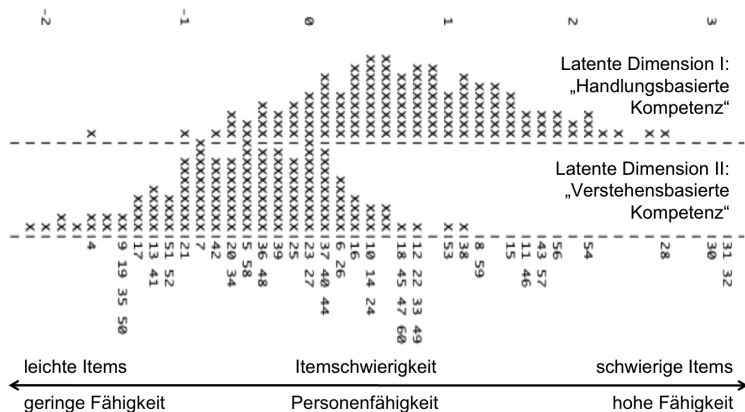
(basierend auf Itemsteps)

Zur Berechnung des zweidimensionalen Modells ist im Input-Bereich „Define model“ das Modell zu beschreiben. Die Zeile „Model item + item*step“ stellt das Kommando für das Partial Credit Model dar (Berücksichtigung der Itemsteps bei ordinalen Daten). Durch die Zuordnung der Items auf zwei latente Dimensionen wird in *ConQuest* zur Berechnung eines mehrdimensionalen Modells das MRCML-Modell als generalisiertes multidimensionales IRT-Modell hinterlegt.

5.2.4.3 Parameterschätzung: Unterschiede zwischen handlungsbasierter und verstehensbasierter Kompetenz

Alle Parameter des zweidimensionalen Modells wurden mit Hilfe der Software *ConQuest* geschätzt. Die verwendete Methode zur Berechnung der Items, Itemsteps und Populationsparameter ist die marginale Maximum Likelihood-Schätzung. Zur Berechnung der Personenparameter sowie der zugehörigen Reliabilitäten werden sowohl der gewichtete Maximum Likelihood-Schätzer (WLE) als auch der auf plausible values basierende Expected a posteriori-Schätzer (EAP/PV) genutzt. Darüber hinaus werden für alle Parameter Standardfehler sowie für die Item- und Itemstepschwierigkeiten Fit-Statistiken zur Verfügung gestellt. Zusammenfassend lassen sich die Schätzer der Item- und Personenparameter in einer Wright Map (Wilson, 2005; Abbildung 5.2.19) darstellen. Die Personenparameter (hier: EAP/PV) sind für die zwei latenten Dimensionen als Histogramm abgebildet; die Itemparameter sind entlang der Logit-Skala angeordnet. In dieser Darstellung sind folglich Personen- und Itemparameter auf eine gemeinsame Skala transformiert.

Abbildung 5.2.19 Wright Map



Anmerkung: Jedes X repräsentiert 2,0 Fälle.

Zur Abbildung der Skalen berufsfachlicher Kompetenz werden zwei latente Dimensionen ineinander gegenübergestellt: die Personenfähigkeit, handlungsbasierte Simulationen innerhalb betrieblicher Situationen lösen zu können (handlungsbasierte Kompetenz), und die Personenfähigkeit, verstehensbasierte Anwendungsaufgaben vor dem Hintergrund betrieblicher Situationen zu bewältigen (verstehensbasierte Kompetenz). Die Skalen sind dann umfassend modelliert, wenn die Testitems, die diese Skalen repräsentieren, auf dem Kontinuum des Personenmerkmals ausreichend streuen. Hierbei ist zu beachten, dass möglichst Items unterschiedlichster Schwierigkeitsgrade gefunden werden sollten. Die Kalibrierung der Items erfolgt dabei auf der Logit-Skala. Diese Skala stellt die Relation zwischen Personenfähigkeit und Itemschwierigkeit dar. Für den vorliegenden Test ist das Item 2 mit einem Itemparameter von -4,297 das leichteste Item und das Item 55 mit einem Schwierigkeitsschätzer von 4,084 das anspruchsvollste Item im Test (aus Darstellungsgründen ist die Logit-Skala der Abbildung 5.2.19 auf den Bereich $-2 \leq x \leq 3$ begrenzt).

Die exakte Zuordnung der Items zu den zwei latenten Dimensionen ist der Abbildung 5.2.19 nicht zu entnehmen; jedoch illustriert diese Darstellung die Vorteile der IRT in folgendem Punkt sehr deutlich: Dadurch, dass die Itemparameter und die Personenparameter auf einer gemeinsamen Skala, der Logit-Skala, transformiert werden, ist das Testverhalten der Personen direkt ersichtlich. Der Populationsmittelwert der Verteilung der handlungsbasierten Kompetenz beträgt 0,750 ($\sigma^2 = 0,612$).

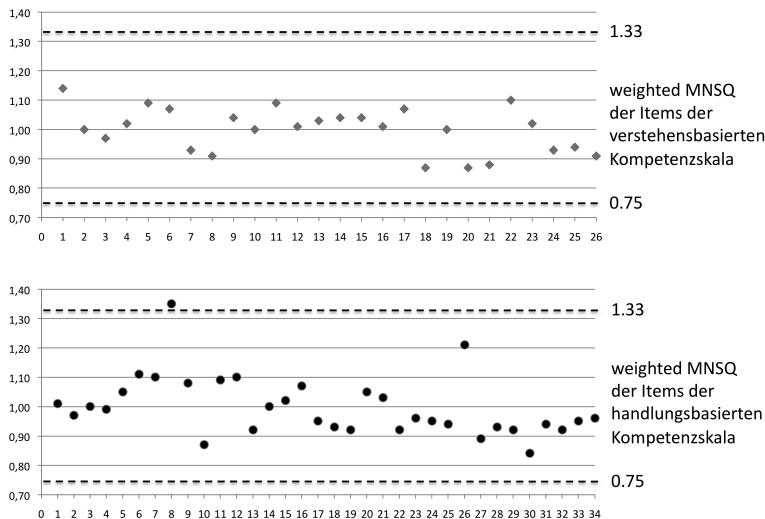
Dieser Wert besagt, dass Items, die einen ebensolchen Schwierigkeitsparameter aufweisen, mit einer Wahrscheinlichkeit von 50 Prozent gelöst werden können. Verfügt eine Person über einen höheren Fähigkeitsparameter, nimmt die Lösungswahrscheinlichkeit zu. Umgekehrt gilt, dass bei Items, die leichter als 0,750 sind, im Mittel eine höhere Wahrscheinlichkeit als $p = 0,5$ besteht, diese Items richtig lösen zu können. Die verstehensbasierte Kompetenz der Auszubildenden weist mit -0,481 ($\sigma^2 = 0,427$) einen niedrigeren Populationsschätzer auf. Dies besagt, dass Items mit einem höheren Schwierigkeitsgrad im Mittel mit einer geringeren Wahrscheinlichkeit als $p = 0,5$ gelöst werden können. In den PISA- und TIMS-Studien sind diese Wahrscheinlichkeiten aufgrund des Mastery-Gedankens von Lernen auf einem höheren Niveau fixiert (Baumert, Bos & Lehmann, 2000, S. 114). Dieses Vorgehen wurde für die vorliegende Studie nicht für zweckmäßig gehalten, da (1) diese Studie keine curricular gesetzten Lernstände erfasst und sich (2) das Testformat der Studie in der Erprobung befindet.

Bei der Interpretation der Populationsschätzer ist anzumerken, dass ein direkter Vergleich zwischen den latenten Dimensionen nicht sinnvoll ist; zwar werden Populationsmittelwerte für zwei verschiedene Fähigkeitsstrukturen geschätzt, diese beziehen sich jedoch auf zwei unterschiedliche Skalen (Wu, Adams, Wilson & Haldane, 2007, p. 94). Diese Argumentation wird zusätzlich durch die geringe Korrelation beider Dimensionen ($r = 0,587$) gestützt. Die Korrelation ist gering genug, um von zwei sich unterscheidenden inhaltlichen Bedeutungen der Skalen ausgehen zu können. Der Korrelationswert zeigt einerseits die hohe praktische

Signifikanz einer zweidimensionalen Lösung im Vergleich zu einem eindimensionalen Modell an, zugleich legt dieser Korrelationswert jedoch auch eine Analyse der beiden Kompetenzskalen im Detail nahe.

Die Fit-Statistik für die Itemparameter ist sehr zufriedenstellend (Abbildung 5.2.20). Die gewichteten Abweichungsquadrate (weighted MNSQ; Wright & Masters, 1982) für das Konstrukt der verstehensbasierten Kompetenz liegen alle innerhalb des zulässigen Intervalls $0,75 \leq wMNSQ \leq 1,33$ (Adams & Khoo, 1996; Bond & Fox, 2001). Alle Infit-Schätzer variieren zwischen 0,87 und 1,14. Für das Konstrukt der handlungsbasierten Kompetenz weist ein Item (Item HK08) mit 1,35 einen zu schlechten Infit auf. Die übrigen 33 Testitems sind im Intervall von 0,82 bis 1,21 symmetrisch um den exakten Fit-Wert von 1,00 verteilt.

Abbildung 5.2.20 Fit-Statistik für die Itemparameter



Als zusätzliches Kriterium des Itemfits wird der T-Wert in *ConQuest* ausgegeben. Die T-Statistik prüft, ob die Abweichungen zwischen erwarteten und beobachteten Häufigkeiten statistisch signifikant sind. Ist der empirisch ermittelte z-Wert größer als der für das $\alpha = 5\%$ -Niveau bei zweiseitigem Test erwartete Testwert ($z = \pm 1,96$), deutet dies auf eine signifikante Abweichung in Richtung einer zu niedrigen Trennschärfe hin. Ein Item (Item HK08) der handlungsbasierten Kompetenzskala und zwei Items (Items VK08 und VK18) der verstehensbasierten Kompetenzskala liegen außerhalb des ermittelten Vertrauensintervalls (Abbildung 5.2.21). Insgesamt können nach Analyse des Itemfits 57 Testitems über das Testformat ALUSIM präsentiert werden; 33 Testitems repräsentieren

die Skala der handlungsbasierten Kompetenz und 24 Testitems die Skala der verstehensbasierten Kompetenz. Ein Vergleich der Itemparameter des eindimensionalen Partial Credit Model und des zweidimensionalen MRCML-Modells zeigt, dass die Itemschätzer weitestgehend identisch sind und nahezu perfekt korrelieren.

Abbildung 5.2.21 *Item-Missfit*

Item	Itemschwierigkeit	weighted MNSQ	Konfidenzintervall	T-Wert
HK08	1,222	1,35	(0,84; 1,16)	3,9
VK08	-0,585	0,91	(0,92; 1,08)	-2,4
VK18	-0,416	0,87	(0,93; 1,07)	-3,5

Anmerkung: HK = handlungsbasierte Kompetenz; VK = verstehensbasierte Kompetenz

Für einen Vergleich der beiden Skalen wurden die Itemparameter einem unabhängigen T-Test unterzogen (Abbildung 5.2.22). Die Itemschwierigkeiten der beiden Skalen unterscheiden sich nicht signifikant ($t = 0,423$; $df = 55$; $p = 0,674$). Der Test zur Messung handlungsbasierter Kompetenz weist im Mittel Items leicht höherer Komplexität auf (Mean = 0,342; SD = 1,967) als der Test der verstehensbasierten Kompetenz (Mean = 0,156; SD = 1,014). Dieser Mittelwertvergleich besitzt jedoch nur geringe Aussagekraft, da sich die zugrundeliegenden Kompetenzskalen in der Breite deutlich voneinander unterscheiden: Die Skala der verstehensbasierten Kompetenz wird von Items im Bereich $-1,409 \leq x \leq 1,730$ beschrieben; die Skala der handlungsbasierten Kompetenz über den Schwierigkeitsbereich $-4,297 \leq x \leq 4,084$.

Abbildung 5.2.22 *T-Test zum Vergleich der Itemschwierigkeiten der Kompetenzskalen*

Latente Dimension	Mean	SD	T-Wert	p-Wert
HK (n=33)	0,342	1,967	0,423	0,674
VK (n=24)	0,156	1,014		

Anmerkung: HK = handlungsbasierte Kompetenz; VK = verstehensbasierte Kompetenz

Die Werte der zentralen Tendenz belegen jedoch, dass für beide Skalen der berufsfachlichen Kompetenz sowohl sehr leichte als auch schwierige Items vorliegen und die Skalen damit umfangreich repräsentiert sind. Insgesamt deuten die Fit-Statistik der Itemparameter sowie die Verteilung der Personenparameter (Histogramm auf Basis der EAP/PV in Abbildung 5.2.19) bereits auf eine angemessene Diskrimination zwischen den Testpersonen hin. Die Wright Map zeigt, dass für das vorliegende Sample die Items gut auf der Logit-Skala verteilt sind. Diese Beobachtung deckt sich mit den relativ kleinen Standardfehlern der Messung der Personenparameter. Für das Konstrukt der handlungsbasierten Kompetenz vari-

iert der Standardfehler der Personen mit vollständigen Daten zwischen 0,35 und 0,46 Logits. Für das Konstrukt der verstehensbasierten Kompetenz sind die Standardfehler der Messung leicht höher (0,38 bis 0,57 Logits). Die relativ niedrigen Messfehler erlauben die Schlussfolgerung, dass es mit dem gegebenen Messinstrument gelingt, die Personen zuverlässig auf verschiedenen qualitativen Ebenen der zu messenden Konstrukte zu verorten.

Abbildung 5.2.23 Reliabilitäten

Latente Dimension	EAP/PV-Reliabilität	WLE-Reliabilität
Berufsfachliche Kompetenz	0,794	0,803
Handlungsbasierte Kompetenz	0,762	0,769
Verstehensbasierte Kompetenz	0,705	0,728

Anmerkung: EAP/PV = expected a posteriori/plausible value; WLE = weighted likelihood estimator

Abbildung 5.2.23 stellt die Reliabilitäten der verschiedenen latenten Dimensionen im Überblick dar. Die Reliabilitäten des zweidimensionalen Modells sind marginal geringer als die Reliabilität der eindimensionalen berufsfachlichen Kompetenz. Die Schätzer der EAP/PV- und WLE-Reliabilitäten unterscheiden sich kaum. Insgesamt werden die Reliabilitätsschätzer als zufriedenstellend bewertet. Auf Basis dieser Werte ist es möglich, Unterschiede zwischen Personen begründet zu analysieren.

Zum Vergleich der Personenfähigkeiten der Skalen handlungsbasierter und verstehensbasierter Kompetenz wurden die Personenschätzer (WLE) mittels eines paarweisen T-Tests für abhängige Stichproben analysiert (Abbildung 5.2.24).

Abbildung 5.2.24 T-Test zum Vergleich der Personenfähigkeiten der Kompetenzskalen

Personenfähigkeit	Mean	SD	T-Wert	p-Wert
HK (n=262)	0,791	1,000	21,143	0,000
VK (n=262)	-0,505	0,848		

Anmerkung: HK = handlungsbasierte Kompetenz; VK = verstehensbasierte Kompetenz

Die Verteilung der Personenparameter in der Wright Map (Abbildung 5.2.19) zeigt bereits, dass sich Lage und Ausprägung der zwei Kompetenzstrukturen unterscheiden. Dieser Unterschied ist signifikant ($t = 21,143$; $df = 261$; $p < ,001$) und die Effektstärke liegt mit $d = 1,39$ in einem sehr hohen Bereich. Bei einem direkten Vergleich der beiden Kompetenzskalen ist das Leistungsvermögen der Auszubildenden im handlungsbasierten Kompetenzbereich folglich über eine Standardabweichung höher ausgeprägt als im verstehensbasierten Kompetenzbereich. Auch im Hinblick auf dieses Ergebnis ist anzumerken, dass aufgrund der unterschiedlichen Skalenbreite nicht direkt auf Leistungsunterschiede geschlos-

sen werden kann. Ein Vergleich zwischen den zwei Kompetenzskalen zeigt jedoch, dass für die Mehrzahl der Probanden die Wahrscheinlichkeit handlungsbasierte Simulationsaufgaben eines höheren Schwierigkeitsgrades zu lösen größer ist als die Lösungswahrscheinlichkeit bei verstehensbasierten Anwendungsaufgaben mit vergleichbarer Schwierigkeit. Werden Analysen auf Basis der Differenzwerte der Personenfähigkeiten (ΔWLE ; Abbildung 5.2.25) durchgeführt, ergibt sich, dass die Mehrzahl der Probanden höhere Leistungen bei der Bewältigung handlungsbasierter Simulationsaufgaben (77,1 %) als bei der Bewältigung verstehensbasierter Anwendungsaufgaben erbringt (ausgedrückt in Logits). Nur drei Probanden zeigen im verstehensbasierten Kompetenzbereich eine bessere Performanz.

Abbildung 5.2.25 Differenzen der Personenfähigkeiten in den Kompetenzskalen (ΔWLE)

Performanz ($WLE_{HK} - WLE_{VK}$)	Logit-Kriterium (ΔWLE)	Anzahl	Prozent (gültig)	Prozent (kumuliert)
$WLE_{HK} < WLE_{VK}$	-0,51	3	1,1	1,1
$WLE_{HK} \approx WLE_{VK}$	$\pm 0,50$	57	21,8	22,9
$WLE_{HK} > WLE_{VK}$	+0,51	202	77,1	100,0

Anmerkung: WLE_{HK} = Personenschätzer der handlungsbasierten Kompetenz; WLE_{VK} = Personenschätzer der verstehensbasierten Kompetenz

21,8 Prozent der Probanden zeigen in beiden Kompetenzbereichen vergleichbare Leistungen. Es kann angenommen werden, dass sich die unterschiedlichen Leistungen der Probanden bei der Zerlegung der Kompetenzskalen in qualitativ begründete Kompetenzniveaus widerspiegeln.

5.2.4.4 Hierarchische Ordnung: Stufen berufsfachlicher Kompetenz

Eine der zentralen Annahmen der Modelle der Rasch-Familie ist, dass sich die zu messenden Konstrukte hierarchisch stufen lassen (vgl. Brown, 2005, p. 22). Hierfür werden die Skalen in begründete Abschnitte zerlegt. Ein jeder Skalenabschnitt wird (1) über die Items, die auf ihm lokalisiert sind, und (2) über die entsprechend zugeordnete Personenfähigkeit beschrieben. Die hierarchische Stufung ist zugleich qualitativ wie empirisch zu begründen. Qualitativ müssen solche Kriterien gewonnen werden, mit denen Anforderungssituationen und damit die Itemschwierigkeiten systematisch variiert werden können. Aus Perspektive der Empirie stellen diese Kriterien Prädiktoren der Itemschwierigkeit dar.

Zur Ermittlung von Kompetenzstufen werden Handlungsmerkmale aus dem beruflichen Arbeitsumfeld verarbeitet, denen differenzierende Kraft im Hinblick auf die Charakterisierung von Anforderungssituationen zugeschrieben werden kann. Die zentrale Idee besteht darin, dass über geeignete Kriterien spezifische Annahmen über die Wirkung einzelner Schwierigkeitsparameter getroffen wer-

den können, um die Frage zu beantworten, wie sich Lösungsprozesse bei einer systematischen Variation des Anspruchsniveaus der Anforderungssituationen verändern. Bei der Entwicklung von Testaufgaben und der entsprechenden Scoring-Verfahren für die Testauswertung ist folglich darauf zu achten, dass die erarbeiteten Anforderungssituationen unterschiedliche Komplexitätsgrade aufweisen müssen: Sind sie so einfach, dass alle Probanden sie sofort oder nach kurzer Einarbeitung ausführen können, oder so anspruchsvoll, dass längere Lern- und Übungsphasen erforderlich sind? Zur Beantwortung dieser Fragen sind kohärente Beziehungen zwischen curricularen Anforderungen, fachspezifischen Ausdifferenzierungen und kognitionspsychologischen Modellen herzustellen (vgl. Wilson, 2005; Winther & Achtenhagen, 2008a; 2009b). Mögliche Kriterien, anhand derer der Anspruchsgehalt einer beruflichen Anforderungssituation variiert werden kann, sind: (1) funktionale Modellierung, (2) inhaltliche Komplexität und (3) die Art der kognitiven Beanspruchung (Abbildung 5.2.26, vgl. Kapitel 3.3). Abbildung 5.2.26 macht deutlich, dass die Anforderungssituationen durch ein Set von Items und dem erwarteten Antwortverhalten auf diese Items definiert sind; die Komplexitätsmerkmale geben folglich die qualitative Struktur des Antwortverhaltens vor. Damit lassen sich Antwortmuster identifizieren, die sich auf weniger oder stark komplexe Items im Test zurückführen lassen.

Abbildung 5.2.26 *Komplexitätsmerkmale von Anforderungssituationen*

Bewertung (Vierstufige Likert-Skala)	Funktionale Modellierung	Inhaltliche Komplexität	Art der kognitiven Taxonomierung
4	Herausragende Modellierungsleistung	Wissenschaftliche Prozeduren	Nutzen und Anwenden
3	Umfangreiche Modellierungsleistung	Theoriekonstruktion	Analysieren und Validieren
2	Teilweise Modellierungsleistung	Systemkohärenzen	Elaborieren und Verstehen
1	Inkorrekte/fehlende Modellierungsleistung	Isolierte Lern-/ Wissensinhalte	Abfragen und Wiedergeben

Anmerkung: 4 = höchste Komplexitätsstufe; 1 = geringste Komplexitätsstufe

Die Komplexitätsstufen, die anhand kognitiver Modelle theoretisch entwickelt werden konnten, sind in die Konstruktion der Anforderungssituationen des ALU-SIM-Assessments als Kriterien der Konstruktrepräsentation direkt eingeflossen. Dies ermöglicht es, die Skalen der berufsfachlichen Kompetenz unter Rückgriff auf die Charakteristik der Items in den verschiedenen Anforderungssituationen in inhalts- und kriterienbezogene Kompetenzstufen zu zerlegen. Hierzu wurden (1) die schwierigkeitsklassifizierenden Merkmale für jedes Item durch fünf Rater beurteilt, um sie anschließend (2) als Prädiktoren der Itemschwierigkeit in einer multiplen Regression zu berechnen.

Dem Rating ging eine Schulung der Rater voran, in der die Kriterien der Konstruktrepräsentation, das Design der einzelnen Items sowie die inhaltliche Be-

deutung der verschiedenen qualitativen Stufen der Beurteilungskriterien erläutert wurden. Zur Veranschaulichung der Übereinstimmung zwischen den Ratern wurde die Intraclass-Korrelation (two-way mixed; ICC [3;1]) jeweils für ein Kriterium über alle Items und Rater ermittelt. In Abbildung 5.2.27 ist die Zusammenfassung dargestellt.

Abbildung 5.2.27 Intraclass-Korrelationskoeffizient (ICC [3;1])

Kriterium der Konstruktrepräsentation (Rateranzahl = 5)	ICC [3;1]
Funktionale Modellierung	0,946
Inhaltliche Komplexität	0,818
Art der kognitiven Taxonomierung	0,877

Anmerkung: two-way mixed; absolute agreement (unjusted); single measure

Die Auswertung zeigt, dass der ICC [3;1] für alle 57 Items, die letztlich in den Test zur Messung berufsfachlicher Kompetenz aufgenommen wurden, im Bereich von 0.818 bis 0.946 liegen. Diese Werte sprechen für eine starke Übereinstimmung zwischen den Ratern. Die Annahmen über die unterschiedlichen Schwierigkeiten der Items lassen sich dazu nutzen, den empirisch geschätzten Itemparameter zu prognostizieren. Hierzu werden im Rahmen einer multiplen Regressionsanalyse aus den drei Kriterien der Konstruktrepräsentation, die jeweils vier Ausprägungen aufweisen, neun operationale Itemparameter (Dummy-Variablen) und drei Referenzkategorien (funktionale Modellierung, inhaltliche Komplexität, kognitive Taxonomierung mit jeweils vier Abstufungen). Das Ergebnis der Regressionsanalyse – abgebildet über die unstandardisierten Regressionsgewichte – ist in Abbildung 5.2.28 dargestellt. Vier Prädiktoren sind hoch bedeutsam für die Vorhersage der Itemschwierigkeit.

Abbildung 5.2.28 Unstandardisierte Regressionsgewichte als Prädiktoren der Itemschwierigkeit

Adj. R ² = 0,749	Regressions- gewicht	Standard- fehler	Beta- gewicht	p
Konstante	-1,479	0,202	---	0,000
Umfangreiche (broad) Modellierungsleistungen	2,128	0,334	0,654	0,000
Analysieren und Validieren (analyze and reasoning)	1,293	0,327	0,362	0,000
Teilweise (incomplete) Modellierungsleistungen	0,840	0,246	0,256	0,001
Systemkohärenz (system coherence)	0,756	0,296	0,228	0,013

Anmerkung: n = 57 (Anzahl der Testitems für beide Kompetenzdimensionen)

Im Hinblick auf die funktionale Modellierung sind Anforderungssituationen, die von den Lernenden eine eigenständige Übersetzung der Situation in die not-

wendigen Schritte des Lösungsprozesses verlangen, besonders anspruchsvoll. Zwei Stufen der funktionalen Modellierung zeigen sich vor diesem Hintergrund im Regressionsmodell signifikant. Besonders stark ist die Itemschwierigkeit von der Notwendigkeit umfangreicher Modellierungsleistungen ($\beta = 0,654$; $p = 0,000$) abhängig. Doch bereits teilweise Modellierungsanforderungen ($\beta = 0,256$; $p = 0,001$) beeinflussen den Anspruchsgehalt eines Items. Die Bedeutung der inhaltlichen Komplexität für die Itemschwierigkeit wird im Prädiktor „Systemkohärenzen“ ($\beta = 0,228$; $p = 0,013$) deutlich. Items, mit denen Zusammenhänge zwischen verschiedenen betriebswirtschaftlichen Konzepten abgebildet werden sollen, sind für die Probanden folglich schwieriger zu lösen. Der kognitive Prozess wird durch den Prädiktor „Analysieren und Validieren“ ($\beta = 0,362$; $p = 0,000$) bedeutsam für die Komplexität einer betrieblichen Anforderungssituation. Die vier Merkmale der Itemschwierigkeit erklären zusammen 74,9 Prozent der Varianz ($F = 45,021$; $p = 0,000$).

In Anlehnung an Hartig (2007) lassen sich die unstandardisierten Regressionsgewichte als Rangwerte der Itemschwierigkeit interpretieren. Ein Item, für das ein Kriterium der Konstruktrepräsentation/Schwierigkeitsmerkmal geratet wurde, ist um den Betrag des Regressionsgewichts (gemessen in Logits) schwieriger als ein Item, das dieses Schwierigkeitsmerkmal nicht trägt. Die Summe der Logits stellt folglich die erwartete Schwierigkeit eines Items im Hinblick auf die schwierigkeitsbeschreibenden Merkmale dar. Der Wert, bei dem ein oder mehrere als signifikant identifizierte Schwierigkeitsmerkmale erstmalig systematisch auftreten, kann als untere Schwelle einer Kompetenzstufe interpretiert werden. Für die Skalen zur Erfassung berufsfachlicher Kompetenz können als Ergebnis der multiplen Regression folglich vier Kompetenzstufen ermittelt werden, die sich als „Kaufmännisches Grund- und Regelwissen“, „Kaufmännisches Handlungswissen“, „Kaufmännisches Analysewissen“ und „Kaufmännisches Entscheidungswissen“ umschreiben lassen. In Abbildung 5.2.29 sind die Kompetenzstufen über die Logit-Grenzen definiert; zusätzlich zu den Schwellenwerten ist die prozentuale Zuordnung der Auszubildenden auf die einzelnen Kompetenzstufen angegeben. Der Anteil der Auszubildenden in den jeweiligen Kompetenzstufen im Hinblick auf die zwei verschiedenen Kompetenzdimensionen unterscheidet sich deutlich: Wenn die Kompetenzstufe II: Kaufmännisches Handlungswissen, als Anspruch des Arbeitsmarktes an Absolventen des Dualen Systems in Anlehnung an die Vorgaben anzusehen ist, können im Bereich der handlungsbasierten Kompetenz 22,6 Prozent und im Bereich der verstehensbasierten Kompetenz 78,2 Prozent der getesteten Auszubildenden diesen Anspruch nicht adäquat erfüllen. Die Ergebnisse im handlungsbasierten Kompetenzbereich sind damit weit positiver ausgefallen, als dies vor dem Hintergrund vergleichbarer Studien hätte angenommen werden können; die Ergebnisse im verstehensbasierten Kompetenzbereich sind eher erwartungskonform und replizieren in hohem Maße u. a. die Befunde des ULME III-Projekts (Untersuchung von Leistungen, Motivation und Einstellungen in der beruflichen Bildung; Lehmann & Seeber, 2007; vgl. auch TOSCA: Köller, Watermann, Trautwein & Lüdtke, 2004). Die Befunde zeigen, dass die Lei-

stungen z. B. im Ausbildungsberuf Industriekaufmann/Industriekauffrau tendenziell eine bimodale Verteilung mit einem Gipfel im unteren Leistungsbereich und einem flachen Kurvenverlauf zum oberen Leistungsspektrum hin aufweisen. Die Leistungsschwächen beziehen sich dabei insbesondere auf Testaufgaben, die das Verstehen und Interpretieren ökonomischer Beziehungen zum Inhalt haben (Lehmann & Seeber, 2007, S. 140; hierzu auch Seeber, 2008).

Abbildung 5.2.29 Zuordnung der Auszubildenden auf den Kompetenzstufen

Kompetenzstufe	Stufen- schwelle	Anteil der Auszubildenden auf den Stufen der handlungsbasierten Kompetenz (in %)	Anteil der Auszubildenden auf den Stufen der verstehensbasierten Kompetenz (in %)
Unter Kompetenzstufe I	-1,479	3,10	35,50
Kompetenzstufe I: Kaufmännisches Grund- und Regelwissen	-0,723	19,50	42,70
Kompetenzstufe II: Kaufmännisches Handlungswissen	0,117	52,70	21,40
Kompetenzstufe III: Kaufmännisches Analysewissen	1,410	23,30	0,40
Kompetenzstufe IV: Kaufmännisches Entscheidungswissen	3,538	1,50	---

Auch für die vorliegende Studie zeigt sich, dass ein Großteil der Auszubildenden (42,70 Prozent) zwar über kaufmännisches Grund- und Regelwissen verfügt, das jedoch zu unflexibel ausgebildet ist, um es variabel im Rahmen beruflicher Anforderungssituationen einsetzen zu können. Anzumerken ist in diesem Zusammenhang, dass das Kompetenzniveaumodell für den Bereich der verstehensbasierten Kompetenz nur bedingt als befriedigend angesehen werden kann; es hat für ein Drittel der Probanden keine Erklärungskraft. Für den Testbereich der handlungsbasierten Kompetenz hat die Graduierung der Fähigkeit eine größere Relevanz. Insbesondere fällt positiv auf, dass nur drei Prozent der Auszubildenden unterhalb der Kompetenzstufe I liegen und dass 52,7 Prozent die Kompetenzstufe II, die als Mindestanspruchsstufe des Arbeitsmarktes definiert werden kann, und weitere 24,80 Prozent die darüber liegenden Kompetenzstufen erreichen konnten. Nachfolgend werden die einzelnen Kompetenzstufen definiert und Beispielitems zur inhaltlichen Veranschaulichung präsentiert.

Beispiel für Kompetenzstufe I: Kaufmännisches Grund- und Regelwissen

Die Kompetenzstufe I: Kaufmännisches Grund- und Regelwissen, wird vorrangig über den Faktor der inhaltlichen Komplexität definiert. Der schwierigkeitsbe-

schreibende Faktor „Systemkohärenz“ weist das geringste Regressionsgewicht auf, das ausgehend von der Konstanten den unteren Schwellenwert der Kompetenzstufe I: Kaufmännisches Grund- und Regelwissen, definiert. Auszubildende auf dieser Kompetenzstufe verfügen über grundlegende ökonomische Begriffs- und Konzeptstrukturen und können Systemkohärenzen (→ Prädiktor der Itemschwierigkeit) herstellen. Hierzu benötigen die Auszubildenden jedoch Hinweise in Form von Modellierungshilfen der Anforderungssituation, da sie nur so in der Lage sind, auf ihre Wissensbestände und prozeduralen Fertigkeiten zurückzugreifen. Folgendes Beispielitem repräsentiert die Kompetenzstufe I für die Skala der verstehensbasierten Kompetenz (Item 20 in der Wright Map; vgl. Abbildung 5.2.19):¹²

Beispielitem Kompetenzstufe I (Itemschwierigkeit: -0,585 Logits)

Aufgabe 2
Die ALUSIM GmbH hat am 29. November 2010 an Ihre Stammkunden eine Liste mit Restbeständen der Sparte Industrieprodukte versendet.
Diese Restbestände wurden den Kunden in einem beiliegenden Schreiben „mit einer Lieferung innerhalb von zwei Wochen“ und „mit einer Vergünstigung von 50 %“ angeboten. Geben Sie in den nachfolgenden Fällen bitte jeweils an, ob ein Kaufvertrag zustande gekommen ist und begründen Sie Ihre Meinung kurz.

2.1
Am 01. 12. 2010 ruft Frau Schridde, Delter Farben GmbH, an und gibt ihre Bestellung ab.

- ☒ Kaufvertrag ist zustande gekommen
- ☐ Kaufvertrag ist nicht zustande gekommen

Begründung:

Das Item bezieht sich auf rechtlich-organisatorische Fragen des Kaufvertrags auf Basis eines verbindlichen Angebots. Dieses Item ist typisch für die Skala der verstehensbasierten Kompetenz. Zur Beantwortung werden deklarative Wissensbestände benötigt, die miteinander in Beziehung zu setzen sind: Liegt ein verbindliches Angebot vor? Welche Fristen gelten für die Annahme eines verbindlichen Angebots? Welche Konsequenzen ergeben sich im Hinblick auf den Kaufvertrag? Der Itemfit des Beispiels ist mit einem weighted MNSQ von 1,06 und einer Itemdiskrimination von 0,35 sehr gut. Das gewählte Beispielitem weist einen Itemschwierigkeitsparameter von -0,585 Logits auf. Dieser liegt leicht unter-

¹² Alle Items der Skala der verstehensbasierten Kompetenz beziehen sich auf die Sequenzen der Simulation insgesamt: Für das exemplarische Beispielitem wurde als Begründung akzeptiert, dass – vorbehaltlich der Verfügbarkeit – der Kaufvertrag zustande gekommen sei, da auf das verbindliche Angebot binnen einer Woche reagiert wurde.

versendeten Preisliste relevant sein könnte, und sie müssen (2) begründet darüber entscheiden, unter welchen Bedingungen welcher Preis tatsächlich in das Bestellsystem einzugeben ist. Das Item erfüllt mit einem weighted MNSQ von 1,10 und einer Itemdiskrimination von 0,20 die Anforderungen zur Aufnahme in den Test der handlungsbasierten Kompetenz. Der Schwierigkeitsparameter ist mit 0,206 Logits geringer als der Populationsmittelwert der handlungsbasierten Kompetenzskala (mean = 0,750; $\sigma^2 = 0,612$). Insgesamt wurde das Beispielitem von 214 Personen mit einer Lösungsquote von 59,81 Prozent bearbeitet.

Beispiel für Kompetenzstufe III: Kaufmännisches Analysewissen

Die Kompetenzstufe III: Kaufmännisches Analysewissen, definiert sich in Abgrenzung zur Kompetenzstufe II hauptsächlich über den Grad der kognitiven Beanspruchung. Der Prädiktor „Analysieren und Validieren“ weist ein Regressionsgewicht von 1,293 auf, das addiert zur Konstanten und den davorliegenden Kompetenzstufen den unteren Schwellenwert von 1,41 Logits für die Kompetenzstufe III ergibt. Von Auszubildenden auf dieser Kompetenzstufe wird erwartet, dass sie über die Fähigkeit verfügen, die Plausibilität einer berufsspezifischen Lösung analysieren und validieren zu können. Die kognitive Taxonomierung stellt folglich für diese Kompetenzstufe den zentralen Schwierigkeitsparameter dar. Dies setzt eine schlussfolgernde Auslegung von Wissen und damit Transferfähigkeit voraus. Das nachfolgende Beispielitem illustriert diese Fähigkeitsstruktur:

Beispielitem Kompetenzstufe III (Itemschwierigkeit: 1,730 Logits)

Einkauf -> Bestellsystem

Daten Katalog Lieferanten Einkauf Extras Hilfe

Leistungen erfassen | LE-Kontrolle

Bestellnummer: 35000 3219 | LE-Datum: 03.12.2012 | Lieferant: 10002 | Name: Colouflex GmbH

Kurztext: Farbsondermischung Piney Grove Green Metallic

Leistungen

Artikel-Nr.	Produktbezeichnung	Menge	Preis pro Mengeneinheit	LE-Einheit
				KW
				KW
				KW
				KW
				KW
				KW
				KW

Erläuterung:
Am Ende einer Kalenderwoche (KW) werden Kontrolllisten der eingegangenen Bestellungen zur abschließenden Überprüfung erstellt.
Die aktuelle Bestellung erhält in der Eingangskontrolle die Bestellnummer 35000 3219.

Die Auszubildenden haben im Vorfeld verschiedene Preiskalkulationen durchgeführt. Sie müssen nun in einem kognitiven Validierungsprozess ermitteln, ob der von ihnen berechnete Einstandspreis tatsächlich relevant für die Eingabe im ERP-System ist. Es wird erwartet, dass die Auszubildenden erkennen, dass hier der Listen-/Angebotspreis der Farbe zu verarbeiten ist und nicht der Einstandspreis der Preiskalkulation.

SENDUNG versenden

Die ALUSIM GmbH benötigt eine Druckfarbe, die bislang nicht im Sortiment enthalten ist. Die Auszubildenden haben in diesem Zusammenhang Einstandspreise im Rahmen einer vergleichenden Preiskalkulation errechnet. Nachdem die Entscheidung für einen Lieferanten gefallen ist, ist die Einkaufsorder im ERP-System zu erfassen. Die Schwierigkeit besteht darin, dass die Auszubildenden in einem kognitiven Validierungsprozess ermitteln müssen, ob der von ihnen im Vorfeld berechnete Einstandspreis oder der Listen-/Angebotspreis zu verarbeiten ist (Item 43 in der Wright Map; vgl. Abbildung 5.2.19). Das Item fittet die Skala der handlungsbasierten Kompetenz mit einem weighted MNSQ von 1,03 und einer Itemdiskrimination von 0,24 sehr gut. Die Itemschwierigkeit liegt mit 1,730 Logits deutlich oberhalb der mittleren handlungsbasierten Kompetenz der Probanden. Von 215 Personen, die dieses Item bearbeitet haben, konnten nur 23 die korrekte Lösung finden (Lösungsquote: 10,70 Prozent).

Beispiel für Kompetenzstufe IV: Kaufmännisches Entscheidungswissen

Verfügen Auszubildende über Kaufmännisches Entscheidungswissen (Kompetenzstufe IV), sind sie in der Lage, im Sinne eines umfassenden Managementprozesses, die berufliche Anforderungssituation im Hinblick auf die Problemformulierung, die Zielsetzungen, die Lösungsmöglichkeiten und die resultierenden Entscheidungen umfassend zu modellieren (Laux, 2007). Die Kompetenzstufe IV enthält nur Items des handlungsbasierten Kompetenzbereichs. Items, die dieser Kompetenzstufe zugeordnet sind, lassen offen, welche Bearbeitungsschritte zur Lösungsfindung gegangen werden müssen und welchen Umfang die Lösung letztlich haben soll. Diese Charakteristik eines unpräzise formulierten Anforderungsgehalts wird in der Kreativitätsforschung beispielsweise als blind pool bezeichnet (Simonton, 2003). Die Schwierigkeit solcher Situationen liegt darin, zunächst auf Basis bestehender Wissensbestände und prozeduraler Fähigkeiten eine so umfassende Modellierung der Situation vorzunehmen (→ Prädiktor der Itemschwierigkeit), dass Lösungsmöglichkeiten ersichtlich werden. Die notwendigen Modellierungsschritte können dabei entsprechend der Situation sehr unterschiedlich beschaffen sein. Vor dem Hintergrund betrieblicher Arbeitsprozesse sind beispielsweise die Sammlung, Komprimierung und Aufbereitung von Informationen, Effizienzkalkulationen einzelner Arbeitsprozessschritte oder die Entscheidung für die optimale Kombination von Arbeitsmitteln möglich. Das nachfolgende Itembeispiel illustriert für die Skala der handlungsbasierten Kompetenz, wie eine Anforderungssituation zunehmend komplexer werden kann und wie dadurch die Lösungschancen der Auszubildenden mit einem geringeren Fähigkeitslevel erheblich reduziert werden können.

Die Auszubildenden sollen auf der Basis von Angeboten und weiteren Zusatzinformationen eine vergleichende Lieferantenauswahl durchführen. Die Hintergründe des Lieferantenvergleichs sind den Auszubildenden bekannt; zwei Angebote für den Vergleich wurden von den Auszubildenden selbst eingeholt. Die Schwierigkeit liegt darin, die extrem große Menge an Einzelinformationen so zu komprimieren, dass alle notwendigen Daten für einen Lieferantenvergleich vor-

liegen: Die Informationsquellen unterscheiden sich, die Lieferbedingungen sind verschieden kodiert, die Abpackungen des Produkts variieren, unterschiedliche Währungen sind zu berücksichtigen etc. Auch das Arbeitsprodukt wird nicht präzise formuliert; als Arbeitsmittel wird ausschließlich eine Excel-Datei ohne Inhalt zur Verfügung gestellt (Item 29 in der Wright Map; vgl. Abbildung 5.2.19). Das Item (weighted MNSQ = 0,95; $d = 0,29$) liegt mit einer Itemschwierigkeit von 3,602 Logits deutlich oberhalb des mittleren Populationsmittelwerts der handlungsbasierten Kompetenzdimension (mean = 0,750; $\sigma^2 = 0,612$). Die Lösungsquote des Items beträgt 6,32 Prozent; 190 Probanden haben die Bearbeitung dieses Items aufgenommen.

Beispielitem Kompetenzstufe IV (Itemschwierigkeit: 3,602 Logits)

Die Informationen in der Lieferantendatei des ERP-Systems des Unternehmens stellen nur einen Ausschnitt aller notwendigen Informationen dar. Insgesamt müssen für den Lieferantenvergleich acht verschiedene Informationsquellen ausgewertet werden.

	Kurz	Lieferbedingungen	Zahlungsbedingungen	Qualität	Termintreue	Ökoaudit (EMAS)
	Büro	ab Werk	008 20 045	1	2	2
	Colour	frei Haus	014 20 030	1	1	2
D	Draht	frei Haus	008 20 020	2	1	2
	Farb	ab Werk	007 20 030	1	1	2
KG	Kunst	ab Werk	000 00 014	1	3	2
	Life	ab Werk	007 20 030	1	1	1
mbH	MPK	ab Werk	010 20 060	1	1	2
	Multi	frei Lager	008 20 045	1	1	2
	Papier	frei Haus	010 20 020	1	2	3
	Plast	frei Haus	000 00 045	1	2	1

Eingabe der Zahlungsbedingungen entsprechend DIN-FIBU:
008 20 045
= 8 Tage 2,0 % Skonto, 45 Tage netto Kasse

Qualität und Termintreue der Lieferanten sind wie Schulnoten vergeben:
1 = ausgezeichnet

Umfang unternehmerischer Verfahren und Maßnahmen, die die negativen Umweltauswirkungen betrieblicher Tätigkeit reduzieren.
Die Ökoaudit-Verordnung entspricht dem Environmental Management and Audit Scheme (EMAS).
Der Ökoaudit der Lieferanten ist wie Schulnoten vergeben:
1 = ausgezeichnet

alu s i m

Die Beispielitems illustrieren insgesamt, wie die Prädiktoren der Itemschwierigkeit als inhaltlicher Anker zur Beschreibung einzelner Kompetenzstufen interpretiert werden können. Im Umkehrschluss zeigen diese Beispiele jedoch auch, dass für eine Testkonstruktion im Vorfeld sehr elaborierte Analysen (1) der Anforderungsbedingungen, (2) der kognitiven Theorien und (3) der inhaltlichen sowie didaktischen Bedingungen durchzuführen sind. Hierbei ist es hilfreich, dass die neuen Assessmentmodelle der angewandten Kognitionswissenschaft Umgebungsanalysen direkt auf psychometrische Messmodelle beziehen (vgl. u. a. Mislevy, 2008; Wilson, 2008; Embretson & Gorin, 2001). Oder anders formuliert: Der Einsatz psychometrischer Modelle setzt eine inhaltliche, didaktische Vorarbeit im Hinblick auf die Itemkonstruktion zwingend voraus, um im Rahmen der Analyse die Ergebnisse sinnvoll interpretieren zu können – auch für die Interpretation ist fachdidaktische Expertise einzufordern.

5.2.4.5 Prädiktoren der Personenfähigkeit

Insbesondere im Hinblick auf ein international vergleichendes Large-Scale Assessment können Effekte von Interesse sein, mit denen die Leistungsfähigkeit der Auszubildenden auf individuelle und institutionale Bedingungen bezogen werden sollen. Hierfür sind Analysen und Auswertungen im Mehrebenenmodell vorzunehmen. Diese Modellierung setzt jedoch für stabile und aussagekräftige Ergebnisse eine sehr große Stichprobe voraus – diese könnte im Rahmen eines VET-LSA realisiert werden; für die vorliegende Studie ist der Stichprobenumfang allerdings zu klein ($N = 264$), so dass im Folgenden auf Basis von Regressionsmodellen nur Tendenzen und mögliche Beziehungen thematisiert werden. Im one-group posttest-only-Design der vorliegenden Studie wurden spezifische Hintergrundinformationen der Auszubildenden erhoben. Hierzu zählen allgemeine Angaben zu Alter, Geschlecht und Bildungsabschluss sowie die Unternehmensgröße und Branche des Ausbildungsbetriebes. Um zu klären, inwieweit individuelle und institutionelle Faktoren die Leistungsfähigkeit prognostizieren können, wurden für die Faktoren Bildungsabschluss, Betriebsgröße und Branche Dummy-Variablen erzeugt. Zusätzlich wurden schulische Leistungen in Form der Noten in den Fächern Deutsch und Englisch erfasst; ein Aufnehmen der Mathematiknote war leider aufgrund des Lernfeldkonzeptes nicht möglich. Aus diesen Individual- und Institutionenmerkmalen erweist sich bei Annahme einer fünfprozentigen Irrtumswahrscheinlichkeit nur die allgemeine Hochschulreife als Prädiktor der Leistungsfähigkeit. Die Ergebnisse sind in Abbildung 5.2.30 dargestellt.

Abbildung 5.2.30 Allgemeine Hochschulreife als Prädiktor der Personenfähigkeit

	β	p	F	R ²
abhängige Variable: Personenparamater (WLE) im handlungsbasierten Kompetenzbereich	0,301	0,025	5,048 (0,002)	0,055
abhängige Variable: Personenparamater (WLE) im verstehensbasierten Kompetenzbereich	0,309	0,023	2,319 (0,046)	0,026

Es zeigen sich darüber hinaus Tendenzen, dass Auszubildende in Großbetrieben über bessere Leistungen sowohl im handlungsbasierten ($\beta = 0,132$; $p = 0,055$) als auch im verstehensbasierten Kompetenzbereich ($\beta = 0,131$; $p = 0,090$) verfügen. Die praktische Bedeutsamkeit aller ermittelten Effekte ist im Hinblick auf die vorliegende Stichprobe jedoch sehr gering, so dass eine weitergehende Interpretation nicht erfolgt.

5.2.4.6 Schlussfolgerungen

Mit der vorliegenden Studie wurde vorrangig der Anspruch verfolgt, Aussagen zur Konstruktvalidität von Simulationsaufgaben zu gewinnen. Hierbei ging es darum, bezogen auf die Erfassung von berufsfachlicher Kompetenz eine Alterna-

tive zu Papier-und-Bleistift-Tests zu entwickeln. Die Analyse der Ergebnisse zeigt, dass mit dem ALUSIM-Assessment verschiedene Validitätskriterien erfüllt werden können.

Hinsichtlich der *Konstruktvalidität* kann geschlussfolgert werden, dass die auf theoretischen Analysen basierenden Dimensionen berufsfachlicher Kompetenz (1) in einem Test umgesetzt werden konnten und (2) sich für das Sample der Auszubildenden im Beruf des Industriekaufmanns/der Industriekauffrau empirisch bestätigen lassen. Die Analyse zeigt, dass die Daten das zweidimensionale Modell gut fitten und dass sowohl die handlungsbasierte als auch die verstehensbasierte Kompetenzdimension entsprechend des Antwortverhaltens der Auszubildenden nach verschiedenen Kompetenzstufen geordnet werden kann. Hierbei ist festzustellen, dass die handlungsbasierten Simulationsaufgaben die Logit-Skala umfassend repräsentieren, während über die verstehensbasierten Anwendungsaufgaben der Test enger konstruiert ist – mit den entsprechenden Konsequenzen für die Graduierung dieser Kompetenzstruktur. Die Fit-Statistik der Itemparameter ist symmetrisch, ohne bemerkenswerte Ausreißer, verteilt und lässt damit auf einen guten Fit der Daten im Hinblick auf die Modellannahmen schließen. Ein Vergleich zwischen den Personenschätzern macht deutlich, dass insgesamt eine höhere Performanz der Auszubildenden im Bereich der handlungsbasierten Kompetenz gemessen werden konnte. Die Personenfähigkeiten in beiden Kompetenzdimensionen sind annähernd normalverteilt, so dass davon ausgegangen werden kann, dass sich Personen auf unterschiedlichen Leistungsstufen verorten lassen. Diese Einschätzung wird durch die guten Reliabilitätskennwerte gestützt. Die *Reliabilität* beider latenter Fähigkeitsstrukturen ist sehr gut, wenn zusätzlich der geringe Standardfehler der Messung in Bezug auf die Konstruktlänge berücksichtigt wird. Diese Reliabilitätsschätzer erlauben es, die Auszubildenden auf verschiedenen Kompetenzniveaus zu lokalisieren. Insgesamt konnten vier Niveaustufen ermittelt werden: Kaufmännisches Grund- und Regelwissen, Kaufmännisches Handlungswissen, Kaufmännisches Analysewissen sowie Kaufmännisches Entscheidungswissen. Inter-Rater-Reliabilitäten wurden ausschließlich bei der Bewertung der Schwierigkeitsstufen, nicht jedoch bei den Itemantworten berücksichtigt. Der ICC [3;1] deutet auf eine hohe Übereinstimmung der Rater im Hinblick auf die Kriterien der Konstruktrepräsentation hin.

Das ALUSIM-Assessment wird als inhaltlich valide bewertet. Zur Stützung der *Inhaltsvalidität* wurden zwei zentrale Schritte gegangen: (1) Zur Entwicklung der Testitems wurden umfassende Beobachtungen an Arbeitsplätzen von Industriekaufleuten durchgeführt, um berufsspezifische Anforderungssituationen zu konstruieren. Der Anspruch bestand einerseits darin, auf der Ebene der inhaltlichen Ausgestaltung den realen betrieblichen Arbeits- und Geschäftsprozessen durch Authentizitätsmodellierungen gerecht zu werden. Hierzu zählt auch, dass Arbeitsmittel in Entsprechung zur betrieblichen Praxis für die Testbearbeitung eingesetzt werden und dass die konstruierten Anspruchssituationen auf einem umfassend modellierten Datensatz, mit dem das simulierte Unternehmen exakt beschrieben werden kann, basieren. Andererseits wurde das Ziel verfolgt, auf der

Ebene der Person auch solche Fähigkeitsstrukturen erheben zu können, die sich auf Handlungsprozesse beziehen lassen. Dies umfasst für eine als authentisch geltende Situationsbewältigung insbesondere auch die notwendigen Interaktionen mit Kollegen, Kunden und Lieferanten. Aspekte wie Informationsgehalt der Kommunikation, Höflichkeit oder auch die Kompatibilität mit den Unternehmenszielen lassen sich als Unterkategorien der Sozialkompetenz auf diesem Wege beschreiben (tätigkeitsspezifische Validität). (2) Die Anforderungssituationen und die zugehörigen Testitems wurden betrieblichen Ausbildern sowie Berufsschulexperten vorgelegt und von diesen als inhaltlich valide klassifiziert. Die unterschiedlichen Itemschwierigkeiten und die Fähigkeiten der Auszubildenden, verschieden schwierige Items zu lösen, können als Belege dafür genommen werden, dass die Testitems kaum floor- und ceiling-Effekte provozieren, so dass das ALUSIM-Assessment sowohl für leistungsschwächere als auch für leistungsstärkere Auszubildende eingesetzt werden kann.

Hinsichtlich der *externen Validität* ist festzustellen, dass das Instrument bezogen auf die handlungsbasierte Komponente bislang nicht direkt mit anderen Assessments verglichen worden ist bzw. verglichen werden konnte. Für den Bereich der verstehensbasierten Kompetenz lassen sich jedoch Parallelen zu den Ergebnissen des Projekts „Untersuchung von Leistungen, Motivation und Einstellungen in der beruflichen Bildung“ (ULME III; Lehmann & Seeber, 2007; Seeber, 2008) herstellen. Beide Studien kommen übereinstimmend zu dem Ergebnis, dass Auszubildende im Beruf des Industriekaufmanns/der Industriekauffrau insbesondere bei Testaufgaben, die das Verstehen und Interpretieren ökonomischer Beziehungen zum Inhalt haben, Leistungsschwächen aufweisen und dass die Leistungsfähigkeit generell auf einem niedrigen Kompetenzniveau konzentriert ist. Für den Bereich der handlungsbasierten Kompetenz liegen bislang keine vergleichbaren Studien für den Bereich der kaufmännischen Ausbildung vor, die eine Prüfung der externen Validität erlaubten. Hier besteht ein Forschungsdesiderat, das insbesondere vor dem Hintergrund der geringen Korrelation zwischen handlungsbasierten und verstehensbasierten Fähigkeitsstrukturen von besonderem Interesse sein kann.

Ein weiteres Validitätskriterium bezieht sich auf die Frage, in welchem Maße der Test erlernbar ist (*consequential validity*). Das ALUSIM-Assessment versteht sich als Fähigkeitstest und weniger als standardisierter Leistungstest. Insbesondere im Hinblick auf Fragen des High-stakes Testing ist jedoch zu prüfen, ob die Ziele des Assessments dadurch unterlaufen werden könnten, wenn gezielt auf das Instrument vorbereitet würde. Dies könnte zur Folge haben, dass die Auszubildenden im Test eine höhere Leistung zeigen, diese jedoch nicht mit einer höheren latenten Fähigkeit korrespondiert (vgl. Brown, 2005, p. 29). Das über handlungsbasierte Simulationsaufgaben gestaltete ALUSIM-Assessment erschwert aufgrund der komplexen Struktur ein „teaching to the test“ und wirkt damit weitgehend einem Memorieren spezifischer Items entgegen. Hinzukommt, dass in den Simulationsaufgaben direkte Handlungen und Entscheidungen von den Auszubildenden eingefordert werden und größtenteils auf theoretische Erläuterungen ver-

zichtet wird. Für den Testbereich der verstehensbasierten Kompetenz ist charakteristisch, dass sich die Testitems direkt auf die Anforderungen in der vorherigen Simulation beziehen, hier jedoch die Nähe zu traditionellen Leistungstests deutlich größer ist.

Insgesamt lässt sich schlussfolgern, dass es die vorgelegten und evaluierten Anforderungssituationen und Testitems des ALUSIM-Assessment erlauben, die berufsfachliche Kompetenz angehender Industriekaufleute sowohl mit ihrer (eher betriebsbezogenen) handlungsbasierten Kompetenz als auch mit ihrer (eher schulbezogenen) verstehensbasierten Kompetenz objektiv, reliabel und valide abzubilden und zugleich begründet nach Kompetenzstufen zu differenzieren.

6 Entwicklungsperspektiven der Kompetenzmodellierung und Kompetenzmessung in der beruflichen Bildung

Im Rahmen dieser Arbeit wurden auf Basis kognitiver Theorien der Bedeutungsumfang sowie die Struktur des Kompetenzbegriffs in der beruflichen Bildung operationalisiert. Ein wesentliches Kriterium der Operationalisierung war die Entscheidung, das in der beruflichen Bildung vertretene Kompetenzverständnis einer umfassenden Handlungskompetenz in eine messbare Modellierung zu übersetzen. Hiermit sind zwei Einschränkungen verbunden: (1) Berufliche Handlungskompetenz wird nicht als eindimensionales komplexes System erfasst, sondern über einzelne klar definierte Konstrukte, die Elemente der Literalität sowie des Berufsspezifischen beinhalten. Dieser Schritt ist zum gegenwärtigen Zeitpunkt notwendig, da sowohl die Methoden der angewandten Statistik als auch die zur Verfügung stehenden Messverfahren eine komplexere Messung (noch) nicht zulassen. (2) Die Modellierung bezieht sich auf den Bereich der kaufmännischen beruflichen Bildung und lässt derzeit keine Annahmen im Hinblick auf die Übertragbarkeit auf andere berufliche Bereiche zu. Diese Einschränkung ist durch den über das Domänenmodell vorgegebenen inhaltlichen Bezugsrahmen der Kompetenzmodellierung zwingend gegeben. Im Hinblick auf die Struktur, zentral jedoch als Grundlage für die fachdidaktisch akzentuierte Graduierung kaufmännischer Kompetenz, werden über die Inhalte und Eigenschaften einer Domäne Operationalisierungsvorgaben gemacht. Die Modellierungsschritte selbst sind hingegen für verschiedene berufliche Domänen sowie für unterschiedlich ausgestaltete Kompetenzauffassungen nutzbar.

Berufliche Kompetenzmodellierung erfordert demnach

- die Analyse der dem Kompetenzverständnis zugrundeliegenden kognitiven Struktur,
- die Zielformulierung im Hinblick auf die mit der Kompetenzmodellierung angestrebte Reichweite bzw. Transferwirkung des Kompetenzverständnisses,
- ein detailliertes konzeptionelles Verständnis der Domäne, das Inhalte, Methoden, Arbeitsweisen sowie fachliche Zugänge umfasst, sowie
- theoretisch begründete Annahmen und Kriterien der Konstruktrepräsentation, mit denen Effekte von Aufgabenmerkmalen auf das Antwortverhalten vorhergesagt werden können und die als Prädiktoren der Leistung die kontextuellen und personalen Anforderungen charakterisieren (vgl. Kapitel 6.1).

Berufliche Kompetenzmessung setzt vor dem Hintergrund der Anforderungen der Modellierung methodische Verfahren voraus, mit denen sich auf der Ebene der Erfassung von Kompetenz domänentypische Handlungsbedingungen und -abläufe widerspiegeln lassen und mit denen auf der Ebene des Messmodells Annahmen über den Zusammenhang von zu erfassender Kompetenz und situationalem Verhalten in der Testsituation berücksichtigt werden. Parallel zur Kompetenzmodellierung sind folglich sowohl diagnostische Instrumente zur Kompe-

tenzerfassung als auch psychometrische Ansätze zur Modellierung des (latenten) Kompetenzkonstruktes zu entwickeln (vgl. Kapitel 6.2).

Der in dieser Arbeit vertretene Ansatz der Kompetenzmodellierung und Kompetenzmessung in der beruflichen Bildung wird im Folgenden im Hinblick auf seine Erweiterungs- und Validierungsmöglichkeiten diskutiert. Dabei stehen insbesondere zwei programmatische Entwicklungsperspektiven im Fokus: die Konstruktion von kompetenzorientierten Lehr-Lernprozessen und deren Evaluation sowie der Aufbau von fachdidaktisch geprägten Aus- und Weiterbildungsangeboten zur Stützung der Lehrprofessionalität (vgl. Kapitel 6.3).

6.1 Forschungsdesiderate der beruflichen Kompetenzmodellierung

In die Modellierung von Kompetenzen fließen i. d. R. drei Fragestellungen ein: Die Frage nach der Dimensionalität und damit nach der Binnenstruktur der zu erfassenden Kompetenz, die Frage nach den Ausprägungen bzw. Stufen der Kompetenz, die mit dem zugrundeliegenden Messverfahren abgebildet werden können, sowie die Frage, wie sich die zu erfassende Kompetenz in Lern- und Entwicklungsprozessen verändert.

6.1.1 Struktur beruflicher Kompetenz

Die Frage nach der Dimensionalität der zu erfassenden Kompetenz ist theoretisch zu bearbeiten und hierbei zentral von den Vorstellungen und Entwicklungslinien der entsprechenden Fachdisziplin geprägt. In der beruflichen Bildung werden unter dem Oberbegriff der beruflichen Handlungskompetenz komplexe Fähigkeitsstrukturen subsumiert, die fachspezifische kognitive Dispositionen, personale Ressourcen sowie übergreifende Handlungs- und Arbeitsmethoden umfassen. Die Terminologie der beruflichen Kompetenzdiskussion verweist in diesem Zusammenhang (1) auf komplexe Handlungen und (2) auf vollständige Prozesse. Damit ist gemeint, dass Handlungskompetenz weder auf der Basis isolierter Leistungssituationen noch auf der Grundlage einzelner personenbezogener Leistungsmerkmale erfasst werden kann. Die Abbildung beruflicher Kompetenz erfordert hingegen Anforderungssituationen, die sowohl Arbeitsabläufe im Sinne beobachtbarer beruflicher Handlungen als auch wissensbasierte Entscheidungsprozesse als Grundlage der Handlungen beinhalten. Um Aussagen darüber zu treffen, wie Auszubildende in derart konstruierten komplexen Situationen agieren, ist darüber hinaus ein Modell davon zu entwickeln, welche personalen Dispositionen ein erfolgreiches Bearbeiten komplexer Anforderungssituationen stützen bzw. als Voraussetzung gegeben sein sollten. Während insbesondere die Forschungen im Zusammenhang mit der Konstruktion und Evaluation komplexer Lehr-Lernumgebungen in der beruflichen Bildung zahlreiche Ansatzpunkte für die Gestaltung komplexer Anforderungssituationen bereitstellen (vgl. hierzu u. a. Kremer, 2007; Achtenhagen, 2002; Bransford, Brown & Cocking, 2000), sind die personalen Ressourcen und internalen Bedingungen für erfolgreiches berufliches Handeln und Verstehen bislang selten in konkreten berufstypischen Anforderungssituationen

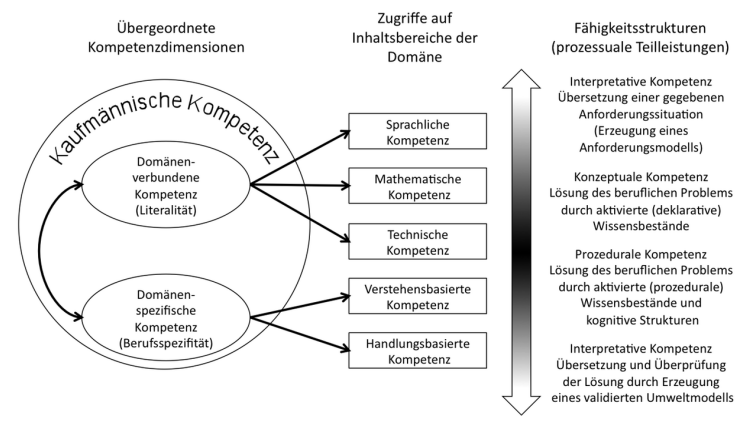
rungssituationen untersucht wurden (vgl. hierzu Winther, 2006). Die Dimensionalität des kaufmännischen Kompetenzkonzepts basiert daher auf der Ebene der personalen Ressourcen auf grundlegenden Annahmen der Kognitionspsychologie und verarbeitet auf der Ebene der Aufgaben berufstypische Situationsmerkmale. Beide Ebenen lassen sich unter Rückgriff auf die Ergebnisse der Studien zur kaufmännischen Kompetenz im domänenspezifischen Kompetenzbereich über handlungs- und verstehensbasierte Berufsspezifika (vgl. Kapitel 5.2.4) und im domänenverbundenen Bereich über sprachliche und mathematische Literalität (vgl. Kapitel 5.1.2 sowie 5.1.3) vor dem Hintergrund konzeptueller, prozeduraler und interpretativer Kompetenzstrukturen zusammenführen. Eine wesentliche Einschränkung der vorgestellten Studien ist, dass domänenspezifische und domänenverbundene Kompetenzen unabhängig voneinander untersucht wurden und die Ergebnisse nicht aufeinander bezogen werden können. Im heuristischen Modell der kaufmännischen Kompetenz stellt domänenverbundene Literalität die verschiedenen Zugänge zur kaufmännischen Domäne dar, die im Sinne des Kompetenzverständnisses von Gelman und Greeno (1989; Greeno, 1998) den Aufbau domänenspezifischer kognitiver Strukturen und Prozesse unterstützen (vgl. Kapitel 2.2.2; Pick-up-Schemes). Aus dieser Perspektive ist von einem direkten Zusammenhang von domänenverbundenen und domänenspezifischen Kompetenzen auszugehen. Dieser Zusammenhang konnte empirisch bislang über Partialkorrelationen im vollzeitschulischen kaufmännischen Eingangsunterricht nachgewiesen werden. Die Ergebnisse bestätigen, dass domänenverbundene Literalität die domänenspezifische Leistung beeinflusst (vgl. Kapitel 5.1.3.1).

Für den kaufmännischen Bereich geht es folglich darum, in welchem Maße die domänenspezifischen Kompetenzen von domänenverbundenen, inhaltlich übergreifenden, und allgemeinen Teilkompetenzen abhängig sind. Diese Frage stellt sich, da das Wissen und Können in berufsspezifischen Handlungen immer auch auf zuvor erworbene Kompetenzen zurückgreift (zurückgreifen müsste). Im bildungspolitischen Raum schlägt sich dieser Zusammenhang beispielsweise in Klagen von Betrieben nieder, nach denen Bewerber um einen Ausbildungsplatz zu geringe schulische Grundkenntnisse mitbrächten und von daher angemessene berufsspezifische Kompetenzen gar nicht erreichten – eine Auffassung, die durch die Diskussion um die PISA-Ergebnisse noch verstärkt wird (vgl. hierzu aber auch Rebmann, Tredop, Klattenhoff, Schulze & Wittrock, 2007). Von daher ist die Rolle der so genannten allgemeinen Inhalte in den beruflichen Schulen immer wieder Gegenstand des Nachfragens: Wie sind die Aufgaben von Deutsch, Mathematik, Englisch zu sehen? Als Inhalte zur Nachbesserung oder zur weiteren Förderung? Am Beispiel der elektrotechnischen Ausbildung konnte qualitativ gezeigt werden, dass sich vor allem mathematische Defizite als schwer überwindbare Barrieren der Fachkompetenzentwicklung erweisen (Nickolaus & Ziegler, 2005; Nickolaus, Gschwendtner & Geißel, 2008; für den kaufmännischen Bereich vgl. u. a. Lehmann & Seeber, 2007; Seeber, 2008). Für die Definition und Erfassung von domänenspezifischen Kompetenzen stellt sich daher die Frage, ob und – falls ja – in welchem Maße sich Zusammenhänge zwischen domänenspezifischen, domänenver-

bundenen und allgemeinen Kompetenzen finden ließen. Damit hängen zwei zentrale Probleme zusammen: Wie beeinflusst allgemeine und domänenverbundene Literalität den Auf- und Ausbau domänenspezifischer Kompetenzen? Wie wären – sofern die erste Frage beantwortet ist – allgemeine und domänenverbundene Kompetenzen, die als nicht zureichend angesehen werden, aufzubauen und zu fördern? Ein Ansatzpunkt, der in diesem Zusammenhang vertreten werden kann – und über die Organisation von Lernfeldern auch teilweise bereits formal vertreten wird – ist, dass dieses nicht separiert, sondern zugleich mit der Vermittlung domänenspezifischer Kompetenzen erfolgen sollte.

Im Hinblick auf die Kompetenzmodellierung ist der Zusammenhang von domänenverbundener Kompetenz (Literalität) und domänenspezifischer Kompetenz (Berufsspezifität) in die prozessuale Struktur der konzeptualen, prozeduralen und interpretativen Kompetenzdimensionen einzubinden. An dieser Stelle sei noch einmal betont, dass sich konzeptuale, prozedurale und interpretative Kompetenzen nicht in ihrer Komplexität unterscheiden, folglich also keine hierarchische Struktur angenommen wird (vgl. Abbildung 6.1). Dies impliziert, dass sich die verschiedenen Kompetenzstrukturen (1) ausschließlich über den Zugriff auf spezifische Inhaltsbereiche charakterisieren lassen (mathematische, sprachliche und technische Kompetenz sowie verstehensbasierte und handlungsbasierte Kompetenz) und dass sie (2) mögliche, sich unterscheidende Fähigkeitsstrukturen darstellen, die zur Anforderungsbewältigung einzusetzen sind (konzeptuale, prozedurale und interpretative Kompetenz).

Abbildung 6.1 Domänenverbundene und domänenspezifische Kompetenz



Die bislang erzielten Ergebnisse zeigen, dass im Hinblick auf die Literalität in kaufmännischen Handlungssituationen die Leistungsfähigkeit der Auszubildenden im Bereich der mathematischen Kompetenz stärker variiert als im Bereich

der sprachlichen Kompetenz (vgl. Kapitel 5.1.3). Mit Blick auf die Berufsspezifität kann festgehalten werden, dass Auszubildende im Ausbildungsberuf des Industriekaufmanns/der Industriekauffrau eine im Mittel höhere handlungsbasierte Kompetenz aufweisen und dass die verstehensbasierten Fähigkeitsstrukturen zum Teil bedenklich gering ausgeprägt sind (vgl. Kapitel 5.2.4). Der technisch akzentuierte Zugriff auf Inhaltsbereiche der kaufmännischen Domäne wurde nicht explizit erfasst.

Um domänenverbundene und domänenspezifische Kompetenz aufeinander zu beziehen, sind Aufgabenformate zu wählen, in denen ausgehend vom Ansatz der Literacy/Numeracy in einen domänenspezifischen Geschäftsvorfall eingeführt wird. Dieses Vorgehen entspricht der betrieblichen Realität, wenn ausgehend von Belegen, Rechnungen, Protokollen etc. eine berufliche Handlung bzw. ein beruflicher Entscheidungsprozess ausgelöst wird. Die Frage, in welchem Umfang domänenbezogene Literalität jedoch die domänenspezifischen Kompetenzen beeinflusst, ist ungeklärt. Auf Basis der Ergebnisse kann angenommen werden, dass auf mathematischen Zugriffen basierende Geschäftsvorfälle stärker beeinflusst sein werden, als dies für Geschäftsvorfälle mit überwiegend sprachlicher Argumentation erwartet werden kann. Der Bereich der interpretativen Kompetenz kann in diesem Zusammenhang als möglicher Engpass identifiziert werden – sprachlich ausgerichtete Problemstellungen sind von einer Mehrzahl der Auszubildenden einfacher zu erfassen und zu validieren als mathematisch orientierte Anforderungssituationen.

Die Ergebnisse der Vorwissenserhebung bestätigen zudem die Annahme, dass sich unter der Voraussetzung zunehmender Domänenspezifität (im Vorwusstest wurde diese über curriculare Bezüge hergestellt) Übergänge zwischen allgemeinen kognitiven Fähigkeiten und domänenspezifischen wissensbasierten Entscheidungen feststellen lassen. Mit zunehmendem Fachbezug scheinen allgemeine Fähigkeiten zur Bewältigung der Anforderungssituation von spezifischen Fachkenntnissen abgelöst zu werden (siehe auch vorwärtsverkettendes vs. rückwärtsverkettendes Vorgehen im Rahmen der Problembearbeitung; Funke, 2004, S. 288f.). Der Übergang von allgemeinen ökonomischen Vorstellungen und Konzepten zu theoriebasierten betriebswirtschaftlichen Modellen ist scheinbar im Rahmen des kaufmännischen Eingangsunterrichts an eine Abnahme der Leistungsfähigkeit im Bereich der domänenbezogenen Literalität gekoppelt. Allgemeine mathematische und sprachliche Fähigkeiten werden zugunsten berufsspezifischer Lösungsalgorithmen aufgegeben, die jedoch noch nicht in dem Maße eingeübt und internalisiert sind, um die gegebenen beruflichen Anforderungssituationen adäquat zu lösen. Im Ergebnis sind Testaufgaben, die dem Literacy/Numeracy-Konzept folgen und kaum curriculare Bezüge aufweisen, von Auszubildenden ohne kaufmännische Vorbildung einfacher zu lösen als von Auszubildenden mit entsprechender Vorbildung (vgl. Kapitel 5.1.3.2).

In einer Studie, in der domänenverbundene und domänenspezifische Kompetenzen integriert erfasst werden, sollten auf der Grundlage von bislang erzielten Ergebnissen die domänenbezogenen und domänenspezifischen Zugänge zu

kaufmännischen Inhalten in einem ersten Schritt als Testskalen konstruiert werden, bei denen eine Auswahl der bislang entwickelten domänenspezifischen Anforderungssituationen um domänenbezogene Literacy/Numeracy-Items ergänzt wird. Dies bedeutet sicherlich eine Reduzierung der betriebspezifischen Realität, da in betrieblichen Arbeitsabläufen selten Verbalisierungen der Eingangs- bzw. Ausgangsdokumente vorgenommen werden. Dem gegenüber steht jedoch eine Stärkung der berufsspezifischen Realität, wenn internalisiert ablaufende Prozesse über geeignete Testaufgaben, mit denen Literalität auf die Berufsspezifika bezogen werden kann, externalisiert und zum Gegenstand berufsauthentischer Anforderungssituationen gemacht werden. In einem zweiten Schritt ist vorstellbar, dieses Verfahren um zusätzliche Items aus den Bereichen Literacy/Numeracy und allgemeiner kognitiver Kompetenzen (Lesen, Schreiben, Rechnen) zu erweitern, um zum einen eine Entkopplung domänenverbundener und allgemeiner kognitiver Kompetenzen von den Inhalten der kaufmännischen Domäne zu erreichen und um damit zum anderen zu Kennwerten zu gelangen, mit denen unterschiedliche berufliche Domänen verglichen werden können. In einem ergänzenden dritten Schritt sind Verfahren der Selbsteinschätzung als Elemente der externen Validierung möglich.

Vorgeschlagen wird folglich ein Methodenmix aus berufsspezifischen Simulationsaufgaben, Performanztests und Selbsteinschätzungen, mit dem es aus heutiger Perspektive allerdings nicht gelingen wird, berufliche Handlungskompetenz in all ihren Facetten und Wechselwirkungen reliabel und objektiv zu erfassen. Eine solche Zielsetzung ist aber weder notwendig noch im Hinblick auf die Operationalisierungsbemühungen zum Konzept der beruflichen Handlungskompetenz wünschenswert. So zeigen einzelne Arbeiten wie z. B. die KOMET-Studie (Berufliche Kompetenzen und berufliche Identität von Auszubildenden in Elektroberufen: Eine „Large-Scale“-Untersuchung an berufsbildenden Schulen in Hessen und Bremen (KOMET); Rauner, Haasler, Heinemann & Grollmann, 2009), dass mit einer Überdimensionierung des Konzeptes beruflicher Handlungskompetenz sowohl die Freiheitsgrade bei der Gestaltung von Testitems als auch die Möglichkeiten der empirischen Validierung unter Rückgriff auf die aktuellen Verfahren der angewandten Statistik drastisch eingeschränkt werden und die Forderung nach reliablen Messverfahren an der Komplexität des zu prüfenden Modells scheitert (scheitern muss).

6.1.2 Graduierung beruflicher Kompetenz

Für unterschiedliche fachliche Domänen und für verschiedene Ebenen der Kompetenzmessung (Kompetenz der Lernenden/Professionalität der Lehrenden) zeichnet sich ab, dass die Graduierung von Kompetenz an kognitiven Prozessen orientiert ist (vgl. u. a. Seeber, 2008; Blömeke, Kaiser & Lehmann, 2008; Nickolaus, 2008). Hiermit sind zwei Probleme verknüpft: Zum einen lässt sich die in der Bloomschen Taxonomie angenommene kumulative hierarchische Struktur der kognitiven Prozesse nicht empirisch validieren (vgl. Madaus, 1994). Zum anderen ist bislang ungeklärt, wie sich die Taxonomie von Bloom, die auf eine in-

putbasierte Lernzieloperationalisierung abzielt, auf kompetenzbasierte und damit outputorientierte Assessments übertragen lässt. Insbesondere im Hinblick auf den zweiten Punkt ist hilfreich, dass sich die neueren Adaptationen der Bloomschen Taxonomie wie die *Taxonomy for Learning, Teaching and Assessing* (Anderson & Krathwohl, 2001) oder die *New Taxonomy of Educational Objectives* (Marzano & Kendall, 2007; 2008) verstärkt auf Prozesse des Lernens und Lehrens beziehen. Bezogen auf den ersten Problempunkt wird in den Adaptationen der Taxonomie von einer tendenziellen Zunahme an Komplexität, nicht jedoch von einer kumulativen hierarchischen Struktur ausgegangen. Die Komplexität der kognitiven Prozesse wird demnach (1) über die Anzahl der eingebundenen und miteinander interagierenden Bearbeitungsschritte und (2) über den Grad der Vertrautheit mit der zugrunde liegenden Lern- und Arbeitsanforderungen bestimmt. Damit liegen operationalisierbare Kriterien vor, mit denen der Grad der kognitiven Auseinandersetzung bestimmt werden kann und mit denen sich das Leistungs- bzw. Lösungsverhalten in spezifischen Anforderungssituationen bewerten lässt. Je intensiver kognitive Auseinandersetzungen in einer Anforderungssituation eingefordert werden, umso komplexer und anspruchsvoller ist diese Situation und umso höher ist die Leistung einer erfolgreichen Anforderungsbewältigung einzuschätzen (vgl. Kapitel 3.3.3).

Die Suchrichtung für die Niveaumodellierung von Kompetenzen ist durch folgende Logik vorgegeben: Gesucht werden Kriterien, die den Anforderungsgehalt von Testaufgaben mit der Folge erhöhen, dass die Lösungswahrscheinlichkeiten einer erfolgreichen Aufgabenbewältigung sinken. Es sind folglich theoretische Modelle zu entwickeln, mit denen sich Komplexitätskriterien – wie beispielsweise die kognitiven Prozesse – auf Verarbeitungs- und Lösungsprozesse beziehen lassen. Um die Charakteristik der Anforderungssituation und die Leistungsfähigkeit der Testperson gemeinsam zu bewerten, sind fachdidaktische Ausarbeitungen sowohl im Hinblick auf die Konstruktion von Testsituationen als auch für die Auswertung der Ergebnisse zwingend notwendig.

Die Kompetenzniveaumodellierung für den Bereich der domänenspezifischen kaufmännischen Kompetenz greift auf drei Ansätze zurück: den Grad der kognitiven Auseinandersetzung im Hinblick auf die in Anforderungssituationen angeregten kognitiven Prozesse, den Umfang der inhaltlichen Komplexität sowie die funktionalen Modellierungsleistungen, die eine Testperson zur Erfassung der Anforderungssituation erbringen muss. Es wird davon ausgegangen, dass mit diesen drei Ansätzen sowohl die Situationsmerkmale als auch die personalen Steuerungssysteme adäquat erfasst werden können, da sich diese Ansätze aus einer Situations- und Personenperspektive begründen und fachdidaktisch bearbeiten lassen. So wird über die Wahl des kognitiven Prozesses, den eine Anforderungssituation ansprechen soll, neben der Formulierung des Items oder des Itemsatzes parallel die Leistung definiert, die die Testperson zur Bewältigung der Anforderungssituation erbringen muss. Gleiches gilt für den Bereich der inhaltlichen Komplexität. Soll das Verstehen von Systemzusammenhängen erfasst werden, sind die verschiedenen Teilbereiche eines betriebswirtschaftlichen Systems in der Anforderungssituation zu definieren.

derungssituation zu verankern, und der Systemzusammenhang ist als Lösung von den Testpersonen einzufordern. Insbesondere im Hinblick auf die Gestaltung komplexer Lehr-Lernumgebungen ist der komplexe Problemeinstieg ein oft eingesetztes Mittel der kaufmännischen Didaktik. Der komplexe Problemeinstieg erfordert auf Seiten der Situation ein offenes Format und i. d. R. ein Überangebot an Informationen, die von den Testpersonen über funktionale Modellierungsleistungen zu ordnen und zu reduzieren sind, um die Anforderungssituation in eine bearbeitbare Problemsituation zu transformieren.

Die Anwendung mehrerer Ansätze hat den Vorteil, dass verschiedene fachdidaktische Fragestellungen systematisch in die Konstruktion von Anforderungssituationen einfließen können und sich in der Analyse letztlich prüfen lässt, welche Merkmale der Anforderungssituationen das Antwortverhalten der Testpersonen systematisch beeinflussen, indem sie die Itemschwierigkeit erhöhen. Die Ergebnisse zeigen, dass die drei gewählten Ansätze gute Prädiktoren der Itemschwierigkeit liefern, mit denen insgesamt 75 Prozent der Gesamtvarianz aufgeklärt werden können (vgl. Kapitel 5.2.4.4). Dieses Ergebnis entsprechend zu interpretieren, ist mit der Schwierigkeit verbunden, dass die Prädiktoren über Raterurteile gewonnen wurden; d. h. nicht die Situationsmerkmale selbst, sondern Fremdeinschätzungen der Situation wurden zum Ausgangspunkt der Niveaumodellierung gemacht. Die Modelle der IRT stellen Varianten zur Verfügung, mit denen verschiedene feststehende Eigenschaften von Testitems modelliert werden können – so z. B. Modelle, die instruktionale Implikationen oder Schätzwahrscheinlichkeiten berücksichtigen können. Jedoch existieren (bislang) keine Modelle, die Variationen in den Situationsparametern adäquat verarbeiten. Die zentrale Herausforderung in diesem Zusammenhang liegt darin, dass mit Modellen insbesondere der Rasch-Familie die Beziehung zwischen Itemschwierigkeit und Antwortverhalten direkt hergestellt wird und man davon ausgeht, dass es ausschließlich die Leistungsfähigkeit einer Testperson ist, die das Antwortverhalten bei einer gegebenen Itemschwierigkeit beeinflusst – und letztlich ist das genau die Information, die erfasst werden soll. Die Konsequenz ist, dass die Kompetenzniveau-modellierung bislang weniger als empirisches Problem gesehen worden ist, sondern vielmehr eine fachdidaktische Herausforderung darstellt. So sind für das Konstrukt der domänenspezifischen kaufmännischen Kompetenz vier Stufen inhaltlich begründbar: „Kaufmännisches Grund- und Regelwissen“, „Kaufmännisches Handlungswissen“, „Kaufmännisches Analysewissen“ und „Kaufmännisches Entscheidungswissen“. Diese Stufen sind das Ergebnis des zur Erfassung des Kompetenzkonstrukts eingesetzten Tests und des Antwortverhaltens der getesteten Stichprobe. Eine der vorrangigen Aufgaben wird es sein, diese Stufen in weiteren Studien zu validieren und entsprechend weiter auszuarbeiten.

Bislang wenig Beachtung findet die These, dass die ex post ermittelten Kompetenzstufen situationssensitiv sind. Hierunter ist zu verstehen, dass das Lern- und Leistungsverhalten von Auszubildenden in verschiedenen Inhaltssequenzen deutliche Unterschiede aufweisen kann. Im Rahmen des DFG-Projekts „Integrierte Kompetenzentwicklung in den beruflichen Fächern des Fachgymnasiums Wirt-

schaft“ konnte beispielsweise gezeigt werden, dass die domänenspezifische Lernleistung eines Auszubildenden stark von den Lerninhalten abhängt und damit zwischen den Lerninhalten erheblich variiert (intraindividueller Vergleich; $F = 53,59$; $p < 0,001$; $\eta^2 = 0,097$; vgl. Winther & Achtenhagen, 2008a, S. 272). Die intraindividuellen Leistungsvariationen in unterschiedlichen Lern- und Leistungssituationen sind dabei maßgeblich von regulativen Vermittlungsleistungen beeinflusst – nur in Anforderungssituationen, die motivational und volitional herausfordernd sind, bringt der Auszubildende seine kognitiven Ressourcen in vollem Umfang ein. Vor dem Hintergrund dieser Befunde ist die Stabilität von Kompetenzstufen diskussionswürdig. Die Identifizierung von Regulationsstereotypen wäre ein möglicher Ansatz, um die Graduierung von Kompetenzen weiter zu klären. Regulationsstereotype können dazu beitragen, die ermittelten Kompetenzstufen stärker auf den Anforderungsgehalt von Testsituationen zu beziehen. Eng damit verbunden ist die Forderung, Testsituationen möglichst motivational und volitional anregend zu gestalten, um die regulativen Vermittlungsleistungen der Auszubildenden zu stützen. Erste empirische Befunde zeigen, dass volitionale und kognitive Kontrolle sowie motivationale Regulation der Auszubildenden in computergestützten Simulationsumgebungen stärker ausgeprägt sind, als dies in überwiegend Paper-Pencil-basierten Lern- und Leistungsumgebungen der Fall ist. Die höhere Ausprägung der regulativen Vermittlungsleistungen wirkt dabei positiv auf die gezeigte domänenspezifische Lernleistung (vgl. Winther, 2007; Winther & Achtenhagen, 2008a).

6.1.3 Entwicklung beruflicher Kompetenz

Im Rahmen der vorliegenden Arbeit wird die kaufmännische Kompetenz von Auszubildenden zu einem bestimmten Zeitpunkt ihrer Kompetenzentwicklung erhoben: in der Studie im vollzeitschulischen Bereich zu Beginn und in der Studie im Bereich des Dualen Systems kurz vor Abschluss des Ausbildungsganges (im dritten Ausbildungsjahr). Diese Einpunkterhebungen lassen weder Rückschlüsse auf Faktoren der Kompetenzentwicklung zu, noch liefern sie Hinweise auf Faktoren, mit denen sich Unterschiede in den Bildungsergebnissen erklären lassen. Um diese Fragen zu beantworten, ist eine Längsschnittuntersuchung durchzuführen. Dabei muss sich eine Erhebung im Längsschnittdesign an einem Modell orientieren, mit dem Kontextbedingungen und Entwicklungsprozesse kaufmännischer Kompetenz abgebildet werden können und das in operationalisierter und damit empirisch validierbarer Form vorliegt. Ein solches Modell existiert für den Bereich der beruflichen Bildung bislang nicht. Mit dem im Rahmen des KOMET-Projekts vertretenen Experten-Novizen-Paradigma (vgl. Rauner, Haasler, Heinemann & Grollmann, 2009, unter Bezug auf Dreyfus & Dreyfus, 1986) wird deutlich gemacht, dass ein zentraler Auftrag beruflicher Bildungsarbeit darin bestehen sollte, eine Enkulturation in eine Expertenkultur anzustreben; allerdings erfüllt die auf der Basis dieses Paradigmas vorgenommene Anordnung verschiedener Lernbereiche auf einem Kontinuum von Orientierungs- und Überblickswissen bis fachsystematischem Vertiefungswissen die Anforderungen eines Kompetenz-

entwicklungsmodells nicht. Im Gegensatz dazu sollten folgende Faktoren die Modellierung von Kompetenzentwicklung in beruflichen Handlungsfeldern leiten:

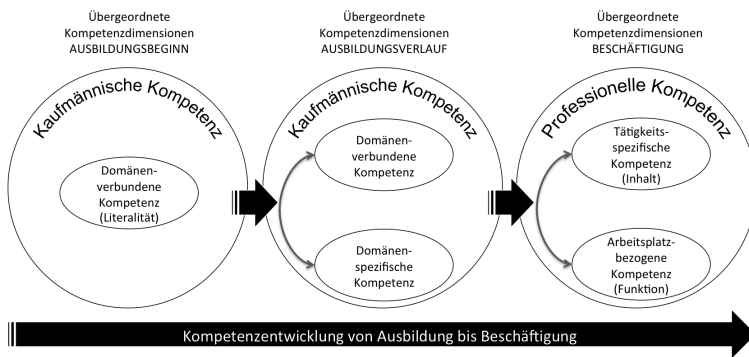
- Entwicklungspsychologische Erklärung der Kompetenzstufen vor dem Hintergrund der Erlern- und Vermittelbarkeit von Kompetenzen;
- detaillierte Erfassung der Kontextfaktoren des Lernens und der internen Vorgänge des Lernenden;
- Abbildung der Übergangsprozesse von allgemeinen kognitiven Fähigkeiten zu domänenspezifischen wissensbasierten Entscheidungen sowie
- Aufzeigen der Enkulturationswege vor dem Hintergrund der Ausbildung von Arbeitsroutinen und informell erworbener Fähigkeiten.

Diese Faktoren machen deutlich, dass die empirische Fundierung von Kompetenzentwicklungsmodellen dann an ihre Grenzen stößt, wenn es um mehr geht als die Herausbildung und Entwicklung von kognitiven Strukturen im Experten-Novizen-Vergleich. Die Modellierung beruflicher Kompetenzentwicklung hat als Erweiterung der Entwicklungsmodelle des allgemeinbildenden Bereichs insbesondere die Frage zu klären, wie der Erwerb von Arbeitsroutinen und informell erworbener Fähigkeiten angemessen zu berücksichtigen ist. Unter Rückgriff auf die ACT-R-Theorie (Anderson, 1994; vgl. Kapitel 2.2.1) kann beispielsweise festgehalten werden, dass Routinen ausgeführt werden können, ohne dass eine entsprechende deklarative Repräsentation der Wissensbasis vorliegen muss. Danach können sich die Zusammenhänge und Umfänge im Kompetenzmodell unter Berücksichtigung von Arbeitsroutinen in der beruflichen Entwicklung dahingehend verändern, dass prozedurale Kompetenzen an Bedeutung gewinnen und konzeptuale sowie interpretative Kompetenzen an Bedeutung verlieren. Konzeptuale Kompetenzen brauchen in standardisierten Arbeitsprozessen nicht als aktivierte Wissensbestände zur Verfügung zu stehen, und interpretative Kompetenzen werden ausschließlich vor dem Hintergrund unbekannter Situationen lösungsrelevant.

Ähnliche Verschiebungen in den Kompetenzstrukturen sind im Hinblick auf die Unterscheidung von domänenspezifischen und domänenverbundenen Kompetenzen zu erwarten. Im Zuge der beruflichen Kompetenzentwicklung werden sich Strategien der Anforderungsbewältigung herausbilden, die von dem einzelnen Auszubildenden/Arbeitnehmer als effektiv bewertet werden. In der Folge ist vorstellbar, dass eine systematische Trennung der Kompetenzbereiche nicht mehr möglich ist. Während im Verlauf der Ausbildung eine Unterscheidung in domänenspezifische und domänenverbundene Fähigkeitsstrukturen sinnvoll ist, kann diese Trennung im Verlauf der beruflichen Kompetenzentwicklung im Rahmen eines Beschäftigungsverhältnisses durch eine Kategorie abgelöst werden, die als professionelle Kompetenz überwiegend über tätigkeitsspezifische und arbeitsplatzbezogene Faktoren charakterisiert werden kann (vgl. Abbildung 6.2). Die zentrale Herausforderung in diesem Zusammenhang wird sein, wie sich die einzelnen Kompetenzdimensionen aufeinander beziehen und in einem Messmodell umsetzen lassen. Hierbei ist hilfreich, dass in dem hier vorgelegten Ansatz die Erfassung der domänenspezifischen Kompetenz am Ausbildungsende über

die Skalen der handlungsbasierten und verstehensbasierten Kompetenz bereits tätigkeitsspezifisch ausgerichtet ist (Kriterium der tätigkeitsspezifischen Validität; vgl. Kapitel 5.2.2.2) und sich vor diesem Hintergrund auch für Erhebungen im Beschäftigungssystem eignet. Die im Rahmen von Beschäftigung zu erfassende tätigkeitsspezifische Kompetenz beschreibt in diesem Zusammenhang die zentralen Inhalte kaufmännischer Aktivitäten, während über die arbeitsplatzbezogene Kompetenz – als Ergänzung – funktionsrelevante Aspekte kaufmännischer Aktivität und damit auch Elemente des Karriereverlaufs berücksichtigt werden. Eine Fokussierung auf tätigkeitsspezifische Kompetenz wird auch durch die Ergebnisse der Feasibility Study VET-LSA gestützt. In der Studie konnte für den Vergleich unterschiedlicher nationaler Bildungsgänge gezeigt werden, dass sich eine inhaltliche Übereinstimmung arbeitsmarktbezogener Anforderungen und betrieblicher Aufgaben nicht für Berufe, wohl aber für berufliche Tätigkeiten erzielen lässt (Baethge & Arends, 2009; für den kaufmännischen Bereich Breuer, Hillen & Winther, 2009). Über welche Zugriffe auf die beruflichen Domänen sich tätigkeitsspezifische und arbeitsplatzbezogene Kompetenzen beschreiben lassen, ist Aufgabe weiterer Forschung.

Abbildung 6.2 Veränderung des Kompetenzmodells unter einer Entwicklungsperspektive



Ein Ausblick auf die entwicklungspsychologische Erklärung der Kompetenzstufen vor dem Hintergrund der Erlern- und Vermittelbarkeit von Kompetenzen muss vage bleiben, da die Kompetenzstufen post-hoc auf Grundlage der testmodell-konform übrig gebliebenen Aufgaben auf Basis von IRT-Modellen definiert werden, wobei die Kompetenzstufengrenzen durch eine Unterteilung der Fähigkeitsskala hinsichtlich der Mindestlösungschancen bestimmt werden. Dieses Verfahren ist sicherlich als defizitär zu bewerten und wird erst durch die fachdidaktische Interpretation der Kompetenzstufen nutzbar. Der zentrale Kritikpunkt ist, dass die so gewonnenen Kompetenzstufen keine Analyse der Leistungsprozesse zulassen, sondern lediglich die Gesamtschwierigkeit bestimmen (vgl. Neubrand,

Klieme, Lüdtke & Neubrand, 2002). Zur Abbildung von Kompetenzentwicklung ist es jedoch nötig, individuelle Zugewinne innerhalb der Kompetenzdimensionen zu identifizieren. Um dies leisten zu können, sind weitere Forschungen und fachdidaktische Ausarbeitungen notwendig, mit denen es gelingen kann, die Anforderungssituationen sowie das Lern- und Arbeitsverhalten von Auszubildenden/Beschäftigten möglichst detailliert zu erfassen, um festzustellen, welche Anforderungssituationen die Auszubildenden/Beschäftigten bei welcher Ausprägung einer Kompetenz bewältigen. Zum jetzigen Zeitpunkt kann empirisch nicht festgestellt werden, inwieweit die bislang identifizierten vier Kompetenzstufen – kaufmännisches Grund- und Regelwissen, kaufmännisches Handlungswissen, kaufmännisches Analysewissen und kaufmännisches Entscheidungswissen – auch dafür geeignet sind, Entwicklungsprozesse abzubilden. Es lässt sich aber die Annahme vertreten, dass über diese Kompetenzstufen Aspekte eines Hineinwachsens in eine Expertenkultur schlüssig nachvollzogen werden können.

6.2 Forschungsdesiderate der beruflichen Kompetenzmessung

Ein zentrales Forschungsdesiderat der empirischen Berufsbildungsforschung lässt sich in Bezug auf Instrumente der Kompetenzdiagnostik skizzieren. Der Entwicklungsbedarf umfasst in diesem Zusammenhang alle Bereiche der beruflichen Aus- und Weiterbildung sowie die verschiedenen Methoden der Kompetenzerfassung. Eng verbunden mit der Entwicklung von Kompetenzmessverfahren für wissenschaftliche Diagnosezwecke ist die Frage, wie sich diese Instrumente für berufliche Assessments und Prüfungsverfahren adaptieren lassen.

6.2.1 Entwicklung und Validierung von Kompetenzmessverfahren

Im Rahmen der vorliegenden Arbeit wurden kaufmännische Kompetenzen über simulationsbasierte Verfahren erfasst, mit denen sich durch Wissens- und Performanzmessungen das Arbeitsverhalten der Auszubildenden beobachten und diagnostizieren lässt. Der Bereich der domänenverbundenen Kompetenz wurde dabei über Simulationsaufgaben im Paper-Pencil-Format erfasst, während im Bereich der domänenspezifischen Kompetenz computerbasierte Simulationen komplexer Arbeits- und Geschäftsprozesse zum Einsatz gekommen sind. Simulationsbasierte Verfahren lassen sich im Hinblick auf berufliche Kompetenzmessung über folgende Kriterien charakterisieren:

- Authentische Abbildung komplexer berufstypischer Situationen, Zahlen, Fakten, Arbeitsmittel und Beziehungen;
- Abbildung vollständiger Arbeits- und Geschäftsprozesse – für den Bereich der kaufmännischen Berufsbildung bedeutet dies, dass fachwissenschaftlich der systemorientierte Ansatz der Betriebswirtschaft zu vertreten ist;
- Standardisierung der Arbeits- und Geschäftsprozesse durch Unternehmensmodellierung.

Die besondere Herausforderung beruflicher Kompetenzmessverfahren besteht darin, dass die konstruierten Anforderungssituationen realen betrieblichen

Vorgängen und Fragestellungen entnommen und damit zwangsläufig einer Realitätsdiskussion unterworfen werden. Mit Hilfe des Konzepts der Authentizität wird die Vermittelbarkeit fachdidaktisch simulierter Realität angestrebt. Insbesondere für den Bereich der domänenspezifischen Kompetenz kann die Auffassung vertreten werden, dass auf Basis computerbasierter Simulationen Testverfahren entwickelt werden können, mit denen berufliche Realität in hohem Maße antizipiert werden kann. Dies gilt in der kaufmännischen Domäne insbesondere vor dem Hintergrund der EDV-gestützten Arbeitstätigkeiten und der über die betrieblichen ERP-Systeme begünstigten hohen Vernetzung der Arbeitsprozesse. Darüber hinaus werden mit simulationsbasierten Verfahren Testumgebungen unterstützt, die komplexen Handlungssituationen gerecht werden können. Komplexe Handlungssituationen zeichnen sich durch ein Abwägen von Zielen aus, fokussieren auf Lösungswege und Handlungsschritte und lassen Raum für Entscheidungsprozesse und damit für Lösungsvariationen.

Forschungsbedarf besteht im Zusammenhang mit computerbasierten Simulationen insbesondere im Hinblick auf die technisch-funktionale Umsetzung vollständiger Arbeits- und Geschäftsprozesse sowie die Optimierung der Auswertungsverfahren. Plattformen, mit denen technologie- bzw. netzwerkbasierte Assessments, die über ein Online-Fragebogensystem hinausgehen, möglich werden, sind derzeit für angewandte Forschung nur beschränkt zugänglich. Mit der open-source Plattform TAO (vgl. hierzu Plichard, Jadoul, Vandenabeele & Latour, 2004, zit. n. Reeff, 2007) steht eine generische Architektur für internetbasiertes Testen zur Verfügung. Die Leistungsfähigkeit der Plattform im Hinblick auf die Abbildung kaufmännischer Arbeits- und Geschäftsprozesse ist jedoch (bislang) nicht gegeben. Interessant sind hingegen die Möglichkeiten der im TAO-System angelegten Datenverwaltung und der Erzeugung von Metadaten als wesentliche Elemente einer erfolgreichen Testdurchführung sowie die verlinkten Verfahren einer automatisierten Datencodierung und -auswertung. In diesem Zusammenhang sind Forschungsk Kooperationen zwischen Assessmentexperten und Softwareentwicklern anzustreben, um den Aufwand und die Kosten für zukünftige computerbasierte Tests zu reduzieren und eine höhere Leistungsfähigkeit der Simulationen zu erreichen.

Bei der inhaltlichen Ausgestaltung kann hingegen auf weitreichende Forschungsergebnisse aus dem Umfeld der Entwicklung und Evaluation komplexer Lehr-Lern-Arrangements zurückgegriffen werden. Die Studie zur Erfassung kaufmännischer Kompetenz im Ausbildungsberuf Industriekaufmann/Industriekauffrau zeigt beispielsweise, dass über die inhaltliche Konzeption des Tests sowohl die Skala der handlungsbasierten Kompetenz als auch die Skala der verstehensbasierten Kompetenz umfassend repräsentiert werden können. Darüber hinaus haben Untersuchungen zu den Effekten von Aufgabenmerkmalen auf das Antwortverhalten der Auszubildenden deutlich gemacht, dass das Konstrukt der berufsfachlichen Kompetenz mit dem eingesetzten Testverfahren valide abgebildet werden kann (Konstruktvalidität). In weiteren Analysen wird ein Fokus darauf liegen, die konkreten beruflichen Anforderungssituationen weiter aufzu-

schlüsseln und hinsichtlich ihrer kognitiven Struktur und ihrer Regeln zu analysieren (vgl. hierzu Anderson, 1993; 1996; Greeno, 1998). Die bislang verwendeten Kriterien zur Identifikation von Itemschwierigkeiten – funktionale Modellierung, inhaltliche Komplexität und Art der kognitiven Taxonomierung – stellen einen bereits empirisch validierten Ansatz dar, der mit Hilfe weiterer Testaufgaben verifiziert bzw. weiterentwickelt werden muss.

Ein insbesondere für den kaufmännischen Bereich gültiges Problem stellt die Prüfung der Kriteriumsvalidität dar. Hierfür wird vorgeschlagen, computerbasierte Simulationen auf Basis einer umfassenden Umternehmensmodellierung mit (standardisierten) Arbeitsproben in einem realen Unternehmen zu vergleichen. Dieser Ansatz ist für kaufmännische Inhalte jedoch nur auf den ersten Blick sinnvoll, wobei zwei Argumente von besonderer Relevanz sind: (1) Die Arbeitsplätze und Arbeitsanforderungen von kaufmännischen Angestellten unterscheiden sich massiv – in Abhängigkeit von der Unternehmensbranche, der Unternehmensgröße, aber auch innerhalb eines einzelnen Unternehmens. Diese Unterschiede, die sich sowohl auf die Arbeitsmittel als auch auf die betrieblichen Lösungen beziehen, verhindern ein Standardisieren von Testumgebungen und damit eine objektive und reliable Messung. (2) Arbeiten, die über Routinetätigkeiten hinausgehen und die Fähigkeit zu eigenverantwortlichem Handeln voraussetzen, dürfen von kaufmännischen Auszubildenden – wenn überhaupt – nur unter Anleitung ausgeführt werden. Diese betriebliche Maßnahme ist dem Umstand geschuldet, dass kaufmännisch relevante Entscheidungen direkte Auswirkungen auf die Beschaffungs-, Produktions- und Vertriebsprozesse des Unternehmens haben und in der Mehrzahl der Fälle auf die Kunden- und Lieferantenbeziehungen ausstrahlen. Zwei Alternativen, mit diesem Problem umzugehen, bieten sich an: Einerseits könnten die Arbeitsproben unabhängig von den realen betrieblichen Abläufen simuliert werden – einschließlich fingierter Störungen und Kommunikationsprozesse. In diesem Fall fielen betriebliche Arbeitsprobe und technologiebasierte Simulation zusammen, mit dem Vorteil, dass über die in der Simulation verarbeiteten Arbeits- und Geschäftsprozesse eine Standardisierung der Testumgebung möglich ist. Andererseits könnte das Arbeitsverhalten in realen betrieblichen Situationen über Fremdeinschätzungen bewertet werden. Hierzu ist eine intensive Schulung externer Beobachter zwingend. Das Problem der Teststandardisierung ist damit jedoch auch nicht zu umgehen und zusätzlich wird die Auswertung und Interpretation der so gewonnenen Daten durch Ratingprozesse beeinträchtigt. Darüber hinaus ist der Aufwand der Datengewinnung extrem hoch, wenn bedacht wird, dass ein Ausbildungsbetrieb in der Industrie im Mittel nicht mehr als drei kaufmännische Auszubildende pro Lehrjahr beschäftigt. Dennoch kann die Annahme vertreten werden, dass auf Fremdeinschätzung basierende Expertenratings valide Kriterien liefern können, die als externes Validitätskriterium insbesondere Aussagen im Hinblick auf domänenspezifische Kompetenzen zu stützen in der Lage wären.

6.2.2 *Blue Prints zur Konstruktion von beruflichen Assessments*

Die im Rahmen der vorliegenden Arbeit vorgenommenen Konstruktionen und Unternehmensmodellierungen sind zu umfangreich, um direkt auf schulische und betriebliche Assessments mit dem Ziel der Individualdiagnostik übertragen zu werden. Dennoch wird der Anspruch erhoben, mit den konstruierten Anforderungen und den erzielten Ergebnissen Hinweise für die Gestaltung von Prüfungsaufgaben ableiten und die Akzeptanz von kompetenzorientierten Prüfungen bei den Beteiligten in Schule und Wirtschaft erhöhen zu können. Dass die Entwicklung und Neugestaltung von schulischen und betrieblichen (Abschluss-)Prüfungen ein wesentliches Forschungsdesiderat darstellt, zeigt sich u. a. daran, dass bislang kaum Aussagen zur Validität und Reliabilität von beruflichen Abschlussprüfungen vorliegen. Auch muss vermutet werden, dass zwischen schriftlichen und mündlichen Prüfungsteilen erhebliche Divergenzen bestehen: Während in den schriftlichen Teilen eine gestauchte Normalverteilungskurve erwartet wird, kann dies für die mündlichen Prüfungsteile nicht angenommen werden. Hier zeigen sich in älteren Analysen ausgeprägte Häufungen in den oberen Leistungskategorien – Befunde, die sich in den Ergebnissen der bislang durchgeführten Fachleistungstests an beruflichen Schulen so nicht bestätigen. Aktuelle Studien, in denen kompetenzbasierte Fachleistungstests eingesetzt wurden, zeigen eine Dominanz der unteren Leistungsbereiche (Lehmann & Seeber, 2007; Winterher & Achtenhagen, 2008; Nickolaus, Gschwendtner & Geißel, 2008), die sich so in den Verteilungen der Ergebnisse der schriftlichen Abschlussprüfungen nicht wiederfindet.

Ein erstrebenswertes „Neben“produkt der Forschungen im Bereich Kompetenzmodellierung und Kompetenzmessung im Hinblick auf ein Large-Scale Assessment in der beruflichen Bildung kann sein, dass sich aus dem Pool der beruflichen Anforderungssituationen Assessmentsysteme als Blue Prints für kompetenzorientierte Assessments in verschiedenen kaufmännischen Ausbildungsprogrammen entwickeln lassen. Ein zentrales Forschungsdesiderat stellt in diesem Zusammenhang die Konstruktion von Testitems sowie deren Präsentation auf Basis eines validen Kompetenzmodells dar (vgl. u. a. Shavelson, 2008). Anhaltspunkte für die Entwicklung von Blue Prints werden im Rahmen des Evidence-Centered Assessment Design-Ansatzes gegeben (u. a. Mislevy, Almond & Lukas, 2003; im Detail vgl. Kapitel 3.1.2). Dieser Ansatz bindet die Assessmententwicklung eng an die Testvalidierung. Dies hat den Vorteil, dass Inhalt und Design des Assessments an den Kontext der Messung sowie an die Probanden gekoppelt sind. Hierfür wird jedes Assessmentsystem über vier Prozessstufen ausgebildet (vgl. Almond, Steinberg & Mislevy, 2002): (1) In einem ersten Schritt erfolgt die Auswahl und Sequenzierung der Anforderungssituationen/Testaufgaben und werden sämtliche Informationen, die zur Aufgabendarstellung benötigt werden, gesammelt. Diese Informationen enthalten insbesondere Hinweise darauf, wie komplex bzw. anspruchsvoll eine Anforderungssituation ist. Das Produkt dieses Prozessschrittes ist eine Datenbank, mit deren Hilfe sich sowohl für verschiedene berufliche Inhalte als auch für unterschiedliche Kontexte und Anspruchsgruppen

passende Aufgabenstellungen finden lassen und in einem Test zusammengestellt werden können. (2) Über einen zweiten Schritt soll der gesamte Prozess der Aufgabendarstellung gesteuert werden. Dies ist insbesondere für die Abbildung kaufmännischer Arbeits- und Geschäftsprozesse sehr sinnvoll, da dieser Prozessschritt nur dann effektiv sein kann, wenn vollständige Datensätze und umfangreiche Modellierungen von beruflichen Situationen vorliegen. Das Produkt dieses Prozesses kann im kaufmännischen Bereich als Baukasten für Unternehmensmodellierungen interpretiert werden. (3) Der dritte Prozessschritt liefert Auswertungshinweise auf der Basis exemplarischer Antworten, während (4) im vierten und letzten Prozessschritt die Auswertung erfolgt.

Mit den Ergebnissen der in dieser Arbeit beschriebenen Studien können die einzelnen Prozessstufen inhaltlich angereichert werden. Jedoch steht eine Erweiterung und zusätzliche Validierung der Befunde noch aus, um tatsächlich eine Plattform für die Gestaltung kaufmännischer Assessments und Abschlussprüfungen im Sinne eines kompletten Assessmentsystems erstellen zu können.

6.3 Kompetenzorientierung in Lehr- und Lernprozessen

Die breit angelegte Forschung im Bereich Kompetenzmessung und Kompetenzmodellierung konzentriert sich bislang vorrangig auf die Instrumentenentwicklung sowie auf die Diagnostik. Sie stellt darüber hinaus auch auf Entwicklungsperspektiven im Hinblick auf die Gestaltung und die Evaluation von Lehr- und Lernprozessen ab. In diesem Zusammenhang wird deutlich, dass sich Kompetenzdiagnostik und Kompetenzvermittlung voneinander unterscheiden. Fragen des Kompetenzaufbaus und des Kompetenzerwerbs sind jedoch insbesondere aufgrund der fehlenden empirischen Befunde zur Kompetenzentwicklung kaum gesichert zu beantworten. Im Hinblick auf die Kompetenzen von Lehrpersonen als Gestalter von Unterrichtsprozessen liegen hingegen erste Befunde vor, in denen insbesondere die fachdidaktischen Kompetenzen der Lehrenden als zentrale Mittler des Lernerfolgs betont werden (vgl. Baumert & Kunter, 2006).

6.3.1 Kompetenzorientierte Unterrichtsgestaltung

Die Diskussion um eine Kompetenzorientierung in schulischen und betrieblichen Lehr-Lernprozessen wird durch die Debatte um Bildungsstandards und den Wechsel im Hinblick auf die Steuerung der Bildungsprozesse begleitet. Die Forschungen zeigen, dass eine Inputsteuerung, in der sich der Lernerfolg über das Angebot von Lerngelegenheiten, die Ausbildung der Lehrenden sowie die Ausstattung der Bildungseinrichtungen definiert, wenig effizient ist. Grubb (2009) stellt beispielsweise über eine kritische Sichtung verschiedener empirischer Studien heraus, dass zwischen finanzieller Ausstattung von Bildungseinrichtungen und dem Outcome der Lernenden kaum Zusammenhänge bestehen und dass das Nutzen von Lerngelegenheiten nicht von den personellen und strukturellen Ressourcen einer Bildungseinrichtung abhängt. Bei Forschungen zur Kompetenzorientierung in Lehr-Lernprozessen werden vor diesem Hintergrund zusätzlich zur

Qualität der Lerngelegenheiten die Qualität der Vermittlung (Prozesssteuerung) sowie die auf empirischer Diagnostik basierenden Kompetenzen der Lernenden (Outputsteuerung) in den Blick genommen, um eine Modifizierung der Lerngelegenheit begründen und umsetzen zu können (vgl. Schwippert, 2005). Das Ziel kaufmännischer Lehr-Lernprozesse kann über ein vertieftes Verständnis der Domäne und über ein sinnhaftes Handeln in domänenspezifischen Situationen definiert werden. Diese Leitidee kaufmännischer Bildungsprozesse wird dabei über das Konzept der beruflichen Handlungskompetenz beschrieben und bezieht sich auf schulische sowie auf betriebliche Lerngelegenheiten. Eine Möglichkeit, die in den Studien zur Erfassung kaufmännischer Kompetenz gewonnenen Ergebnisse auf eine Entwicklung von Lehr-Lernprozessen und damit auf die Gestaltung von Lerngelegenheiten zu beziehen, besteht darin, die identifizierten Kompetenzstufen als fachdidaktische Zielkategorien zu formulieren. Damit ist der Aufbau eines kaufmännischen Grund- und Regelwissens als Ausgangspunkt der Enkulturationsprozesse in eine berufliche Expertenkultur zu sehen. Kaufmännisches Grund- und Regelwissen stellt in diesem Zusammenhang die Voraussetzung für den Erwerb kaufmännischen Handlungswissens dar, das als arbeitsmarktbezogene Vorgabe beruflicher Ausbildungsprogramme definiert werden kann. Die beruflichen Entwicklungen und der Kompetenzerwerb in der beruflichen Domäne sind zu einem großen Teil durch Kontext- und Arbeitsprozesswissen geprägt, das auf beruflichen Erfahrungen basiert und die Grundlage für kaufmännisches Analyse- und Entscheidungswissen darstellt. Während also die empirisch ermittelten Kompetenzstufen das Ergebnis einer fachdidaktisch geprägten Iteminterpretation sind (vgl. Kapitel 5.2.4.4), werden über die nachfolgend beschriebenen Zielkategorien Gestaltungselemente von Lernprozessen herausgearbeitet.

Zielkategorie I: Aufbau eines kaufmännischen Grund- und Regelwissens

Kompetenzorientierte Lehr-Lernprozesse zielen auf die Stützung eines individuellen Lernens und Leistens ab. Die Studie zur Erfassung kaufmännischer domänenverbundener Kompetenz liefert Befunde, die die Bedeutung ökonomischer Literalität für zukünftige Lehr-Lernprozesse in der kaufmännischen Domäne unterstreichen. Die im beruflichen Bildungssystem starke Heterogenität der Schülerschaft – im Hinblick auf Alter und Vorbildung der Auszubildenden – schlägt sich auch in großen Varianzen in den Vorwissensleistungen nieder. Es lassen sich Lernende identifizieren, die auf Alltagsniveau argumentieren, und solche, die bereits über ein ausgebautes Verständnis ökonomischer Zusammenhänge verfügen. In diesem Zusammenhang sind ein Angleichen der Wissensbasis und das Überwinden naiver sowie missconcept-behafteter Alltagstheorien von besonderer Relevanz. Unter Rückgriff auf die Kompetenzstufe I: Kaufmännisches Grund- und Regelwissen, sind solche Anforderungssituationen in die Unterrichtsgestaltung einzubringen, die auf ökonomische Begriffs- und Konzeptstrukturen abstellen und insbesondere Systemkohärenzen betonen. Damit soll die für ein sinnvolles Handeln in kaufmännischen Situationen notwendige deklarative Wissensbasis aufgebaut werden.

In den Richtlinien und Lehrplänen kaufmännischer Schulen ist für den Eingangunterricht eine auf das systemische Verstehen hin ausgerichtete Unterrichtsgestaltung bereits formal vorgesehen (vgl. Niedersächsisches Kultusministerium, 2001; KMK, 2000). Darüber hinaus liegen mit komplexen Fallstudien und multimedialen Lehr-Lehr-Arrangements umfassende Materialien vor, mit denen über verschiedene, authentisch an der Berufsrealität orientierte Situationen grundlegende Begriffe und Konzepte der betriebswirtschaftlichen Theorie vermittelt werden können. Die Ergebnisse belegen jedoch, dass gerade im Hinblick auf die verstehensbasierte Kompetenz mehr als ein Drittel der Auszubildenden nicht über ein auf Systemkohärenzen basierendes Verständnis der betriebswirtschaftlichen Theorie verfügt. Die betrieblichen Arbeits- und Geschäftsprozesse werden hingegen von der Mehrzahl der Auszubildenden in ihren Zusammenhängen erfasst (vgl. Kapitel 5.2.4.4). Für die Unterrichtsgestaltung lassen sich auf Basis dieser Befunde folgende Schlussfolgerungen ableiten: Es ist einerseits eine intensivere Vermittlung betriebswirtschaftlicher Theorien anzustreben; ein Prozess der insbesondere vor dem Hintergrund der Bachelorisierung betriebswirtschaftlicher Studiengänge neu zu bewerten und aufzugreifen ist. Andererseits sind neue Formen der Dekontextualisierung ökonomischen Wissens zu entwerfen. Studien amerikanischer wirtschaftswissenschaftlicher Fakultäten zeigen, dass sich über Ansätze des Entrepreneurship die wahrgenommene inhaltliche Relevanz sowie das Interesse an betriebswirtschaftlichen Inhalten steigern lässt; damit verändern sich auch die Struktur des domänenspezifischen Wissens sowie die Anwendungsbezüge (vgl. u. a. Program for Executive Education at the Stanford Graduate School of Business).

Zielkategorie II: Erwerb eines kaufmännischen Handlungswissens

Für ein adäquates Handeln in schulischen und betrieblichen Situationen ist eine Aktivierung bestehender Wissensbestände und Strukturen unabdingbar. Doch selbst bei ausreichender Wissensbasis ist ein angemessenes Handeln nicht zwangsläufig die Folge. Die Ursachen dafür können in nicht ausreichend ausgeprägten Metaprozessen liegen, die eine erfolgreiche Umsetzung des Wissens in konkrete Handlungskontexte behindern (vgl. hierzu Artelt, Demmrich & Baumert, 2001). Die Ergebnisse der Studien zur Erfassung kaufmännischer Kompetenz lassen den Schluss zu, dass insbesondere Defizite im Lernbereich (inhaltspezifische Vorkenntnisse) eine bewusst-strategische Bearbeitung der Lernanforderung erschweren, da ohne inhaltspezifisches Wissen Lernprozesse nicht befriedigend reguliert werden können. Insbesondere die fehlende Adaptationsfähigkeit der kognitiven Strukturen der Auszubildenden trägt dazu bei, dass Anforderungen der Kompetenzstufe II: Kaufmännisches Handlungswissen, von einer Vielzahl der Auszubildenden nicht angemessen bearbeitet werden können.

Die Herausforderung bei Aufgaben auf diesem Kompetenzniveau besteht darin, berufliche Anforderungen in bearbeitbare Problemstellungen zu übersetzen und die gegebenen Informationen flexibel zu nutzen. Dies setzt voraus, dass fachliches Wissen dekontextualisiert vorliegt, auf eine Vielzahl von Situationen über-

tragbar ist und in diesen auch genutzt werden kann. Insbesondere aus der Situated Learning-Forschung lassen sich wichtige Bedingungen für den Wissenstransfer ableiten. Transfer kann umfassend und mit hohem Allgemeingrad in einem ersten Zugriff als Anwendung von Wissen in neuen Situationen definiert werden (vgl. Messner, 1978; Mandl, Prenzel & Gräsel, 1991). Es besteht die Gefahr, dass in der Lernsituation erworbenes Wissen auf eben diese Situation fixiert bleibt. Um dies zu vermeiden, sollten Kasuistik und Systematik des Unterrichts in eine Balance gebracht werden (vgl. u. a. Achtenhagen, 2002). Eine planvolle Unterrichtssequenzierung, die dem Prinzip konkreter vor abstrakter Handlungsfolge verpflichtet sein sollte, stellt sicher, dass dieselben Inhalte in verschiedenen Kontexten vermittelt werden (vgl. CTGV, 1992). Multiple Kontexte erleichtern es dem Lernenden, von relevanten Merkmalen der Konzepte zu abstrahieren und eine flexible Wissensrepräsentation zu entwickeln. Die Flexibilität bei der Anwendung des Gelernten wird darüber hinaus durch multiple Perspektiven gesichert. Über multiple Kontexte und Perspektiven werden Lerngelegenheiten geschaffen, die helfen, einen sicheren Umgang mit betriebswirtschaftlichen Konzepten einzuüben, indem die Lernerfahrungen an informationsreiche, kohärente und realistische Anforderungssituationen gebunden sind (Verhoeven, Schnotz & Paas, 2009). Empirische Studien belegen in diesem Zusammenhang, dass authentische und komplexe Lernumgebungen die Lernenden darin schulen, Anforderungssituationen zu bewerten, die Problemstellung zu identifizieren und zu definieren. Darüber hinaus können sie dazu beitragen, dass Lernende über ein weitreichendes Repertoire an Lösungsstrategien verfügen (u. a. Goldman, 2009; Schacter & Fagnano, 1999).

Diese Ergebnisse lassen sich sowohl auf schulische als auch auf betriebliche Lernprozesse anwenden. Für schulisches Lernen bestätigen die Ergebnisse, dass die Entwicklungen und Evaluationen komplexer Lernumgebungen weiter von Bedeutung sind. Damit verbunden ist die Forderung, die in den Konstruktionskriterien komplexer Lernumgebungen verankerten Impulse einer auf systemisches Verstehen ausgerichteten Unterrichtsgestaltung auch auf die Lehrbücher des kaufmännischen Unterrichts anzuwenden. Für betriebliches Lernen ist zu betonen, dass Lernen am Arbeitsplatz nicht per se kohärent und informationsreich ist, sondern in erheblichem Maße auf die transparente Darstellung betrieblicher Arbeits- und Geschäftsprozesse angewiesen ist. Nur so kann es den Lernenden gelingen, kaufmännisches Handlungswissen aufzubauen, das sich dann flexibel in wechselnden Anforderungssituationen anwenden und umsetzen lässt.

Zielkategorien III und IV: Ausbau eines kaufmännischen Analyse- und Entscheidungswissens

Der Unterrichtsgestaltung mit Hilfe komplexer Lernumgebungen wird auch im Hinblick auf den Erwerb von kaufmännischem Analyse- und Entscheidungswissen Bedeutung beigemessen. Zur Beschreibung dieser Stufen kaufmännischer Kompetenz sind domänenverbundenes und domänenbezogenes Wissen – im Sinne eines systemischen Verstehens fachbezogener und fachspezifischer Konzepte

und Prozeduren – um ein auf vielfältige Anforderungssituationen bezogenes Kontextwissen und um ein spezifisches Arbeitsprozesswissen zu ergänzen. Die Kombination aus Fachwissen, Kontextwissen und Prozesswissen wird in der Literatur auch als Erfahrungswissen bezeichnet, das erst durch Einübung von Routinen und durch Wiederholung von berufstypischen Anforderungen erworben werden kann (Paulsen, 1993; Hacker, 1992). Der Erwerb und der Ausbau von Erfahrungswissen kann insbesondere durch prozessorientierte Lern- und Arbeitsvorgänge gestützt werden, da hierdurch die vollständige Planbarkeit von betrieblichen Prozessen in Frage gestellt werden kann. Beispiele für den Unterrichtsalltag wären Störungen, die in die Beschaffungs-, Produktions- und Vertriebsprozesse eines Unternehmens eingebunden werden. Diese Störungen können zahlreiche unternehmensinterne und -externe Folgen haben, die i. d. R. nicht standardisiert sind und im Ergebnis auch nicht eindeutig bearbeitet werden können. Weder die Komplexität der Einflussgrößen ist vollständig beschreib- oder berechenbar, noch ist es möglich, alle Einflüsse oder Ereignisse in komplexen Unternehmensprozessen zu antizipieren. In der Folge wird für kaufmännische Anforderungssituationen neben einem planmäßig-rationalen auch ein erfahrungsgeleitetes subjektivierendes Handeln bedeutsam (Böhle & Weihrich, 2009; Böhle, Pfeiffer, Seysay-Tegethoff & Nese, 2004).

Während auf der Ebene der Kompetenzstufe III: Kaufmännisches Analysewissen, planmäßig-rationales Handeln eingefordert wird, ist für die Kompetenzstufe IV: Kaufmännisches Entscheidungskwissen, insbesondere erfahrungsgeleitetes Verhalten für ein erfolgreiches Bearbeiten der Anforderungssituationen notwendig. Planmäßig-rationales Handeln setzt eine schlussfolgernde Auslegung von Fachwissen voraus, um die in vielfältigen Anforderungssituationen erzeugten Lösungen im Hinblick auf ihre Anwendbarkeit und Verwertbarkeit analysieren und validieren zu können. Erfahrungsgeleitetes-subjektivierendes Handeln kann hingegen als dialogisch-explorativ beschrieben werden. Dies bedeutet, dass in kaufmännischen Entscheidungssituationen Imaginationen komplexer Prozesse und deren Effekte analysiert und validiert werden müssen, um zu einer Lösungsfindung zu kommen. Diese Imaginationen sind vorrangig assoziativ und umfassen im Sinne eines komplexen Managementprozesses die Problemformulierung, die Zielsetzungen, die Lösungsmöglichkeiten sowie die resultierenden Entscheidungen einschließlich einer Abschätzung der zu erwartenden Ergebnisse des Handelns.

Für die Gestaltung von Unterricht bedeutet dies, dass neben der Konstruktion von Lernumgebungen, in deren Zentrum Unternehmensprozesse stehen, auch Methoden zu implementieren sind, mit denen insbesondere der Erwerb und der Ausbau von Erfahrungswissen gefördert werden können. Hierzu zählen Übungsaufgaben, die ein entdeckend-exploratives Vorgehen begünstigen und für ein Abbilden der Systemzusammenhänge ein assoziativ-bildhaftes Denken erfordern. Derart komplexe Aufgabenstellungen setzen ein fundiertes kaufmännisches Grund- und Regelwissen sowie ein ausgeprägtes kaufmännisches Handlungswissen voraus.

6.3.2 Befunde zur Lehrprofessionalität

Eine wesentliche Herausforderung im Hinblick auf die Gestaltung von kompetenzorientierten Unterrichtsprozessen ist die Abstimmung der Methoden des Unterrichts mit den im Unterricht zu vermittelnden Kompetenzen. Oder anders formuliert: Wie ist ein Unterricht zu gestalten, mit dem beispielsweise erfahrungsgeleitete-subjektivierende Kompetenz im Sinne eines komplexen Managementverhaltens effizient ausgebildet werden kann, und welche Fähigkeiten benötigt ein Lehrender, um Lernende substantiell in ihrem Lernprozess unterstützen zu können? Die Beantwortung dieser Fragestellung wird dadurch erschwert, dass bislang ungenügend gesicherte Informationen darüber vorliegen, welche Eigenschaften eine effektive instruktionale Unterstützung aufweisen sollte und aus welchen Gründen einige Instruktionen positive Effekte auf das Lernen und Leisten haben und andere nicht. Schnotz und Kürschner (2007) betonen in diesem Zusammenhang, dass die kognitiven Lehr- und Lerntheorien keine einfachen „rules-of-thumb“ zur Verfügung stellen, mit denen der Einfluss des Lehrens auf die Leistungsfähigkeiten der Lernenden gesichert vorausgesagt werden könnte.

In einer auf umfangreichem Datenmaterial basierenden Meta-Analyse stellen Seidel und Shavelson (2007) instruktionale Einflussgrößen auf das Lernen vor. Die Analyse ist aus zwei Gründen bemerkenswert: Zum einen werden Indikatoren der Lehreffektivität entworfen, mit deren Hilfe Vergleiche zwischen theoretisch unterschiedlich akzentuierten Studien möglich sind. Dies erlaubt beispielsweise, Ergebnisse aus kognitiv geprägten Lehr- und Lernansätzen auf behavioristische Ansätze zu übertragen. Zum anderen werden in der Meta-Analyse Ergebnisse von Studien zusammengefasst, deren forschungsmethodische Bearbeitung von Korrelationsbeschreibungen bis zum experimentellen Interventionsdesign reicht. Folgende Ergebnisse können als besonders zentral herausgestellt werden: Die Beschreibung der Effektivität von Instruktionsmerkmalen greift zu kurz, wird ausschließlich der übergreifende Outcome des Lernens bewertet. Für diesen Fall weisen sämtliche Prädiktoren (z. B. Verstärkung, Feedback, Lernzeit) verschwindend geringe Effektstärken auf. Diese erhöhen sich signifikant, wenn der Outcome des Lernens auf zentrale, das Lernen beschreibende Prozesse bezogen wird. Der Outcome wird vor diesem Hintergrund über folgende Kategorien bestimmt (vgl. Seidel & Shavelson, 2007, p. 470): Die Kategorie Lernprozess fokussiert auf die Regulation von Lernaktivitäten vor dem Hintergrund des Wissenserwerbs einschließlich der für den Wissenserwerb notwendigen Vermittlungsleistungen (kognitive und motivationale Auseinandersetzung, Einsatz von Lernstrategien). Über die Kategorien motivational-affektiver Outcome und kognitiver Zuwachs werden hingegen eher Langzeiteffekte erfasst, wie sie in standardisierten Leistungs- und Kompetenztests abgebildet werden. Die Ergebnisse zeigen, (1) dass instruktionale Kriterien i. d. R. stärker auf die Effektivität des Lernprozesses als auf die summative kognitive Lernleistung ausstrahlen (zum Vergleich der Lernleistungsprädiktoren auf state- und trait-Ebene des Lernens siehe Winther, 2006) und dass (2) eine domänenspezifische Lernprozessgestaltung der stärkste Prädiktor der Lernleistung ist – unabhängig von der untersuchten Domäne, der un-

tersuchten Schulform und -stufe oder der Kategorie des Outcome des Lernens (Seidel & Shavelson, 2007, p. 476). Unter domänenspezifischer Lernprozessgestaltung ist zu verstehen, dass Lehrende Lernumgebungen schaffen sollen, in denen das Lernen auf die Inhalte, Methoden und Arbeitsweisen eines konkreten Inhalts- und/oder Fachbereiches abgestellt ist. Voraussetzung dafür ist ein tiefgehendes Verständnis der fachlichen Inhalte und Methoden, das in pädagogisch sinnvoller Form angewendet werden kann (vgl. hier die Unterscheidung zwischen *subject-matter content knowledge* und *pedagogical content knowledge*; Shulman, 1987; White & Frederiksen, 1998; Baumert & Kunter, 2006).

Diese exemplarischen Befunde – auch wenn zu berücksichtigen ist, dass die domänenspezifischen Merkmale der Unterrichtsqualität in den in die Meta-Analyse eingeflossenen Studien größtenteils über distale Indikatoren, wie Kursbesuche oder Trainingsmaßnahmen, erfasst wurden (vgl. u. a. die Studie von McGuinness, McGuinness & Donohue, 1995) – begünstigen eine Neuausrichtung der theoretischen Perspektive der Lehrprofessionalität abseits von generalisierten Erziehungserwartungen (vgl. hierzu die Beiträge in Zlatkin-Troitschanskaia, Beck, Sembill, Nickolaus & Mulder, 2009). Nach Baumert und Kunter (2006, S. 473) stellen die didaktische Vorbereitung und die Inszenierung von Unterricht die zentralen Anforderungen für den Lehrerberuf und das professionelle Kompetenzprofil von Lehrenden dar. Die Autoren folgen damit der Argumentation Shulmans, der kontinuierlich anmahnt, dass in einer psychologisch verengten Unterrichtsforschung, den Gegenständen des Unterrichts – im Sinne fachlicher Inhalte – zu wenig Beachtung geschenkt wird und die ausschließliche Berücksichtigung generisch pädagogischer Kompetenzen den Blick auf professionelles Lehrerhandeln verstellt (Shulmann, 1987; 1998; 2004; vgl. hierzu u. a. Minnameier, 2005). So belegen Studien zum professionellen Handlungswissen, dass die im Unterricht eingesetzten fachdidaktischen Kenntnisse und Fähigkeiten der Lehrenden weitgehend von ihrem konzeptuellen Fachwissen abhängen (für einen Überblick vgl. Baumert & Kunter, 2006). Auf der Basis dieser Befunde lässt sich festhalten, dass (1) die Ausprägung und Adaptationsfähigkeit domänenspezifischen Wissens das Repräsentations- und Handlungsrepertoire im Sinne fachdidaktischer Expertise mit positiven Effekten für die (domänenspezifische) Unterrichtsgestaltung erhöht und dass folglich (2) Untersuchungen zur Lehrprofessionalität domänenspezifisch anzulegen sind.

Belastbare empirische Ergebnisse zum fachwissenschaftlichen und fachdidaktischen Wissen und Können von Lehrenden in spezifischen Domänen liegen bislang kaum vor. Für den deutschsprachigen Raum, als maßgeblich kann das auf das Professionswissen von Mathematiklehrern ausgerichtete Projekt COACTIV (Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die Entwicklung mathematischer Kompetenz; Kunter, Klusmann, Dubberke, Baumert, Blum, Brunner, Jordan, Krauss, Löwen, Neubrand & Tsai, 2006) angesehen werden. Im Rahmen dieses Projekts, das an die nationale Ergänzung der PISA-Studie 2003 gebunden ist, werden die inhaltlichen, fachdidaktischen und pädagogischen Fähigkeiten von Lehrkräften konzeptualisiert und hinsichtlich ih-

rer Bedeutung für eine erfolgreiche, auf professionelle Handlungskompetenz ausgerichtete Berufsausübung analysiert. In den Ergebnissen wird Folgendes deutlich:

- Domänenspezifisches Fachwissen (*subject-matter content knowledge*) und fachdidaktische Fähigkeiten (*pedagogical content knowledge*; Shulman, 1987) sind empirisch abgrenzbare Konstrukte. D. h., sie bilden verschiedene Wissensfacetten der Lehrprofessionalität ab, die strukturell mit weiteren Aspekten der Lehrprofessionalität assoziiert sind. So geben Lehrende, die über ein hohes fachwissenschaftliches und fachdidaktisches Wissen verfügen, in geringerem Maße transmissive lerntheoretische Überzeugungen an. Insbesondere die fachdidaktische Fähigkeit der Lehrenden erweist sich als signifikanter Prädiktor für eine kognitiv-herausfordernde und konstruktiv-unterstützende Unterrichtsgestaltung und ist – mediiert über die Unterrichtsmerkmale – substantiell bedeutsam für die Leistung der Lernenden (zusammenfassend vgl. Kunter, Klusmann & Baumert, 2009).
- Fachwissen und fachdidaktische Fähigkeiten werden domänenspezifisch ausgebildet und an den Inhalten der jeweiligen Domäne für Professionshandeln deutlich. Hierbei zeigt sich, dass ein intensiveres Fachstudium sowohl für ein höheres Fachwissen als auch für ein ausgeprägteres fachdidaktisches Wissen ein Indikator ist. Seifried und Ziegler (2009) merken in diesem Zusammenhang an, dass die Modellierung domänenspezifischen Professionswissens neben dem Inhaltsbereich auch die Adressaten, den institutionellen Rahmen sowie das Ziel- bzw. Bezugssystem des Unterrichts beinhalten sollte, und werfen die allgemeine Frage auf, inwieweit sich die Befunde aus dem Bereich der Mathematik auf berufliche Bereiche übertragen lassen. Hierzu ist anzumerken, dass insbesondere die Frage der Domänenspezifität in der beruflichen Lehrerbildung solange zweitrangig bleiben wird, solange nicht die substantielle Bedeutung des fachwissenschaftlichen und fachdidaktischen Wissens für die professionelle Kompetenz von Lehrenden angemessen in den Diskussionen der Lehreraus- und -weiterbildung Berücksichtigung findet (vgl. hierzu u. a. Neuweg, 2005).
- Die Berufserfahrung hat keinen Einfluss auf die Ausprägung des Fachwissens und der fachdidaktischen Fähigkeiten. Insbesondere dieser empirische Befund steht im scheinbaren Widerspruch zu generalisierenden Annahmen der Lehreraus- und -weiterbildung. Terhart (u. a. 2002; 2006; 2009) vertritt z. B. die These, dass sich Professionalität von Lehrkräften im Verlauf der Berufsbiographie vollzieht. Wenn aber Kompetenzentwicklung – wird der Expertiseforschung gefolgt – an strukturierte Lerngelegenheiten gebunden ist, wird die Perspektive einer zeitlich ausgerichteten, berufsbiographischen Kompetenzgenese nicht verworfen, sondern im Hinblick auf die Notwendigkeit einer inhaltsspezifischen Ausdifferenzierung gestützt. Terhart betont in diesem Zusammenhang insbesondere eine stärkere Positionierung der Fachdidaktik in der ersten Phase der Lehrerbildung und schlägt gleichzeitig vor, andere fachbezogene Studieninhalte zu kürzen und

das Studienangebot strenger auf das spätere Berufsfeld zu beziehen (vgl. Terhart, 2009, S. 434). Vor einer Reduzierung fachwissenschaftlicher Inhalte kann in diesem Zusammenhang nur gewarnt werden – eine fachwissenschaftliche Aufwertung der Lehreraus- und -weiterbildung müsste das Ziel sein. Empirisch zeigt sich das Fachwissen als Mittler der fachdidaktischen Fähigkeiten. Oder anders formuliert: Fachdidaktisches Wissen kann nicht ohne Fachwissen erworben und angewendet werden (Baumert & Kunter, 2006; Kunter, Klusmann & Baumert, 2009).

Im Hinblick auf kompetenzorientierte Gestaltung von Unterricht als eine zentrale Anforderung an Lehrprofessionalität lassen sich aus den beschriebenen Befunden folgende Schlussfolgerungen ziehen: Insgesamt sind die derzeit vorliegenden empirischen Befunde, die insbesondere aus der Mathematiklehrerbildung stammen (vgl. hier auch Blömeke, Kaiser & Lehmann, 2008) nicht geeignet, um generalisierende Rückschlüsse auf andere Domänen zu ziehen. Sie stärken jedoch die für die berufliche Bildung weitgehend akzeptierte Annahme, dass in allen Phasen der Lehrerbildung strukturierte Lerngelegenheiten zur Verfügung zu stellen sind, deren Fokus auf der fachwissenschaftlichen und fachdidaktischen Ausbildung liegen sollte. Die Beurteilung der Qualität der Lehrerbildung sollte dabei evidenzbasiert (vgl. Weber & Achtenhagen, 2009; Oser, 2001) und unter Berücksichtigung kompetenzorientierter Unterrichtsentwicklung (vgl. Oelkers & Reusser, 2008) erfolgen.

Diese Forderungen sind empirisch zu validieren. Im Hinblick auf die Ergebnisse der Studien zur Erfassung kaufmännischer Kompetenz ist beispielsweise von besonderem Interesse, über welche Merkmale der Unterrichtsgestaltung und des professionellen Lehrerhandelns sich die Graduierung verstehensbasierter und handlungsbasierter Kompetenz erklären lässt – Informationen die für Lern- und Lehrprozesse in schulischen und betrieblichen Kontexten von Bedeutung sind (van Buer & Zlatkin-Troitschanskaia, 2007). Allerdings ist zu berücksichtigen, dass der Kompetenzerwerb und die Kompetenzentwicklung von Auszubildenden im beruflichen Bereich zu einem großen Teil auch von den Ausbildungsleistungen der Betriebe getragen wird. In der Machbarkeitsstudie für ein Berufsbildungs-PISA (VET-LSA; Baethge, Achtenhagen, Arends, Babic, Bathge-Kinsky & Weber, 2006) werden diese Zusammenhänge von institutionellen und individuellen Ausbildungsbedingungen für den Erwerb von Kompetenzen und deren Verwertung diskutiert; dabei wird herausgestellt, welche umfassenden Informationen – auch solche zu Gestaltungsmerkmalen der Ausbildungsprozesse – für ein Benchmark nationaler Berufsbildungssysteme gegeben sein müssen. Schumann und Eberle (2009) schlagen vor, Erhebungen zur Professionalität des Ausbildungspersonals an die Konzeption eines VET-LSA zu binden, um den Kausalzusammenhang zwischen Professionalität der Lehrpersonen und beruflicher Kompetenz der Auszubildenden unter Berücksichtigung der Lernorte und der Zielsetzung der Ausbildung erfassen zu können. Ein wesentliches Desiderat in diesem Zusammenhang stellt die konzeptionelle Modellbildung dar, die als zentrales Element der Grundlagenforschung in diesem Bereich zu sehen ist.

Literatur

- Abs, H. J. (2007). Überlegungen zur Modellierung diagnostischer Kompetenz bei Lehrerinnen und Lehrern. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 63–84). Münster: Waxmann.
- Achtenhagen, F. (1998). Prospekt einer multimedial gestützten Einführung in die Betriebswirtschaftslehre. Ein Mastery-Learning-Projekt für Industriekaufleute. In R. Bogaschewski & U. Götzke (Hrsg.), *Unternehmensplanung und Controlling. Festschrift zum 60. Geburtstag von Jürgen Bloech* (S. 357–371). Heidelberg: Physika.
- Achtenhagen, F. (2001). Komplexe Lehr-Lern-Arrangements/-Umgebungen. In G. Kramer & K. Kiepe (Hrsg.), *Jahrbuch Ausbildungspraxis 2001* (S. 265–270). Köln: Fachverlag Deutscher Wirtschaftsdienst.
- Achtenhagen, F. (2002). Das Unternehmen als komplexes ökonomisches und soziales System. In F. Achtenhagen (Hrsg.), *Forschungsgeleitete Innovation der kaufmännischen Berufsausbildung – insbesondere am Beispiel des Wirtschaftsgymnasiums* (S. 49–74). Bielefeld: Bertelsmann.
- Achtenhagen, F. (2003). Konstruktionsbedingungen für komplexe Lehr-Lern-Arrangements und deren Stellenwert für eine zeitgemäße Wirtschaftsdidaktik. In F.-J. Kaiser & H. Kaminski (Hrsg.), *Wirtschaftsdidaktik* (S. 77–97). Bad Heilbrunn: Klinkhardt.
- Achtenhagen, F. (2004). Prüfung von Leistungsindikatoren für die Berufsbildung sowie zur Ausdifferenzierung beruflicher Kompetenzprofile nach Wissensarten. In BMBF (Hrsg.), *Bildungsreform Band 8: Expertisen zu den konzeptionellen Grundlagen für einen Nationalen Bildungsbericht – Berufliche Bildung und Weiterbildung/Lebenslanges Lernen*. Bonn: BMBF.
- Achtenhagen, F. (2005). Prüfung von Leistungsindikatoren für die Berufsbildung sowie zur Ausdifferenzierung beruflicher Kompetenzprofile nach Wissensarten. In M. Baethge, K.-P. Buss & C. Lanfer (Hrsg.), *Expertisen zu den konzeptionellen Grundlagen für einen Nationalen Bildungsbericht – Berufliche Bildung und Weiterbildung/Lebenslanges Lernen* (S. 11–32). Berlin: BMBF.
- Achtenhagen, F. (2006). Europäischer Bildungsraum als Auftrag der Berufsbildungsforschung. In M. Eckert & A. Zöller (Hrsg.), *Der europäische Berufsbildungsraum – Beiträge der Berufsbildungsforschung* (S. 287–313). Bielefeld: Bertelsmann.
- Achtenhagen, F., Preiß, P. & Weber, S. (2004). Fachdidaktische Grundlagen des Ökonomieunterrichts. In H. Kaminski (Hrsg.), *Lernmodule "Ökonomische Bildung online"*. Gütersloh: Bertelsmann Stiftung.
- Achtenhagen, F. & Winther, E. (2006). Möglichkeiten des Kompetenzaufbaus und seiner Erfassung bei Schülerinnen und Schülern in der kaufmännischen Erstausbildung. In G. Minnameier & E. Wuttke (Hrsg.), *Berufs- und wirtschaftspädagogische Grundlagenforschung. Lehr-Lern-Prozesse und Kompetenzdiagnostik. Festschrift für Klaus Beck* (S. 345–360). Frankfurt a. M.: Peter Lang.
- Achtenhagen, F. & Winther, E. (2007). *Ergebnisse des DFG-Projekts "Integrierte Förderung von Selbst-, Sach- und Methodenkompetenz in den berufsbezogenen Fächern des Wirtschaftsgymnasiums". Abschlussbericht zur Einreichung bei der DFG*. Professur für Wirtschaftspädagogik der Georg-August Universität Göttingen.
- Achtenhagen, F. & Winther, E. (2008). Wirtschaftspädagogische Forschung zur beruflichen Kompetenzentwicklung. In N. Jude, J. Hartig & E. Klieme (Hrsg.), *Kompetenzerfassung in pädagogischen Handlungsfeldern: Theorien, Konzepte und Methoden. Bildungsreform-Reihe des BMBF, Band 26* (S. 117–140). Bonn, Berlin: BMBF.
- Adams, R. J. & Khoo, S. T. (1996). *Quest*. Melbourne, Australia: Australian Council for Educational Research.
- Adams, R. J. & Wilson, M. (1996). The Random Coefficients Multinomial Logit Model. In G. Engelhard & M. Wilson (Eds.), *Objective Measurement: Theory into Practice*. Norwood, NJ: Ablex.
- Adams, R. J., Wilson, M. & Wang, W.-C. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21 (1), 1–23.
- Adams, R. J. & Wu, M. (Eds.) (2002). *PISA 2000 Technical Report*. Paris: OECD.
- Aebli, H. (1978). Von Piagets Entwicklungspsychologie zur Theorie der kognitiven Sozialisation. In G. Steiner (Hrsg.), *Piaget und die Folgen* (S. 604–627). Zürich: Kindler.

- Aebli, H. (1981). *Denken: Das Ordnen des Tuns. Band II*. Stuttgart: Klett-Cotta.
- Alexander, P. A. (1992). Domain Knowledge: Envolving Themes and Emerging Concerns. *Educational Psychologist*, 27 (1), 33–51.
- Alexander, P. A., Schallert, D. L. & Hare, V. C. (1991). Coming to Terms: How Researchers in Learning and Literacy Talk about Knowledge. *Review of Educational Research*, 61, 315–343.
- Almond, R. G. & Mislevy, R. J. (1999). Graphical Models and Computerized Adaptive Testing. *Applied Psychological Measurement*, 23, 223–237.
- Almond, R. G., Steinberg, L. S. & Mislevy, R. J. (2002). Enhancing the Design and Delivery of Assessment Systems: A Four-Process Architecture. *Journal of Technology, Learning, and Assessment*, 1 (5), 138–153.
- Amabile, T. M. (1996). *Creativity in Context*. Boulder, CO: Westview Press.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (Eds.) (1999). *Standards for Educational and Psychological Testing*. Washington, DC: National Academy Press.
- Ames, C. (1992). Classrooms: Goals, Structures, and Student Motivation. *Journal of Educational Psychology*, 84, 261–271.
- Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1987). Skill Acquisition: Compilation of Weak-Method Problem Solutions. *Psychological Review*, 94, 192–210.
- Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Anderson, J. R. (1996). ACT. A Simple Theory of Complex Cognition. *American Psychologist*, 51 (4), 355–365.
- Anderson, J. R., Simon, H. A. & Reder, L. M. (1997). Situative versus Cognitive Perspectives: Form vs. Substance. *Educational Researcher*, 26, 18–21.
- Anderson, L. W. & Krathwohl, D. R. (with Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al.) (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY: Longman.
- Anderson, M. (1992). *Intelligence and Development: A Cognitive Theory*. Oxford: Blackwell.
- Arnold, R. (1997). Von der Weiterbildung zur Kompetenzentwicklung In Arbeitsgemeinschaft Qualifikations-Entwicklungs-Management (Hrsg.), *Kompetenzentwicklung 97: Berufliche Weiterbildung in der Transformation – Fakten und Visionen* (S. 253–307). Münster: Waxmann.
- Artelt, C., Demmrich, A. & Baumert, J. (2001). Selbstreguliertes Lernen. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Deutsches PISA-Konsortium) (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 271–298). Opladen: Leske + Budrich.
- Attwell, P. (1990). What is Skill? *Work and Occupations*, 17, 422–448.
- Australian Education Council (Ed.) (1992). *Mathematics Profiles, Levels 1 – 6*. Melbourne, Australia: Curriculum Corporation.
- Bader, R. & Müller, M. (2002): Leitziel der Berufsbildung: Handlungskompetenz: Anregungen zur Ausdifferenzierung des Begriffs. *Die berufsbildende Schule*, 54 (6), 176–182.
- Baethge, M., Achtenhagen, F., Arends, L., Babic, E., Baethge-Kinsky, V. & Weber, S. (2006). *Berufsbildungs-PISA – Machbarkeitsstudie*. Stuttgart: Steiner.
- Baethge, M. & Arends, L. (Eds.) (2009). *Feasibility Study VET-LSA. A Comparative Analysis of Occupational Profiles and VET Programmes in 8 European Countries. International Report*. Göttingen: Soziologisches Forschungsinstitut.
- Baethge, M. & Bartelheimer, P. (2005). *Deutschland im Umbruch. Ergebnisse und Perspektiven des sozioökonomischen Berichtsansatzes*. SOFI-Mitteilungen, 33. Soziologisches Forschungsinstitut Göttingen.
- Baethge, M., Solga, H. & Wieck, M. (2007). *Berufsbildung im Umbruch – Signale eines überfälligen Aufbruchs*. Berlin: Friedrich-Ebert-Stiftung.
- Baethge, M. & Wieck, M. (2006): Berufliche Bildung in der Bildungsberichterstattung. *Zeitschrift für Erziehungswissenschaft*, 10, Beiheft 6, 163–185.
- Baker, F. B. (1992). *Item Response Theory. Parameter Estimation Techniques*. New York, NY: Marcel Dekker.

- Barnett, R. (1994). *The Limits of Competence: Knowledge, Higher Education and Society*. Buckingham: Open University Press.
- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.) (2000). *Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schul Laufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Deutsches PISA-Konsortium) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Baumert, J. & Schümer, G. (2002). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb im nationalen Vergleich. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000 - Die Länder der Bundesrepublik Deutschland im Vergleich* (S. 159–202). Opladen: Leske + Budrich.
- Baxter, G. P. (1995). Using Computer Simulations to Assess Hands-on Science Learning. *Journal of Science Education and Technology*, 4 (1), 21–27.
- Beaton, A. E. & Allen, N. L. (1992). Interpreting Scales through Scale Anchoring. *Journal of Educational Statistics*, 17, 191–204.
- Beck, K. (2001). WBT – Wirtschaftskundlicher Bildungstest. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren* (S. 559–562). Lengerich: Pabst Science Publishers.
- Beck, K. (2005). Ergebnisse und Desiderate zur Lehr-Lern-Forschung in der kaufmännischen Berufsausbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 101, 533–556.
- Biggs, J. B. & Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York, NY: Academic Press.
- Björnsvold, J. (2000). *Assessing Non-Formal Learning: European Developments and Paradoxes*. CEDEFOP, Thessaloniki.
- Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for Learning*. London: Open University Press.
- Black, P. & Wiliam, D. (2004). The Formative Purpose: Assessment Must First Promote Learning. In M. Wilson (Ed.), *Towards Coherences Between Classroom Assessment and Accountability. 103rd Yearbook of the National Society for the Study of Education* (pp. 20–50). Chicago, IL: The University of Chicago Press.
- Blömeke, S., Kaiser, G. & Lehmann, R. (2008) (Hrsg.). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematik-studierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Münster: Waxmann.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (Eds.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York, NY: David McKay.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, C. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolf, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- BMBF (Hrsg.) (2002). *5-Punkte-Programm „Zukunft Bildung“*. Berlin.
- Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch-Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Böhle, F., Pfeiffer, S. & Sevsay-Tegethoff, N. (Hrsg.) (2004). *Die Bewältigung des Unplanbaren - Fachübergreifendes erfahrungsgeleitetes Arbeiten und Lernen*. Wiesbaden: VS Verlag.
- Böhle, F. & Wehrich, M. (Hrsg.) (2009). *Handeln unter Unsicherheit*. Wiesbaden: VS Verlag.
- Böhner, M. & Straka, G. A. (2005). Bankwirtschaftliche Kompetenz – Konzept und standardisierte Erfassung. *bwp@ – Berufs- und Wirtschaftspädagogik online*, 8.

- Bohlinger, S. (2008). Competence as the Core Element of the European Qualifications Framework. *European Journal of Vocational Training*, 43, 96–118.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*, 6. Aufl. Heidelberg: Springer.
- Brand, W., Hofmeister, W. & Tramm, T. (2005). Auf dem Weg zu einem Kompetenzstufenmodell für die berufliche Bildung – Erfahrungen aus dem Projekt ULME. *bwp@ – Berufs- und Wirtschaftspädagogik online*, 8.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (Eds.). (2000). *How People Learn. Brain, Mind, Experience and School*. Washington, DC: National Academy Press.
- Bransford, J. D. & Stein, B. S. (1984). *The Ideal Problem Solver: A Guide for Improving your Thinking, Learning, and Creativity*. New York, NY: W. H. Freeman
- Breuer, K., Hillen, S. & Winther, E. (2009). Comparative International Analysis of Occupational Tasks and Qualification Requirements for the Labour Market and Assessment Tasks at the End of VET in Participating Countries – Business and Administration. In M. Baethge & L. Arends (Eds.), *Feasibility Study VET-LSA. A Comparative Analysis of Occupational Profiles and VET Programmes in 8 European Countries. International Report*. Göttingen: Soziologisches Forschungsinstitut.
- Breuer, K. & Höhn, K. (1999). *Wirtschaftsmodellversuch: Entwicklung und Implementation eines Qualitätsförderungssystems für die handlungsorientierte Abschlussprüfung zum Versicherungskaufmann/zur Versicherungskauffrau auf Grundlage der Ausbildungsverordnung vom 8.2.1996. Abschlussbericht*. Karlsruhe: BiBB.
- Briggs, D. & Wilson, M. (2004). An Introduction to Multidimensional Measurement using Rasch-Models. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch-Measurement* (pp. 322–342). Maple Grove, MN: JAM Press.
- Brown, N. J. S. (2005). *The Multidimensional Measure of Conceptual Complexity*. Berkeley Evaluation and Assessment Research Center: University of California, Berkeley.
- Brunner, M. (2005). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Dissertation: Humboldt-Universität zu Berlin.
- van Buer, J. & Zlatkin-Troitschanskaia, O. (2007). Diagnostische Lehrerexpertise und adaptive Steuerung unterrichtlicher Entwicklungsangebote. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 381–400). Frankfurt a. M.: Peter Lang.
- van Buer, J. & Wagner C. (Hrsg.) (2007). *Qualität von Schule. Ein kritisches Handbuch*. Frankfurt a. M.: Peter Lang.
- Business Council of Australia (2002). *Employability Skills for the Future*. Australian Chamber of Commerce and Industry, DEST.
- Byrnes, J. P. (2008). *Cognitive Development and Learning in Instructional Contexts*. Bosten et al.: Pearson.
- Carnevale, A. P., Gainer, J. L. & Meltzer, A. S. (1991). *Workplace Basics: The Essential Skills Employers Want*. San Francisco: Jossey-Bass Publishers.
- Carroll, J. B. (1987). New Perspectives in the Analysis of Abilities. In R. R. Ronning, J. A. Glover & J. C. Conoley (Eds.), *The Influence of Cognitive Psychology on Testing* (pp. 267–284). Hillsdale, NJ: Lawrence Erlbaum.
- CEDEFOP (Ed.) (2003). *Terminology of Vocational Training Policy: A Multilingual Glossary for an Enlarged Europe*. http://www.europass.cedefop.eu.int/img/dynamic/c-313/cv-1_en_USglossary_4030_6k.pdf (Letzter Abruf: 20. September 2008).
- Chang, S.-H., Lin, P.-C. & Lin, Z. C. (2007). Measures of Partial Knowledge and Unexpected Responses in Multiple-Choice Tests. *Educational Technology & Society*, 10 (4), 95–109.
- Chase, W. G. & Simon, H. A. (1973). Perception in Chess. *Cognitive Psychology*, 4, 55–81.
- Chi, M. T. H. (1978). Knowledge Structure and Memory Development. In R. Siegler (Ed.), *Children's Thinking: What Develops?* (pp. 73–96). Hillsdale, NJ: Lawrence Erlbaum.
- Chi, M. T. H. (1985). Interactive Roles of Knowledge and Strategies in the Development of Organized Sorting and Recall. In S. F. Chipman, J. W. Segal & R. Glaser (Eds.), *Thinking and Learning Skills* (pp. 457–484). Hillsdale, NJ: Lawrence Erlbaum.
- Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P. & Glaser, R. (1989). Self-explanations: How Students Study and Use Example Learning to Solve Problems. *Cognitive Science*, 13, 145–182.

- Chi, M. T. H., Glaser, R. & Farr, M. J. (1988). *The Nature of Expertise*. Hillsdale, NJ: Lawrence Erlbaum.
- Chi, M. T. H., Hutchinson, J. & Robin, A. (1989). How Inference About Novel Domain-Related Concepts Can Be Constrained by Structured Knowledge. *Merrill-Palmer Quarterly*, 35, 27–62.
- Chickering, A. W. & Reisser, L. (1993). *Education and Identity*. San Francisco: Jossey-Bass Publishers.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1968). *Language and Mind*. New York, NY: Harcourt, Brace & World.
- Chomsky, N. (1977). *Reflexionen über die Sprache*. Frankfurt am Main: Suhrkamp.
- Chomsky, N. (1980). Rules and Representations. *The Behavioral and Brain Sciences*, 3, 1–61.
- Chomsky, N. (1981). *Regeln und Repräsentationen*. Frankfurt am Main: Suhrkamp.
- Clement, U. (2003). Competency Based Education and Training – eine Alternative zum Ausbildungsberuf. In R. Arnold (Hrsg.), *Berufsbildung ohne Beruf?* (S. 129–156). Hohengehren: Schneider.
- Cohen, R. S. & Wartofsky, M. W. (1983). *Epistemology, Methodology, and the Social Sciences*. Hingham, MA: Kluwer.
- Collins, A., Brown, J. S. & Newman, S. (1989). Cognitive Apprenticeship: Teaching Students the Craft of Reading, Writing, and Mathematics. In L. B. Resnick (Ed.), *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser* (pp. 89–114). Hillsdale, NJ: Lawrence Erlbaum.
- Commission of the European Communities (Ed.) (2005). *Towards a European Qualifications Framework for Lifelong Learning*. Brussels.
- Connel, M. W., Sheridan, K. & Gardner, H. (2003). On Abilities and Domains. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The Psychology of Abilities, Competencies, and Expertise* (pp. 126–155). Cambridge: Cambridge University Press.
- Cook, T. D. & Campbell, D. (Eds.) (1979). *Quasi-Experimentation. Design & Analysis Issues for Field Settings*. Chicago, IL: Rand McNally College Publishing.
- CTGV (1992). The Jasper Series as an Example of Anchored Instruction: Theory, Program, Description, and Assessment Data. *Educational Psychologist*, 27, 291–315.
- De Boeck, P. & Wilson, M. (Eds.) (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York, NY: Springer.
- De Corte, E., Op't Eynde, P. & Verschaffel, L. (2002). "Knowing What to Believe": The Relevance of Students' Mathematical Beliefs. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing* (pp. 297–320). Mahwah, NJ: Lawrence Erlbaum.
- Dehnbostel, P. & Lindemann, H.-J. (2007): Kompetenzen und Bildungsstandards in der schulischen und betrieblichen Berufsbildung. In: P. Dehnposter, H.-J. Lindemann & C. Ludwig (Hrsg.), *Lernen in der Arbeit in Schule und Betrieb* (S. 179–197). Münster: Waxmann.
- Department of Education and Science (Ed.) (1987). *National Curriculum Task Group on Assessment and Testing: A Report*. London: HMSO.
- Descy, P. & Tessaring, M. (2001). *Training and Learning for Competence. Second Report on Vocational Training Research in Europe: Executive Summary*. Luxembourg: Office for Official Publications of the European Communities.
- Deutsche Forschungsgemeinschaft (Hrsg.) (1990). *Berufsbildungsforschung an den Hochschulen der Bundesrepublik Deutschland. Denkschrift*. Weinheim: VCH.
- Dietrich, H. (2003). Scheme Participation and Employment Outcome of Young Unemployed People. Empirical Findings from Nine European Countries. In Policy Press (Eds.), *Youth Unemployment and Social Exclusion in Europe* (pp. 83–108). Bristol: Policy Press.
- Downing, S. M. & Haladyna, T. M. (Eds.) (2004). *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum.
- Doyle, W. (2000). *Authenticity*. Paper presented at the AERA Annual Meeting, New Orleans.
- Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over Machine*. New York: The Free Press.
- Dubs, R., Euler, D., Rüegg-Stürm, J. & Wyss, C. E. (2004). *Einführung in die Managementlehre*. Band 1. Bern, Stuttgart, Wien: Haupt Verlag.

- Elliott, J. (1989). Appraisal of Performance or Appraisal of Persons. In H. Simons & J. Elliott (Eds.), *Rethinking Appraisal and Assessment* (pp. 80–99). Milton Keynes: Open University Press.
- Ellström, P.-E. (1997). The many Meanings of Occupational Competence and Qualification. In W. J. Nijhof & J. N. Streumer (Eds.): *Key Qualifications in Work and Education* (pp. 39–50). Dordrecht: Kluwer Academic Publishers.
- Embretson, S. E. (1985). Multicomponent Latent Trait Models for Test Design. In S. E. Embretson (Ed.), *Test Design. Developments in Psychology and Psychometrics* (pp. 195–218). Orlando, FL: Academic Press.
- Embretson, S. E. (1998). A Cognitive Design System Approach to Generating Valid Tests: Application to Abstract Reasoning. *Psychological Methods*, 3, 380–396.
- Embretson, S. E. (1999). Issues in the Measurement of Cognitive Ability. In S. E. Embretson & S. L. Hershberger (Eds.), *The New Rules of Measurement* (pp. 1–15). Mahwah, NJ: Lawrence Erlbaum.
- Embretson, S. E. (2002). Generating Abstract Reasoning Items with Cognitive Theory. In S. Irvine & P. Kyllonen (Eds.), *Generating Items for Cognitive Tests: Theory and Practice* (pp. 35–60). Mahwah, NJ: Lawrence Erlbaum.
- Embretson, S. E. & Gorin, J. S. (2001). Improving Construct Validity with Cognitive Psychology Principles. *Journal of Educational Measurement*, 38 (4), 343–368.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response-Theorie for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Erpenbeck, J. & von Rosenstiel, L. (2003). *Handbuch Kompetenzmessung*. Stuttgart: Schäffer-Poeschel.
- European Commission (Ed.) (2008). *European Qualification Framework for Lifelong Learning (EQF)*. Luxembourg: Office for Official Publications of the European Communities.
- European State Secretariat for the Economy (Eds.) (2005). *Youth Unemployment*. Europaen Union: Brüssel.
- Evans, J. S. B. T., Newstead, S. E. & Byrne, R. M. J. (1993). *Human Reasoning: The Psychology of Deduction*. Hove, UK/Hillsdale, NJ: Lawrence Erlbaum.
- Evers, F. T., Rush, J. C. & Berdrow, I. (1998). *The Bases of Competence. Skills for Lifelong Learning and Employability*. San Francisco: Jossey-Bass Publishers.
- Faulstich, P. (1997). Kompetenz – Zertifikate – Indikatoren. In Arbeitsgemeinschaft Qualifikations-Entwicklungs-Management (Hrsg.), *Kompetenzentwicklung 97: Berufliche Weiterbildung in der Transformation – Fakten und Visionen* (S. 141–196). Münster: Waxmann.
- Fend, H. (1981). *Theorie der Schule*. München, Wien, Baltimore: Urban & Schwarzenberg.
- Fitts, P. M. & Posner, M. I. (1967). *Human Performance*. Belmont, CA: Brooks & Cole.
- Fleishman, E. A., Costanza, D. P & Marshall-Mies, J. (1999). Abilities. In N. G. Peterson, M. D. Mumford, W. C. Borman, P. R. Jeanneret & E. A. Fleishman (Eds.), *An Occupational Information System for the 21st Century: The Development of O*NET* (pp. 175–195). Washington, DC: American Psychological Association.
- Forman, S. L. & Steen, L. A. (1999). *Beyond Eighth Grade: Functional Mathematics for Life and Work*. Berkeley, CA: National Center for Research in Vocational Education.
- Frederiksen, J. R. & White, B. Y. (2004). Designing Assessments for Instruction and Accountability: An Application of Validity Theory to Assessing Scientific Inquiry. In M. Wilson (Ed.), *Towards Coherences Between Classroom Assessment and Accountability. 103rd Yearbook of the National Society for the Study of Education* (pp. 74–104). Chicago, IL: The University of Chicago Press.
- Frensch, P. A. (2006). Kognition. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 19–28). Göttingen: Hogrefe.
- Funke, J. (2004). Komplexes Problemlösen: Möglichkeiten eines deduktivistischen Vorgehens. In E. Erdfelder & J. Funke (Hrsg.), *Allgemeine Psychologie und deduktivistische Methodologie* (S. 281–300). Göttingen: Vandenhoeck & Ruprecht.
- Funke, J. & Frensch, P. A. (Hrsg.) (2006). *Handbuch der Allgemeinen Psychologie – Kognition*. Göttingen: Hogrefe.
- Gagné, R. M. (1977). *The Conditions of Learning*. New York, NY: Holt, Rinehart & Winston.

- Gagné, R. M. & Briggs, L. (1974). *Principles of Instructional Design*. New York, NY: Holt, Rinehart & Winston.
- Gelman, R. & Greeno, J. G. (1989). On the Nature of Competence: Principles for Understanding in a Domain. In L. B. Resnick (Ed.), *Knowing, Learning and Instruction. Essays in Honor of Robert Glaser* (pp. 125–186). Hillsdale, NJ: Lawrence Erlbaum.
- Gelman, S. A. (1996). Concepts and Theories. In R. Gelman & T. Kit-Fong Au (Eds.), *Perceptual and Cognitive Development* (pp. 117–184). San Diego, CA: Academic Press.
- Getsch U. & Preiß, P. (2003a). *Modellunternehmen Kettenfabrik A & S GmbH – Grundkurs Rechnungswesen – belegorientiert. Geschäftsjahre 2010-2014. Belege und Grafiken zum Bearbeiten und Lösungsheft*. Troisdorf: Bildungsverlag Eins.
- Getsch U. & Preiß, P. (2003b). *Modellunternehmen A & S GmbH – Grundkurs Rechnungswesen – belegorientiert. Anlagenkartei, Hauptbücher und Quartalsberichte für die Geschäftsjahre 2010-2014. Arbeitsheft und Lösungsheft*. Troisdorf: Bildungsverlag Eins.
- Gillen, J. & Kaufhold, M. (2005). Kompetenzanalysen – Kritische Reflexion von Begrifflichkeiten und Messmöglichkeiten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 101 (3), 364–378.
- Glaser, R. (1963). Instructional Technology and the Measurement of Learning Outcomes: Some Questions. *American Psychologist*, 18, 519–521.
- Glaser, R. (1976). Cognitive Psychology and Instructional Design. In D. Klahr (Ed.), *Cognition and Instruction* (pp. 303–316). Hillsdale, NJ: Lawrence Erlbaum.
- Glaser, R. (1981). The Future of Testing: A Research Agenda for Cognitive Psychology and Psychometrics. *American Psychologist*, 36, 923–936.
- Glaser, R., Lesgold, A. & Lajoie, S. (1987). Toward a Cognitive Theory for the Measurement of Achievement. In R. R. Ronning, J. A. Glover & J. C. Conoley (Eds.), *The Influence of Cognitive Psychology on Testing* (pp. 41–86). Hillsdale, NJ: Lawrence Erlbaum.
- Gobet, F. & Simon, H. A. (2000). Five Seconds or Sixty? Presentation Time in Expert Memory. *Cognitive Science*, 24 (4), 651–682.
- Goldman, S. R. (2009). Commentary: Explorations of Relationships Among Learners, Tasks, and Learning. *Learning and Instruction*, 19 (5), 451–454.
- Goldman, S. R. & Pellegrino, J. W. (1984). Deduction about Induction: Analyses of Developmental and Individual Differences. In R. J. Sternberg (Ed.), *Advances in Psychology of Human Intelligence* (pp. 149–197). Hillsdale, NJ: Lawrence Erlbaum.
- Greeno, J. G. (with the Middle School Mathematics Through Applications Project Group) (1998). The Situativity of Knowing, Learning, and Research. *American Psychologist*, 53, 5–26.
- Greeno, J. G. (1994). Gibson's Affordance. *Psychological Review*, 101, 336–342.
- Greeno, J. G., Riley, M. S. & Gelman, R. (1984). Conceptual Competence and Children's Counting. *Cognitive Psychology*, 16, 94–143.
- Gruber, H. (1999). *Erfahrung als Grundlage kompetenten Handelns*. Bern: Huber.
- Gruber, H. & Mandl, H. (1996). Das Entstehen von Expertise. In J. Hoffmann & W. Kintsch (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C: Theorie und Forschung, Serie II Kognition, Band 7, Lernen* (S. 583–615). Göttingen: Hogrefe.
- Gustafsson, J.-E. & Undheim, J. O. (1996). Individual Differences in Cognitive Functions. In D. Berliner & R. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 186–242). New York, NY: Macmillan.
- Hanf, G., Blötz, U., Le Mouillour, I., Mucke, K., Mettin, G., Rein, V. & Reymers, M. (2006). *Fachlicher Prüfbericht zu den Grundbegriffen und Deskriptoren des Entwurfs für einen Europäischen Qualifikationsrahmen*. <http://www.bibb.de/de/25717.htm> (Letzter Abruf: 18. November 2008).
- Hannan, D. F. & Werquin, P. (2000). Education and Labour Market Change: The Dynamics of Education to Work Transitions in Europe. In Cedefop (Eds.), *Second Report on Vocational Training Research in Europe: Background Report* (pp. 91–136). Luxembourg: Office for Official Publications of the European Communities.
- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S. 83–99). Weinheim: Beltz.
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Berlin: Springer.

- Hashway, R. M. (1998). *Assessment and Evaluation of Developmental Learning. Qualitative Individual Assessment and Evaluation Models*. Westport: Praeger.
- Hasselhorn, M. (1998). Metakognition. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 348–351). Weinheim: Beltz.
- Hasselhorn, M. (2000). Metakognition als ein zentrales Konzept lebenslangen Lernens. In F. Achtenhagen & W. Lempert (Hrsg.), *Lebenslanges Lernen im Beruf – seine Grundlegung im Kindes- und Jugendalter* (Bd. 3) (S. 41–53). Opladen: Leske + Budrich.
- Hays, R. T. (2006). *The Science of Learning: A Systems Theory Approach*. Boca Raton, FL: Brown-Walker Press.
- Heidegger, G. & Petersen, W. (2004). *Accreditation of Non-Formal and Informal Learning in Germany*. EPANIL (LdV project) first meeting, October 2004.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Beltz Test.
- Hensge, K., Görmär, G., Lorig, B., Molitor, H. & Schreiber, D. (2008). *Kompetenzstandards in der Berufsausbildung. Forschungsprojekt 4.3.201, Zwischenbericht*. Bonn: Bundesinstitut für Berufsbildung.
- v. Hentig, H. (2002). *Bildung*, 4. Auflage, Weinheim und Basel: Beltz.
- Hirschfeld, L. A. & Gelman, S. A. (1994). Toward a Topography of Mind: An Introduction to Domain Specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 3–35). Cambridge, MA: Cambridge University Press.
- Höft, S. & Funke, U. (2001) Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 135–174). Göttingen: Hogrefe.
- Hofer, B. K. & Pintrich, P. R. (Eds.) (2002). *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*. Mahwah, NJ: Lawrence Erlbaum.
- Hoffmann, W. (1922). *Die Reifezeit: Probleme der Entwicklungspsychologie und Sozialpädagogik*. Leipzig: Quelle und Meyer.
- Hofmeister, W. (2005). Erläuterung der Klassifikationsmatrix zum ULME-Kompetenzstufenmodell. *bwp@ – Berufs- und Wirtschaftspädagogik online*, 8.
- Hofstadter, D. R. (1985). *Metamagical Themas: Questing for the Essence of Mind and Pattern*. New York, NY: Basic.
- International Life Skills Survey (ILSS) (2000). *Policy Research Initiative*. Statistics Canada.
- IPN (Leibniz-Institut für die Pädagogik der Naturwissenschaften) (Hrsg.) (2003). *Ziele und Perspektiven naturwissenschaftlicher Bildung*. Kiel: IPN.
- Isen, A. M. (2000). Positive Affect and Decision Making. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotions* (pp. 417–435). New York, NY: Guilford.
- John, E. G. (2004). *Das Wirtschaftsgymnasium: Realität und bildungspolitischer Anspruch*. Rinteln: Merkur.
- Jude, N. & Klieme, E. (2007). Sprachliche Kompetenz aus Sicht der pädagogisch-psychologischen Diagnostik. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S. 9–22). Weinheim: Beltz.
- Kilpatrick, L., Swafford, J. & Findell, B. (Eds.) (2001). *Adding it up: Helping Children Learn Mathematics*. National Research Council. Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Kim, J. K. & Nicewander, W. A. (1992). Ability Estimation for Conventional Tests. *Psychometrika*, 58 (4), 587–599.
- Kirschner, P., v. Vilsteren, P., Hummel, H. & Wigman, M. (1997) The Design of a Study Environment for Acquiring Academic and Professional Competence. *Studies in Higher Education*, 22 (2), 151–171.
- Klein, G. & Hoffman, R. R. (2008). Macrocognition, Mental Models, and Cognitive Tasks Analysis Methodology. In J. M. Schraagen, L. G. Militello, T. Ormerod & R. Lipshitz (Eds), *Naturalistic Decision Making and Macrocognition* (pp. 57–80). Aldershot: Ashgate.
- Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, 56, 10–13.

- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Frankfurt am Main: DIPF.
- Klieme, E. & Hartig, J. (2008). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaften. Sonderheft 8/2007* (S. 11–29). Wiesbaden: VS Verlag.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 139–190). Opladen: Leske + Budrich.
- KMK (Hrsg.) (2000a). *Handreichungen für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz (KMK) für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland: Bonn.
- KMK (Hrsg.) (2000b). *Vereinbarung zur Gestaltung der gymnasialen Oberstufe in der Sekundarstufe II* (Beschluss der Kultusministerkonferenz vom 7.7.1972 i. d. F. vom 16.6.2000 – Anlagen nach dem Stand der Fortschreibung vom 18.03.2005). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland: Bonn.
- KMK (Hrsg.) (2001). *Erste Konsequenzen aus den Ergebnissen der PISA-Studie*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland: Bonn.
- KMK (Hrsg.) (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Bildungsabschluss, Beschluss vom 04.12.2003*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland: Bonn.
- Köller, O., Watermann, R., Trautwein, U. & Lüdtke, O. (Hrsg.) (2004). *Wege zur Hochschulreife in Baden-Württemberg. TOSCA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien*. Opladen: Leske + Budrich.
- Kreitzer, A. E. & Madaus, G. F. (1994). Empirical Investigations of the Hierarchical Structure of the Taxonomy. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's Taxonomy. A Forty-year Retrospective. 93rd Yearbook of the National Society for the Study of Education. Part II* (pp. 64–81). Chicago, IL: The University of Chicago Press.
- Kremer, H.-H. (2007). Lernen in medienbasierten Lernumgebungen – Modelversuch Kool. *Die Berufsbildende Schule*, 59, 216–220.
- Kretzschmar, J. R. (1912). *Entwicklungspsychologie und Erziehungswissenschaft: eine pädagogische Studie auf entwicklungstheoretischer, ethnologischer und kulturhistorischer Grundlage*. Leipzig: Wunderlich.
- Kuhl, J. (1996). Wille und Freiheitserleben. Formen der Selbststeuerung. In J. Kuhl & H. Heckhausen (Hrsg.), *Motivation, Volition und Handlung. Enzyklopädie der Psychologie* (C/IV/4, S. 665–765). Göttingen: Hogrefe.
- Kunter, M., Klusman, U. & Baumert, J. (2009). Professionelle Kompetenz von Mathematiklehrkräften: Das COACTIV-Modell. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messungen* (S. 153–166). Weinheim: Beltz.
- Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A., Krauss, S., Löwen, K., Neubrand, M. & Tsai, Y.-M. (2006). Linking Aspects of Teacher Competence to Their Instruction. In M. Prenzel (Ed.), *Studies on the Educational Quality of Schools. The Final Report on the DFG Priority Programme* (pp. 39–59). Münster: Waxmann.
- Langford P. E. & Hunting, R. (1994). A Representational Communication Approach to the Development of Inductive and Deductive Reasoning. In A. Demetriou & A. Efklides (Eds.) *Intelligence, Mind and Reasoning: Perspectives on Individual Differences. Advances in Psychology* (pp. 193–233). Amsterdam: Psychology Press.
- Laux, H. (2007). *Entscheidungstheorie*. 7. Aufl. Berlin, Heidelberg, New York, NY: Springer.

- Lave, J. & Wenger, E. (1990). *Situated Learning: Legitimate Peripheral Participation*. Palo Alto, CA: Institute for Research in Learning.
- Lehmann, R. & Seeber, S. (Hrsg.) (2007). *ULME III. Untersuchungen von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen*. Hamburg: HIBB.
- Lehmann, R., Seeber, S. & Hunger, S. (2006). *ULME II – Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der teilqualifizierenden Berufsfachschulen*. Hamburg: Behörde für Bildung und Sport.
- Leschinsky, A. & Cortina, K. S. (2008). Zur sozialen Einbettung bildungspolitischer Trends in der Bundesrepublik. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 21–52). Reinbek: Rowohlt.
- Lesgold, A., Lajoie, S., Logan, D. & Eggan, G. (1990). Applying Cognitive Task Analysis and Research Methods to Assessment. In N. Frederikson, R. Glaser, A. Lesgold & M. Shafto (Eds.), *Diagnostic Monitoring of Skill and Knowledge Acquisition* (pp. 234–268). Hillsdale, NJ: Lawrence Erlbaum.
- Lienert, G. A. (1969). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lindemann, H.-J. (2008). Prüfungsaufgabenerstellung – Qualifizierung von Lehrkräften für diese Aufgabe. *bwp@ – Berufs- und Wirtschaftspädagogik online, Spezial 4*.
- Linhares, A. & Brum, P. (2007). Understanding Our Understanding of Strategic Scenarios: What Role Do Chunks Play? *Cognitive Science*, 31 (6), 989–1007.
- Lisbon-to-Copenhagen-to-Maastricht Consortium Partners (Eds.) (2004): *Achieving the Lisbon Goal: The Contribution of VET – Final Report to the European Commission*. http://europa.eu.int/comm/education/policies/2010/studies/maastricht_en.pdf (Letzter Abruf: 20. September 2008).
- Luchins, A. S. (1942). Mechanization in Problem Solving. *Psychological Monographs*, 54, 4–36.
- Mandl, H., Gruber, H. & Renkl, A. (1995). Situiertes Lernen in multimedialen Lernumgebungen. In L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia* (S. 167–178). Weinheim: Psychologie Verlags Union.
- Mandl, H., Prenzel, M. & Gräsel, C. (1991). *Das Problem des Lerntransfers in der betrieblichen Weiterbildung (Forschungsbericht Nr. 1)*. München: Ludwig-Maximilians-Universität, Institut für Empirische Pädagogik und Pädagogische Psychologie.
- Markman, A. B. (1999). *Knowledge Representation*. Mahwah, NJ: Lawrence Erlbaum.
- Markowitsch, J. & Luomi-Messerer, K. (2008). Development and Interpretation of Descriptors of the European Qualifications Framework. *European Journal of Vocational Training*, 43, 33–58.
- Marzano, R. J. & Kendall, J. S. (2007). *The New Taxonomy of Educational Objectives*. Thousand Oaks, CA: Corwin Press.
- Marzano, R. J. & Kendall, J. S. (2008). *Designing & Assessing Educational Objectives. Applying the New Taxonomy*. Thousand Oaks, CA: Corwin Press.
- Masters, G. N. (1982). A Rasch Model for partial Credit Scoring. *Psychometrika*, 47, 149–174.
- Mayer, R. E. (2004). Teaching of Subject Matter. *Annual Review of Psychology*, 55, 715–744.
- McCloy, R., Waugh, G. & Medsker, G. (1998). *Determining the Occupational Reinforcer Patterns for O*NET Occupational Units*. Alexandria, VA: Human Resources Research Organization.
- McGuinness, D., McGuinness, C. & Donohue, J. (1995). Phonological Training and the Alphabet Principle: Evidence for Reciprocal Causality. *Reading Research Quarterly*, 30, 830–852.
- Mertens, D. (1974). Schlüsselqualifikationen: Thesen zur Schulung für eine moderne Gesellschaft. *Mitteilungen aus Arbeitsmarkt- und Berufsforschung*, 7, 36–43.
- Messick, S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments. *Educational Researcher*, 23 (2), 13–23.
- Messner, H. (1978). *Wissen und Anwenden. Zur Problematik des Transfers im Unterricht*. Stuttgart: Klett.
- Minnameier, G. (2005). Wissen und Können im Kontext inferentiellen Denkens. In H. Heid & C. Harteis (Hrsg.), *Verwertbarkeit. Ein Qualitätskriterium (erziehungs-)wissenschaftlichen Wissens?* (S. 183–203). Wiesbaden: VS Verlag.

- Mislevy, R. J. (1994). Evidence and Inference in Educational Assessment. *Psychometrika*, 59, 439–483.
- Mislevy, R. J. (2007). Cognitive Psychology and Educational Assessment. *Educational Measurement*, 128, 257–305.
- Mislevy, R. J. (2008). Issues of Structure and Issues of Scale in Assessment from a Situative/Sociocultural Perspective. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel & L. J. Young (Eds.), *Assessment, Equity, and Opportunity to Learn* (pp. 259–294). Cambridge, MA: Cambridge University Press.
- Mislevy, R. J., Almond, R. G. & Lukas, J. F. (2004). *A Brief Introduction to Evidence-Centered Design (CSE Technical Report 632)*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, UCLA.
- Mislevy, R. J. & Haertel, G. D. (2006). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice*, 25 (4), 6–20.
- Mislevy, R. J. & Riconscente, M. M. (2004). Evidence-Centered Assessment Design. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 61–90). Mahwah, NJ: Lawrence Erlbaum.
- Mislevy, R. J. & Riconscente, M. M. (Eds.) (2005). *Evidence-Centered Assessment Design: Layers, Structures, and Terminology. PADI Technical Report 9*. Menlo Park, CA: SRI International.
- Mislevy, R. J., Steinberg, L. S. & Almond, R. G. (2003). On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3–66.
- Moore, S. (1996). Estimating DIF with the RCML Model. In G. Engelhard & M. Wilson (Eds.), *Objective Measurement III: Theory into Practice* (pp. 219–238). Norwood, NJ: Ablex.
- Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H. & Young, L. J. (Eds.) (2008). *Assessment, Equity, and Opportunity to Learn*. Cambridge, MA: Cambridge University Press.
- Mühlemann, S., Wolter, S. C., Fuhrer, M. & Wüest, A. (2007). *Lehrlingsausbildung – ökonomisch betrachtet. Ergebnisse der zweiten Kosten-Nutzen-Studie*. Chur, Zürich: Rüegger.
- Müller, K., Fürstenau, B. & Witt, R. (2007). Ökonomische Kompetenz sächsischer Mittelschüler und Gymnasiasten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 103 (2), 227–247.
- Mummendey, H. D. (2003). *Die Fragebogen-Methode*. Göttingen: Hogrefe.
- Murnane, R. & Levy, F. (1996). *Teaching the New Basic Skills: Principles for Educating Children to Thrive in a Changing Economy*. New York, NY: Free Press.
- Murphy, P. K. & Alexander, P. A. (2006). *Understanding How Students Learn. A Guide for Instructional Leaders*. Thousand Oaks, CA: Corwin Press.
- National Center for Education Statistics (NCES) (1993). *Adult Literacy in America. Report of the National Adult Literacy Survey (NALS)*. Washington, DC: U.S. Department of Education.
- National Council on Education and the Disciplines (NCED) (2001). *The Case for Quantitative Literacy*. Washington, DC: U.S. Department of Education.
- Neubrand, M., Klieme, E., Lüdtke O. & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung. *Unterrichtswissenschaft. Zeitschrift für Lernforschung*, 30 (2), 100–119.
- Neuweg, G. H. (2005). Emergenzbedingungen pädagogischer Könnerschaft. In H. Heid & C. Harteis (Hrsg.), *Verwertbarkeit. Ein Qualitätskriterium (erziehungs-)wissenschaftlichen Wissens?* (S. 205–228). Wiesbaden: VS Verlag.
- Nickolaus, R. (2008). Vorstellungen zur Modellierung beruflicher Handlungskompetenz und erste Versuche zu ihrer empirischen Prüfung. In R. Nickolaus & H. Schanz (Hrsg.), *Didaktik der gewerblich-technischen Berufsbildung* (S. 87–102). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Nickolaus, R., Gschwendtner, T. & Abele, S. (2009). *Die Validität von Simulationsaufgaben am Beispiel der Diagnosekompetenz von Kfz-Mechatronikern. Vorstudie zur Validität von Simulationsaufgaben im Rahmen eines VET-LSA*. Abschlussbericht für das Bundesministerium für Bildung und Forschung: Universität Stuttgart.
- Nickolaus, R., Gschwendtner, T. & Geißel, B. (2008): Modellierung und Entwicklung beruflicher Fachkompetenz in der gewerblich-technischen Erstausbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104, 48–73.

- Nickolaus, R., Knöll, B. & Gschwendtner, T. (2007). Innovations in Vocational Education and Difficulties in Their Empirical Substantiation. *European Journal of Vocational Training*, 40, 22–37.
- Nickolaus, R. & Ziegler, B. (2005): Der Lernerfolg schwächerer Schüler in der beruflichen Ausbildung im Kontext methodischer Entscheidungen. In: P. Gonon, F. Klauser, R. Nickolaus & R. Huisinga, R. (Hrsg.), *Kompetenz, Kognition und neue Konzepte der beruflichen Bildung* (S. 161–175) Wiesbaden: VS Verlag.
- Niedersächsisches Kultusministerium (2001). *Rahmenrichtlinien für die Unterrichtsfächer Betriebswirtschaft mit Rechnungswesen/Controlling, Informationsverarbeitung und Volkswirtschaft im Fachgymnasium – Wirtschaft*. (Stand: Mai 2001). Hannover.
- Noris, N. (1991). The Trouble with Competence. *Cambridge Journal of Education*, 21 (3), 330–341.
- OECD (Ed.) (1995). *Literacy, Economy and Society. Results of the First International Adult Literacy Survey*. Paris: OECD.
- OECD (Ed.) (1999). *Measuring Student Knowledge and Skills. A New Framework for Assessment*. Paris: OECD.
- OECD (Ed.) (2000). *From Initial Education to Working Life. Making Transitions Work*. Paris: OECD.
- OECD (Ed.) (2001). *Lernen für das Leben – Erste Ergebnisse von PISA 2000*. Paris: OECD.
- OECD (Ed.) (2002). *Manual for the PISA 2000 Database*. Paris: OECD.
- OECD (Ed.) (2003). *The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris: OECD.
- OECD (Ed.) (2005). *PISA 2003 Data Analysis Manual*. Paris: OECD.
- OECD (Ed.) (2008). *The OECD Programme for the International Assessment of Adult Competencies (PIAAC)*. Paris: OECD.
- Oelkers, J. & Reusser, K. (2008). *Expertise: Qualität entwickeln – Standards sichern – mit Differenz umgehen*. Berlin: BMBF.
- Opwis, K. (2006). Mathematische und Computer-Modellierung. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 726–731). Göttingen: Hogrefe.
- Oser, F. (2001). Standards: Kompetenzen von Lehrpersonen. In F. Oser & J. Oelkers (Hrsg.), *Die Wirksamkeit der Lehrerbildungssysteme* (S. 215–342). Zürich: Rüegger.
- Paulsen, M. B. & Wells, C. T. (1998). Domain differences in the epistemological beliefs of college students. *Research in Higher Education*, 39 (4), 365–382.
- Pellegrino, J., Chudowsky, N. & Glaser, R. (Eds.) (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. National Research Council Committee on the Foundations of Assessment. Washington, DC: National Academy Press.
- Perkins, D. N. (1992). *Smart Schools: From Training Memories to Educating Minds*. New York, NY: Free Press.
- Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R. & Fleishman, E. A. (Eds.). *An Occupational Information System for the 21st Century: The Development of O*NET*. Washington, DC: American Psychological Association.
- Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 451–502). San Diego, CA: Academic Press.
- Pintrich, P. R. (2004). A Conceptual Framework for assessing Motivation and Self-Regulated Learning in College Students. *Educational Psychology Review*, 16, 385–407.
- Pintrich, P. R. & de Groot, E. V. (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Educational Psychology*, 82, 33–40.
- Pollock, J. L. (1999). *Contemporary Theories of Knowledge*. Lanham: Rowman & Littlefield Publishers.
- Popham, W. J. (1993). *Educational Evaluation*. Needham Heights, MA: Allyn and Bacon.
- Preiß, P. (2005). Entwurf eines Kompetenzkonzepts für den Inhaltsbereich Rechnungswesen/Controlling. In P. Gonon, F. Klauser, R. Nickolaus & R. Huisinga (Hrsg.), *Kompetenz, Kognition und neue Konzepte der beruflichen Bildung* (S. 67–85). Wiesbaden: VS Verlag.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J. & Schiefele, U. (Hrsg.) (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche*. Münster: Waxmann.

- Prenzel, M., Walter, O. & Frey, A. (2007). PISA misst Kompetenzen. Eine Replik auf Rindermann (2006): Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, 58 (2), 128–136.
- Pressley, M., Borkowski, J. G. & Schneider, W. (1987). Cognitive Strategies: Good Strategy Users Coordinate Metacognition and Knowledge. *Annals of Child Development*, 4, 89–129.
- Raffe, D. (2005). National Qualifications Framework as Integrated Qualifications Frameworks. *SAQA Bulletin*, 8, 24–36.
- Raffe, D. (2008). The Concept of Transition System. *Journal of education and work*, 21 (4), 277–296.
- Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Copenhagen: The Danish Institute of Educational Research.
- Rauner, F., Grollmann, P. & Martens, T. (2007). Messen beruflicher Kompetenz(entwicklung). *ITB-Forschungsberichte*, 21. Bremen: Institut Technik und Bildung.
- Rauner, F., Haasler, B., Heinemann, L. & Grollmann, P. (Hrsg.) (2009). *Messen beruflicher Kompetenzen. Teilband 1: Grundlagen und Konzeption des KOMET-Projektes*. Münster: Lit Verlag.
- Rebmann, K., Tredop, D., Klattenhoff, K., Schulze, G. & Wittrock, M. (2007). *Empirische Studien zur Ausbildungsreife im Urteil von Lehrkräften und Unternehmensvertreter(inne)n aus der Weser-Ems-Region*. Projektberichte und Materialien zur Berufs- und Wirtschaftspädagogik, Heft 37.
- Reeff, J.-P. (2007). Technische Lösungen für ein computer- und internetbasiertes Assessmentsystem. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. Bildungsforschung Band 20* (S. 81–91). Bonn, Berlin: BMBF.
- Reetz, L. (1989a). Zum Konzept der Schlüsselqualifikationen in der Berufsbildung. *Berufsbildung in Wissenschaft und Praxis*, 5, 3–10.
- Reetz, L. (1989b). Zum Konzept der Schlüsselqualifikationen in der Berufsbildung. *Berufsbildung in Wissenschaft und Praxis*, 6, 24–30.
- Reetz, L. (1990). Zur Bedeutung der Schlüsselqualifikationen in der Berufsbildung. In L. Reetz & T. Reitmann (Hrsg.), *Schlüsselqualifikationen* (S. 16–35). Hamburg: Feldhaus.
- Reetz, L. (1999). Zum Zusammenhang von Schlüsselqualifikationen – Kompetenzen – Bildung. In T. Tramm, D. Sembill, F. Klauser & E. G. John (Hrsg.), *Professionalisierung kaufmännischer Berufsbildung* (S. 32–51). Frankfurt: Peter Lang.
- Reich, R. B. (1991). *The Work of Nations: Preparing Ourselves for the 21st Century Capitalism*. New York, NY: Vintage.
- Reif, F. (2008). *Applying Cognitive Science to Education. Thinking and Learning in Scientific and Other Complex Domains*. Cambridge, MA: MIT Press.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? *Psychologische Rundschau*, 57, 69–86.
- Roberts, L., Wilson, M. & Draney, K. (1997). *The SEPUP Assessment System: An Overview*. University of California, Berkeley: BEAR Report Series, SA-97-1.
- Rohwer, W. D. & Sloane, K. (1994). Psychological Perspectives. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's Taxonomy. A Forty-year Retrospective. 93rd Yearbook of the National Society for the Study of Education. Part II* (pp. 41–63). Chicago, IL: The University of Chicago Press.
- Rosendahl, J., Fehring, G. & Straka, G. A. (2008). Lernkompetenz bei Bankkaufleuten in der beruflichen Erstausbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104 (2), 201–214.
- Rost, J. (2006). Zum Einsatz der Item-Response-Theorie für die Messung berufsbezogener Kompetenzen im Rahmen der Studie „Berufsbildungs-PISA“ – Expertise. In M. Baethge, F. Achtenhagen, L. Arends, E. Babic, V. Baethge-Kinsky & S. Weber (Hrsg.), *Berufsbildungs-PISA – Machbarkeitsstudie* (S. XXXIV–XXXVII). Stuttgart: Steiner.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. 2. Aufl. Bern: Huber.
- Roth, H. (1971). *Pädagogische Anthropologie. Bd. 2: Entwicklung und Erziehung. Grundlagen einer Entwicklungspädagogik*. Hannover: Schroedel.
- Rüegg-Stürm, J. (2004). Das neue St. Galler Management Modell. In R. Dubs, D. Euler, J. Rüegg-Stürm & C. E. Wyss (Hrsg.), *Einführung in die Managementlehre* (S. 65–134). Bern: Haupt.

- Rychen, S. & Salganik, L. H. (Eds.) (2001). *Defining and Selecting Key Competencies*. Göttingen: Hogrefe.
- Rychen, S. & Salganik, L. H. (Eds.) (2003). *Key Competencies for a Successful Life and a Well-Functioning Society*. Göttingen: Hogrefe.
- Schacter, J. & Fagnano, C. (1999). Does Computer Technology Improve Student Learning and Achievement. How, When and Under What Conditions? *Journal of Educational Computing Research*, 20, 329–343.
- Scheeres, H. & Hager, P. (1994). *Competences and the Curriculum*. <http://ftp.swin.edu.au/pub/aare/aare94/conf94/scheh94.233> (Letzter Abruf: 17 November 2008).
- Schelten, A. (2004). Schlüsselqualifikationen. *Wirtschaft und Berufserziehung. Zeitschrift für Berufsbildung*, 56 (4), 11–13.
- Schneider, W. (1997). The Impact of Expertise on Performance: Illustration from Developmental Research on Memory and Sports. *High Ability Studies*, 8, 7–18.
- Schnotz, W. & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, 19, 469–508.
- Schommer, M. (1993). Epistemological Development and Academic Performance among Secondary Students. *Journal of Educational Psychology*, 85 (3), 406–411.
- Schommer, M. & Dunnell, P. A. (1994). A Comparison of Epistemological Beliefs between Gifted and Non-Gifted Highschool Students. *Roeper Review*, 16 (3), 207–210.
- Schraagen, J. M., Militello, L. G., Ormerod, T. & Lipshitz, R. (Eds.) (2008). *Naturalistic Decision Making and Macrocognition*. Aldershot: Ashgate.
- Schumann, S. & Eberle, F. (2009). Überlegungen zur Erfassung von langfristigen Effekten der Professionalität von Lehrenden. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messungen* (S. 717–728). Weinheim: Beltz.
- Schwippert, K. (2005). Vergleichende Lernstandsuntersuchungen, Bildungsstandards und die Steuerung von schulischen Bildungsprozessen. *bwp@ – Berufs- und Wirtschaftspädagogik online*, 8.
- Seeber, S. (2005). Zur Erfassung und Vermittlung berufsbezogener Kompetenzen im teilqualifizierenden Bildungsgang „Wirtschaft und Verwaltung“ an Hamburger Berufsfachschulen. *bwp@ – Berufs- und Wirtschaftspädagogik online*, 8.
- Seeber, S. (2007). Zur Anforderungsstruktur eines Fachleistungstests für Auszubildende des Berufs Einzelhandelskaufmann/Einzelhandelskauffrau. In D. Münk, J. van Buer, K. Breuer & T. Deißinger (Hrsg.), *Hundert Jahre kaufmännische Ausbildung in Berlin* (S. 184–193). Opladen: Budrich.
- Seeber, S. (2008). Ansätze zur Modellierung beruflicher Fachkompetenz in kaufmännischen Ausbildungsberufen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104, 74–97.
- Seifried, J., Sembill, D., Nickolaus, R. & Schelten, A. (2005). Analysen systemischer Wechselwirkungen beruflicher Bildungsprozesse – Forschungsstand und Forschungsperspektiven beruflicher Bildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 101, 601–618.
- Seifried, J. & Ziegler, B. (2009). Domänenbezogene Professionalität. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messungen* (S. 83–92). Weinheim: Beltz.
- Seidel T. & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research*, 77 (4), 454–499.
- Selin, B. (2008). The Proposal for a European Qualifications Framework. Making it a Reality – Possibilities and Limitations. *European Journal of Vocational Training*, 43, 4–18.
- Sembill, D. & Seifried, J. (2007). Selbstorganisiertes Lernen und Unterrichtsqualität. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 401–412). Frankfurt a. M.: Lang.
- Shavelson, R. J. (2008). Reflections on Quantitative Reasoning: An Assessment Perspective. In B. L. Madison & L. A. Steen (Eds.), *Calculation vs. Context: Quantitative Literacy and Its Implications for Teacher Education* (pp 27–47). Mathematical Association of America.

- Shavelson, R. J. (2007). *Reflections on the Assessment of Job Performance and Competence*. Paper presented at the Herbsttagung Sektion Berufs- und Wirtschaftspädagogik, Göttingen.
- Shavelson, R. J., Baxter, G. P. & Pine, J. (1991). Performance Assessment in Science. *Applied Measurement in Education*, 4, 347–362.
- Shavelson, R. J., Baxter, G. P. & Pine, J. (1992). Performance Assessments: Political Rhetoric and Measurement Reality. *Educational Researcher*, 21 (4), 22–27.
- Shavelson, R. J., Ruiz-Primo, M. A. & Wiley, E. W. (1999). Note on Sources of Sampling Variability in Science Performance Assessments. *Journal of Educational Measurement*, 36, 61–71.
- Shavelson, R. J., Young D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M. K. & Yin, Y. (2008). On the Impact of Formative Assessment on Student Motivation, Achievement, and Conceptual Change. *Applied Measurement in Education*, 21 (4), 335–359.
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, 86, 420–428.
- Shulman, L. S. (1986). Paradigms and Research Programs in the Study of Teaching: A contemporary Perspective. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 3–36). New York, NY: Macmillan.
- Shulman, L. S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, 57 (1), 1–22.
- Shulman, L. S. (1998). Theory, Practice, and the Education of Professionals. *The Elementary School Journal*, 98 (5), 511–526.
- Shulman, L. S. (2004). *The Wisdom of Practice. Essays on Teaching, Learning, and Learning to Teach*. San Francisco, CA: Jossey-Bass.
- Siemon, J. (2003). *Evaluation eines komplexen Lehr-Lern-Arrangements. Eine netzwerk- und inhaltsanalytische Studie am Beispiel der Einführung in ein Modellunternehmen*. Wiesbaden: DUV.
- Simonton, D. K. (2003). Scientific Creativity as Constrained Stochastic Behavior: The Integration of Product, Person, and Process Perspectives. *Psychological Bulletin*, 129, 475–494.
- Singley, M. K. & Anderson, J. R. (1989). *The Transfer of Cognitive Skill*. Cambridge, MA: Harvard University Press.
- Sloane, P. F. E. (2005). Standards von Bildung – Bildung von Standards. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 101, 484–496.
- Sloane, P. F. E. & Dilger, B. (2005). The Competence Clash – Dilemmata bei der Übertragung des 'Konzepts der nationalen Bildungsstandards' auf die berufliche Bildung. *bwp@ – Berufs- und Wirtschaftspädagogik online*, 8.
- Sonntag, K. & Schaper, N. (1999). Förderung beruflicher Handlungskompetenz. In K. Sonntag (Hrsg.), *Personalentwicklung in Organisationen* (S. 211–260). Göttingen: Hogrefe.
- Spada, H. & Wichmann, S. (1996). Kognitive Determinanten der Lernleistung. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich D Praxisgebiete, Serie 1 Pädagogische Psychologie, Band 2 Psychologie des Lernens und der Instruktion* (S. 119–152). Göttingen: Hogrefe.
- Stamm, M. (2006). Youth Unemployment. Outline of a Psycho-Social Perspective. *European Journal of Vocational Training*, 39, 105–114.
- Sternberg, R. J. (1988). A Three-Faceted Model of Creativity. In R. J. Sternberg (Ed.), *The Nature of Creativity* (pp. 124–147). New York, NY: Cambridge University Press.
- Sternberg, R. J. & Ben-Zeev, T. (2001). *Complex Cognition. The Psychology of Human Thought*. New York, Oxford: Oxford University Press.
- Stevens, R. & Wineburg, S. (2005). Comparative Understanding of School Subjects: Past, Present, and Future. *Review of Educational Research*, 75(2), 125–157.
- Stiller, L. (1998). Schlüsselqualifikationen – Neuordnung/Ordnungsmittel. In E. Wittmann & J. van Buer (Hrsg.), *Schlüsselqualifikationen zwischen bildungspolitischem Anspruch, wissenschaftlicher Grundlegung und wissenschaftsadäquater Umsetzung* (S. –15). Berlin: Humboldt Universität.
- Straka, G. A., Fehring, G. & Rosendahl, J. (2008). *Modellvalidierung beruflicher Fachkompetenz und ihre Veränderung während der Ausbildung von angehenden Bankkaufleuten*. Vortrag Herbsttagung der Sektion für Wirtschaftspädagogik. Darmstadt.

- Straka, G. A. & Macke G. (2009). Berufliche Kompetenz: Handeln können, wollen und dürfen. Zur Klärung eines diffusen Begriffs. *Berufsbildung in Wissenschaft und Praxis – BWP*, 3, 15–21.
- Sweller, J. (1988). Cognitive Load during Problem Solving: Effects on Learning. *Cognitive Science*, 12, 257–285.
- Sweller, J., van Merriënboër, J. J. G. & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10, 251–296.
- Tenorth, H.-E. (2004). Bildungsstandards und Kerncurriculum. Systematischer Kontext, bildungstheoretische Probleme. *Zeitschrift für Pädagogik*, 50 (5), 650–661.
- Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz*. Münster: Institut für Schulbildung und allgemeine Didaktik.
- Terhart, E. (2006). Was wissen wir über gute Lehrer? *Pädagogik*, 58 (5), 42–47.
- Terhart, E. (2009). Erste Phase: Lehrerbildung an der Universität. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messungen* (S. 425–438). Weinheim: Beltz.
- Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge, MA: Cambridge University Press.
- Tramm, T. (1992). Grundzüge des Göttinger Projektes “Lernen, Denken, Handeln in komplexen Situationen – unter Nutzung neuer Technologien in der kaufmännischen Berufsausbildung”. In F. Achtenhagen & E. G. John (Hrsg.), *Mehrdimensionale Lehr-Lern-Arrangements: Innovation der kaufmännischen Ausbildung* (S. 1–27). Wiesbaden: Gabler.
- Trautwein, U., Lüdtke, O. & Beyer, B. (2004). Rauchen ist tödlich, Computerspiele machen aggressiv? Allgemeine und theorie-spezifische epistemologische Überzeugungen bei Studierenden unterschiedlicher Fachrichtungen. *Zeitschrift für Pädagogische Psychologie*, 18, 187–199.
- TREE (Hrsg.) (2008). *Projekt-Dokumentation 2000-2008*. Bern/Basel: TREE.
- U. S. Department of Labor Employment and Training Administration (Ed.) (2006). *Testing and Assessment: A Guide to Good Practice to Workforce Investment Professionals*. Washington, DC: ETA.
- Van der Linden, W. J. & Hambleton, R. K. (Eds.) (1997). *Handbook of Modern Item Response Theory*. New York, NY: Springer.
- Van de Rijt, B., Godfrey, R., Aubrey, C., van Luit, J. E. H., Ghesquiere, P., Hasemann, K., Tancing, S., Kavkler, M. & Tzouriado, M. (2003). The Development of Early Numeracy in Europe. *Journal of Early Childhood Research*, 1 (2), 155–180.
- Van de Rijt, B., van Luit, J. E. H. & Pennings, A. H. (1999). The Construction of the Utrecht Early Mathematical Competence Scales. *Educational and Psychological Measurement*, 59, 289–309.
- Verhoeven, L., Schnotz, W. & Paas, F. (2009). Guest Editorial: Cognitive Load in Interactive Knowledge Construction. *Learning and Instruction*, 19 (5), 369–375.
- Viteles, M. S. (1922). *Job Specifications and Diagnostic Tests of Job Competency Designed for the Auditing Division of a Railway Company* (a Psychological Study in Industrial Guidance), IX of Experimental Studies in Psychology and Pedagogy. Philadelphia: University Press.
- Voss, J. F., Blais, J., Means, M. L., Greene, T. R. & Ahwesh, E. (1986). Informal Reasoning and Subject Matter Knowledge in the Solving of Economics Problems by Naive and Novice Individuals. *Cognition and Instruction*, 3, 269–302.
- Voss, J. F. & Post, T. A. (1988). On the Solving of Ill-Structured Problems. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The Nature of Expertise* (pp. 261–285). Hillsdale, NJ: Lawrence Erlbaum.
- Walter, O. (2005). *Kompetenzmessung in den PISA-Studien. Simulation zur Schätzung von Verteilungsparametern und Reliabilitäten*. Lengerich: Pabst.
- Wang, S. (2001). Precision of Warm's Weighted Likelihood Estimates for a Polytomous Model in Computerized Adaptive Testing. *Applied Psychological Measurement*, 25 (4), 317–331.
- Wang, W., Wilson, M. & Adams, R. J. (1997). Rasch Models for Multidimensionality Between Items and Within Items. In M. Wilson & G. Engelhard (Eds.), *Objective Measurement: Theory into Practice* (Vol. 4, pp. 139–155). Norwood, NJ: Ablex Publishing.
- Warner Weil, S., Wildermeersch, D., Jansen, T. & Percy-Smith, B. (2005). *Unemployed Youth and Social Exclusion in Europe. Learning for Inclusion?* Aldershot: Asgate.
- Weber, S. & Achtenhagen, F. (2009). Forschungs- und evidenzbasierte Lehrerbildung. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messungen* (S. 477–488). Weinheim: Beltz.

- Weinert, F. E. (1999). *Konzepte der Kompetenz*. Paris OECD.
- Weinert, F. E. (2001a). Concept of Competence: A Conceptual Clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies*. Seattle: Hogrefe und Huber.
- Weinert, F. E. (2001b). Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 17–32). Weinheim: Beltz.
- Weinert, F. E. (Hrsg.) (2001c). *Leistungsmessung in Schulen*. Weinheim: Beltz.
- Westera, W. (2001). Competences in Education: A Confusion of Tonques. *Journal of Curriculum Studies*, 33 (1), 75–88.
- White, B. & Frederiksen, J. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction*, 16 (1), 3–118.
- Whitley, S. E. (1980). Multicomponent for Latent Trait Models for Ability Tests. *Psychometrika*, 45, 479–494.
- Wigdor, A. K. & Green, B. F. (Eds.) (1991). *Performance Assessment for the Workplace*. National Research Council. Washington, DC: National Academy Press.
- Wilson, M. (Ed.) (2004a). *Towards Coherences Between Classroom Assessment and Accountability. 103rd Yearbook of the National Society for the Study of Education*. Chicago, IL: The University of Chicago Press.
- Wilson, M. (2004b). Assessment, Accountability and the Classroom: A Community of Judgement. In M. Wilson (Ed.), *Towards Coherences Between Classroom Assessment and Accountability. 103rd Yearbook of the National Society for the Study of Education* (pp. 1–19). Chicago, IL: The University of Chicago Press.
- Wilson, M. (2004c). On Choosing a Model for Measuring. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch-Measurement* (pp. 11–42). Maple Grove, MN: JAM Press.
- Wilson, M. (2005). *Constructing Measures: An Item-response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum.
- Wilson, M. (2008). Cognitive Diagnosis Using Item Response Models, *Journal of Psychology*, 216 (2), 74–88.
- Wilson, M. & Draney, K. (2004). Some Links between Large-Scale and Classroom Assessments: The Case of the BEAR Assessment System. In M. Wilson (Ed.), *Towards Coherences Between Classroom Assessment and Accountability. 103rd Yearbook of the National Society for the Study of Education* (pp. 132–154). Chicago, IL: The University of Chicago Press.
- Wilson, M. & Sloane, K. (2000). From Principles to Practice: An Embedded Assessment System. *Applied Measurement in Education*, 13 (2), 181–208.
- Winterton, J., Delamare-Le Deist, F. & Stringfellow, E. (2006). *Typology of Knowledge, Skills and Competences: Clarification of the Concept and Prototype*. Luxembourg: Office of the Official Publications of the European Communities.
- Winther, E. (2006). *Motivation in Lernprozessen. Konzepte in der Unterrichtspraxis von Wirtschaftsgymnasien*. Wiesbaden: DUV.
- Winther, E. (2007). Selbstreguliertes Lernen: Theorien, empirische Befunde und fachdidaktische Implikationen. *Kölner Zeitschrift für Wirtschaft und Pädagogik*, 22 (42), 35–62.
- Winther, E. & Achtenhagen, F. (2007). *Kompetenzmessung im kaufmännisch-verwaltenden Bereich: Hintergründe der Modellierung und Validierung für den Ausbildungsberuf Bankkauf-frau/-mann*. Workshop VET-LSA, 15. Oktober, Göttingen.
- Winther, E. & Achtenhagen, F. (2008a). Personale traits und selbst-regulative states zur Beschreibung von Unterrichtsprozessen. *Unterrichtswissenschaft. Zeitschrift für Lernforschung*, 36 (3), 255–280.
- Winther, E. & Achtenhagen, F. (2008b). Kompetenzstrukturmodell für die kaufmännische Bildung. Adaptierbare Forschungslinien und theoretische Ausgestaltung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104, 511–538.
- Winther, E. & Achtenhagen, F. (2008c). ‚Konzeptuale Kompetenz‘ und ‚Selbstregulation‘ als Grundlagen einer berufsbezogenen Kompetenzforschung. In D. Münk, P. Gonon, K. Breuer & T. Deißinger (Hrsg.), *Modernisierung der Berufsbildung. Neue Forschungserträge und Perspektiven der Berufs- und Wirtschaftspädagogik* (S. 100–110). Opladen: Barbara Budrich.

- Winther, E. & Achtenhagen, F. (2009a). Measurement of Vocational Competencies – a Contribution to an International Large-Scale-Assessment on Vocational Education and Training. *Empirical Research in Vocational Education and Training*, 1, 8–106.
- Winther, E. & Achtenhagen, F. (2009b). Skalen und Stufen kaufmännischer Kompetenz. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105 (4), 521–556.
- Winther, E. & Achtenhagen, F. (2010). Berufsfachliche Kompetenz: Messinstrumente und empirische Befunde zur Mehrdimensionalität beruflicher Handlungskompetenz. *Berufsbildung in Wissenschaft und Praxis*, 1/2010, 18–21.
- Witt, R. (2006). Kompetenzstufenmodelle zur Messung ökonomischer Bildung. In G. Minnameier & E. Wuttke (Hrsg.), *Berufs- und wirtschaftspädagogische Grundlagenforschung. Lehr-Lern-Prozesse und Kompetenzdiagnostik* (S. 407–420). Frankfurt a. M.: Peter Lang.
- Wolf, D. P. (1989). Portfolio Assessment: Sampling Student Work. *Educational Leadership*, 46, 35–39.
- Wolter, S. C. (2007). *Kosten und Nutzen der Berufsausbildung aus Sicht der Betriebe*. Pressegespräch BBT. Bern.
- Wright, B. D. & Douglas, G. A. (1977). Conditional versus Unconditional Procedures for Sample-free Item Analysis. *Educational and Psychological Measurement*, 37, 47–60.
- Wright, B. D. & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago, IL: MESA Press.
- Wright, B. D. & Stone, M. H. (1979). *Best Test Design*. Chicago, IL: MESA Press.
- Wu, M. (2004). Student performance in problem solving compared with performance in mathematics, reading and science. In OECD (Ed.), *Problem solving for tomorrow's world. First measures of cross-curricular competencies from PISA 2003* (pp. 49–58). Paris: OECD.
- Wu, M. (2005). The Role of Plausible Values in Large-Scale Surveys. *Studies in Educational Evaluation*, 31 (2-3), 114–128.
- Wu, M. L., Adams, R. J., Wilson, M. & Haldane, S. A. (2007). *ACER ConQuest. Version 2.0. Generalised Item Response Software*. Camberwell: ACER Press.
- Wuttke, E. & Wolf, K. D. (2007). Developing an Instrument for Identifying a Person's Ability to Solve Problems – Results of a Pilot Study. *European Journal of Vocational Training*, 41, 84–102.
- Yuan, K., Steedle, J., Shavelson, R., Alonzo, A. & Oppezzo, M. (2006). Working Memory, Fluid Intelligence, and Science Learning. *Educational Research Review*, 1, 83–98.
- Zlatkin-Troitschanskaia, O., Beck, K., Sembill, D., Nickolaus, R. & Mulder, R. (Hrsg.) (2009). *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung*. Weinheim: Beltz.

Wie sind berufliche Kompetenzen messbar und welche Kompetenzmodelle erweisen sich als plausibel für die Beschreibung beruflicher Lern- und Arbeitsprozesse? Wie können Kompetenzen schon in der Ausbildung sichtbar gemacht werden, damit unter anderem berufliche Abschlüsse in Europa miteinander verglichen werden können?

Der Band untersucht die Struktur und Reichweite von Kompetenzen in kaufmännisch-beruflichen Anforderungssituationen. Die Autorin entwickelt Instrumente zur Messung beruflicher Kompetenzen, die eine systematische Erfassung von beruflichen Handlungen und Argumentationsmustern möglich machen. Grundlage hierfür ist ein kaufmännisches Domänenmodell, das sowohl für die Entwicklung von Instrumenten der Kompetenzmessung als auch für die Gestaltung von Ausbildungsprogrammen nutzbar gemacht werden kann. Nach einer Einführung in die probabilistische Testtheorie stellt sie psychometrische Modelle und Verfahren der beruflichen Kompetenzmessung für einen vollzeitschulischen sowie für einen dualen Bildungsgang vor.