

## **Zur Erfassung und Vermittlung berufsbezogener Kompetenzen im teilqualifizierenden Bildungsgang „Wirtschaft und Verwaltung“ an Hamburger Berufsfachschulen**

---

### **1 Struktur des Beitrages**

Im folgenden Beitrag werden Möglichkeiten zur Erfassung und Skalierung von Kompetenzdimensionen exemplarisch anhand eines wirtschaftsberuflichen Tests diskutiert, der bei den Jugendlichen teilqualifizierender Bildungsgänge an Hamburger Berufsfachschulen im Bereich Wirtschaft und Verwaltung eingesetzt wurde. Im zweiten Teil des Beitrages wird mehr-ebenenanalytisch geprüft, ob die so geschätzten individuellen Kompetenzen in unterschiedlichen Lerngruppen differenziellen Vermittlungsprozessen unterlagen. Die in diesem Zusammenhang dargestellten Resultate stellen Zwischenschritte aus gegenwärtig laufenden Analysen dar, die noch eine weitere Ausdifferenzierung erfahren werden.

### **2 Kontext und Anlage der Studie**

#### **2.1 Teilqualifizierende Bildungsgänge an Hamburger Berufsfachschulen**

In die teilqualifizierenden Bildungsgänge der Hamburger Berufsfachschulen münden Jugendliche ein, die aus unterschiedlichen Gründen beim Übergang in eine berufliche Ausbildung scheiterten. Überwiegend handelt es sich um Schülerinnen und Schüler, die aufgrund individueller Wissens- und Kompetenzlücken oder wegen ihrer Herkunft am Ausbildungsstellenmarkt benachteiligt sind. So hat beispielsweise nur etwa die Hälfte der Jugendlichen bei Eintritt in den teilqualifizierenden Bildungsgang in Mathematik das Leistungsniveau von Hamburger Hauptschülern zu Beginn von Klassenstufe 9 erreicht und nur etwa 18 Prozent das der Hamburger Realschüler gleicher Klassenstufe. Auch die Lesekompetenz weist markante Leistungsrückstände auf. Ein Großteil der hier untersuchten Schülerinnen und Schüler kommt aus bildungsferneren Elternhäusern; der Anteil von Jugendlichen mit Migrationshintergrund liegt mit rund 50 Prozent deutlich über dem Durchschnitt an Hamburger Schulen.

Darüber hinaus sind in diesen Bildungsgängen durchaus auch Jugendliche zu finden, die aufgrund ihrer Persönlichkeitsmerkmale und Lernvoraussetzungen eine Ausbildung erfolgreich absolvieren könnten, deren Berufswahlprozesse jedoch noch nicht abgeschlossen sind. Insgesamt betrachtet zeigen die Befunde zu den Lernausgangslagen ein zwar heterogenes, aber dennoch vornehmlich ungünstiges Leistungsprofil in Bezug auf jene zentralen und grundlegenden Kompetenzen, die für die individuellen Lern- und Lebenschancen als bedeutsam gelten (dazu PRENZEL, DRECHSEL, CARSTENSEN & RAMM 2004, 17ff.).

Die Ausbildung in teilqualifizierenden Bildungsgängen an Berufsfachschulen zählt zu den indirekten Übergangswegen von der Schule in eine Berufsausbildung; sie ist dem sog. ‘Chancenverbesserungssystem’ (BAETHGE, BUSS & LANFER 2003, 44) zuzurechnen. Die teilqualifizierende Berufsfachschule in Hamburg hat entsprechend den Rahmenvereinbarungen über die Berufsfachschulen der Kultusministerkonferenz einen doppelten Bildungsauftrag zu erfüllen: zum einen soll den Schülerinnen und Schülern die Chance zur Erweiterung ihrer bisher erworbenen allgemeinen Bildung eingeräumt werden, und zum anderen werden berufsbezogene Inhalte aus einem oder mehreren anerkannten Ausbildungsberufen eines Berufsfeldes vermittelt. Der Besuch der teilqualifizierenden Berufsfachschule wird nicht auf eine spätere Ausbildungszeit angerechnet. Mit der erfolgreichen Beendigung des zweijährigen Programms sind die Jugendlichen im Besitz eines in seinen Berechtigungen der mittleren Reife gleichgestellten Abschlusses.

## **2.2 Zur Anlage der Studie**

Mit den beiden Studien „Untersuchung der Lernstände, Motivation und Einstellungen am Beginn der beruflichen Ausbildung“ (ULME I; vgl. LEHMANN, IVANOV, HUNGER & GÄNSFUSS 2005) und der Studie „Untersuchung der Lernstände, Motivation und Einstellungen am Ende der teilqualifizierenden Bildungsgänge an Berufsfachschulen“ (ULME II) liegen sowohl Längsschnittdaten für die Entwicklung der allgemeinen Fachleistungen in den Bereichen Deutsch-Leseverständnis, Mathematik und Englisch als auch Querschnittdaten für die inzwischen erworbenen berufsbezogenen Kompetenzen in insgesamt sechs Berufsfeldern vor. Zum einen geht es also um die Feststellung der Lernentwicklung in den genannten Fähigkeitsbereichen und um empirisch abgesicherte Informationen, in welchem Grade die dort erlangten und dem Realschulabschluss gleichgestellten Berechtigungen auch dem Leistungsprofil des mittleren Abschlusses entsprechen. Zum anderen sollen berufsbezogene Kompetenzen erfasst und auf ihre Struktur hin analysiert werden. Insbesondere diesem zuletzt genannten Ziel dient die hier vorgelegte Arbeit.

Insgesamt wurden 3.038 Schülerinnen und Schüler, die im Schuljahr 2002/03 die zweijährige Bildungsmaßnahme begonnen hatten, längsschnittlich untersucht. Es handelt sich dabei um eine Vollerhebung dieses spezifischen Angebots der Hamburger Berufsfachschulen. Gegen Ende des Schuljahrs 2003/04 konnten noch 1.762 Jugendliche, die die Maßnahme vollständig durchlaufen hatten, in die Abschlusserhebungen einbezogen werden. Die übrigen Jugendlichen hatten den Bildungsgang offenbar vorzeitig abgebrochen. Das heterogene Leistungsspektrum dieser Gruppe zu Beginn des Bildungsgangs legt den Schluss nahe, dass der Abbruch aus durchaus sehr unterschiedlichen Gründen erfolgt sein wird.

Das Berufsfeld Wirtschaft und Verwaltung stellt mit 1.517 Jugendlichen die stärkste Gruppe dar; rund die Hälfte der Jugendlichen, die im Schuljahr 2002/03 in einen solchen teilqualifizierenden Bildungsgang eintraten, entfielen auf dieses Segment. Insgesamt konnten 886 Jugendliche gegen Ende der Ausbildung an den Handelsschulen noch erreicht werden; am wirtschaftsberuflichen Test nahmen davon 792 Jugendliche teil.

### **3 Dimensionen und Skalierung des wirtschaftsberuflichen Tests**

#### **3.1 Zum Aufbau des Tests**

Der wirtschaftsberufliche Fachleistungstest wurde unter Federführung des Instituts für Berufs- und Wirtschaftspädagogik an der Universität Hamburg entwickelt und bezieht sich auf die curricularen Schwerpunkte der kaufmännischen Fachgebiete des Bildungsgangs. Von den 21 Testaufgaben setzen sich 17 Aufgaben aus jeweils vier, zwei weitere Aufgabe aus fünf bzw. sechs Wahr-Falsch-Paaren zusammen. Zwei Aufgaben enthalten neben zu beurteilenden Aussagen auch Aufgaben zur Berechnung ökonomischer Kennzahlen im offenen Antwortformat. Zwei weitere Aufgaben sind im Modus von Zuordnungsaufgaben konzipiert. Die jeweils zu einer Aufgabe zusammengestellten Aussagen beziehen sich auf ein wirtschaftsberufliches Sachgebiet bzw. auf einen Themenschwerpunkt. Insgesamt besteht der Test aus 95 Einzel-Items. Die kognitiven Anforderungen des Tests umfassen das Wiedererkennen erlernter ökonomischer Begriffe, Kategorien, Sachverhalte und Wechselwirkungen sowie das Anwenden erworbenen Wissens auf neue Situationen. Weiterhin wird bei den Testaufgaben zwischen Faktenwissen, konzeptionellem und prozeduralem Wissen unterschieden (vgl. auch den Beitrag von BRAND, HOFMEISTER & TRAMM in dieser Ausgabe). Das Klassifikationsraster bezieht folglich die Testaufgaben auf (fach)didaktische Kriterien und stellt einen wichtigen Ausgangspunkt für die Konstruktion der Testaufgaben und die inhaltliche Beschreibung von Kompetenzen dar (vgl. auch die im Rahmen von PISA 2003 entwickelten Kompetenzcluster, die die Modellierung des Anspruchsniveaus der Aufgaben auf einem theoretischen Niveau unterstützen sollen in BLUM, NEUBRAND, EHMKE, SENKBEIL, JORDAN, ULFIG & CASTENSEN 2004, 50).

#### **3.2 Zur Dimensionierung und Skalierung des Tests**

Sollen die Befunde aus Leistungstests im Hinblick auf curriculare, didaktische und psychologische Fragestellungen interpretiert werden, so ist in jedem Fall eine inhaltliche Interpretation der erworbenen berufsbezogenen Kompetenzen erforderlich. Dabei wird die jeweilige Kompetenz an einem „Kriterium wohldefinierter Operationen gemessen, die ihrerseits bestimmte Lernziele repräsentieren können“ (KLIEME, BAUMERT, KÖLLER & BOS 2000, 114). Zudem setzt dieses Vorgehen Skalierungsmethoden voraus, die Fähigkeitsschätzungen nach Maßgabe fachlich bestimmter Anforderungsstufen ermöglichen (vgl. dazu BAUMERT, BOS & LEHMANN 2000, 27).

Dementsprechend wurde hier zunächst geprüft, inwiefern der wirtschaftsberufliche Fachleistungstest unter Nutzung von Methoden der Item-Response-Theory (vgl. FISCHER & MOLENAAR 1995, ROST 2004) skalierbar ist. Dabei werden bekanntlich die Fähigkeiten der Probanden unter Berücksichtigung der als populationsunabhängig geltenden Schwierigkeit der gelösten Aufgaben ermittelt, während gleichzeitig der Schwierigkeitsgrad der Testaufgaben auf demselben Maßstab bestimmt wird. Aufgabenschwierigkeiten und Schülerfähigkeiten sind über die Kategorie der Lösungswahrscheinlichkeit miteinander verknüpft. Die Schwierigkeit eines Items ist durch die Lage des Wendepunktes der so genannten Item-Cha-

rakteristik-Kurven definiert, an dem die von der Personenfähigkeit abhängende Lösungswahrscheinlichkeit 0,5 beträgt (vgl. hierzu ROST 2004, 115ff.). Durch eine einfache Transformation der Personenfähigkeiten wurde in dem hier diskutierten Test erreicht, dass Aufgabenschwierigkeit und Personenfähigkeit einer Lösungswahrscheinlichkeit von 0,65 entsprechen. Von der Beherrschung oder Bewältigung eines Anforderungsniveaus wird hier (ähnlich wie beispielsweise in TIMSS und in der PISA-Untersuchung) also erst dann gesprochen, wenn 'auf lange Sicht' etwa zwei von drei entsprechend schwieriger Aufgaben gelöst würden.

In einem ersten Schritt erfolgte also die Skalierung des Tests auf Basis des eben skizzierten einparametrischen dichotomen Raschmodells, das ein Latent-Trait-Modell darstellt, bei dem die manifeste Reaktion auf jedes Item in der geschilderten Weise auf eine spezifische Personenfähigkeit und eine ebenso bestimmte Itemschwierigkeit zurückgeführt wird (vgl. dazu BAUMERT, KÖLLER, LEHRKE & BROCKMANN 2000, 60ff.). Die Skalierbarkeit nach dem Rasch-Modell impliziert paarweise Unabhängigkeit der einzelnen Aufgaben, ferner die Abwesenheit (bedeutsamer) Rate-Tendenzen und das Fehlen von Effekten, nach denen einzelne Probandengruppen und/oder einzelne Items durch spezielle Antwortmuster charakterisiert sind. Demnach muss eine eindeutige und in allen Teilpopulationen gültige Rangfolge der Aufgaben in Bezug auf die Schwierigkeit existieren, was in der Regel an unterschiedliche Anforderungsmerkmale der Aufgaben gebunden ist (vgl. dazu KLIEME, BAUMERT, KÖLLER & BOS 2000, 111ff.).

Unter Verwendung des Computerprogramms ConQuest wurden über das einparametrische logistische Modell zwei Varianten geprüft und miteinander verglichen: ein eindimensionaler und ein zweidimensionaler Ansatz. Dabei ergab sich, dass der eindimensionale Ansatz zufriedenstellende Eigenschaften aufweist, insbesondere eine hohe interne Konsistenz ( $\alpha = 0,83$ ), während der zweidimensionale Ansatz einige Anhaltspunkte für die Existenz einer zusätzlichen Dimension lieferte. Diese Aufspaltung in zwei Dimensionen kann als Unterscheidung von spezifisch berufsbezogenen Kompetenzen und eher allgemeinem ökonomischem Wissen interpretiert werden. Bei Annahme dieser Lösung betrüge die latente Korrelation zwischen den beiden Dimensionen  $r = 0,62$ . Allerdings ist diese zweite Dimension mit zu wenigen Items besetzt, sodass ihre interne Konsistenz  $\alpha = 0,61$  nicht überschreitet. Deshalb erscheint es einstweilen nicht angezeigt, mit dieser Variante zu arbeiten. Es wird jedoch als aussichtsreich erachtet, durch Entwicklung zusätzlicher Items die Messung jener zweiten Dimension zu verbessern, so dass in einer künftigen Untersuchung umfassendere Kompetenzprofile erfasst werden können. Bei der Bestimmung von übergreifenden sowie spezifischen Kompetenzen in verschiedenen kaufmännischen Berufen, die in den Abschlussmessungen an den Berufsschulen im Jahre 2005 in Hamburg bereits eingeleitet sind (ULME III), dürfte eine Unterscheidung mehrerer Dimensionen nahe liegen.

## 4 Anforderungsmerkmale, Schwierigkeit von Aufgaben und Fähigkeitsniveaus

### 4.1 Testanforderungen und Aufgabenschwierigkeit

Von den 95 verfügbaren Items erfüllen 80 Items die Gütekriterien, die die Grundlage für eine Schätzung von Personen- und Itemparameter bilden, namentlich die der homogenen und ausreichenden Diskriminanz. Die Schwierigkeitskennwerte dieser 80 Items reichen von 54 bis 201 auf einer Skala, die über eine lineare Transformation der ursprünglichen Personenkennwerte (der so genannten 'Logits') so festgelegt wurde, dass der Mittelwert für die hier untersuchte Gruppe 100 und die Standardabweichung 25 Punkte beträgt. Die mittlere Itemschwierigkeit beträgt 114 Punkte auf derselben Skala (vgl. Abbildung 1).

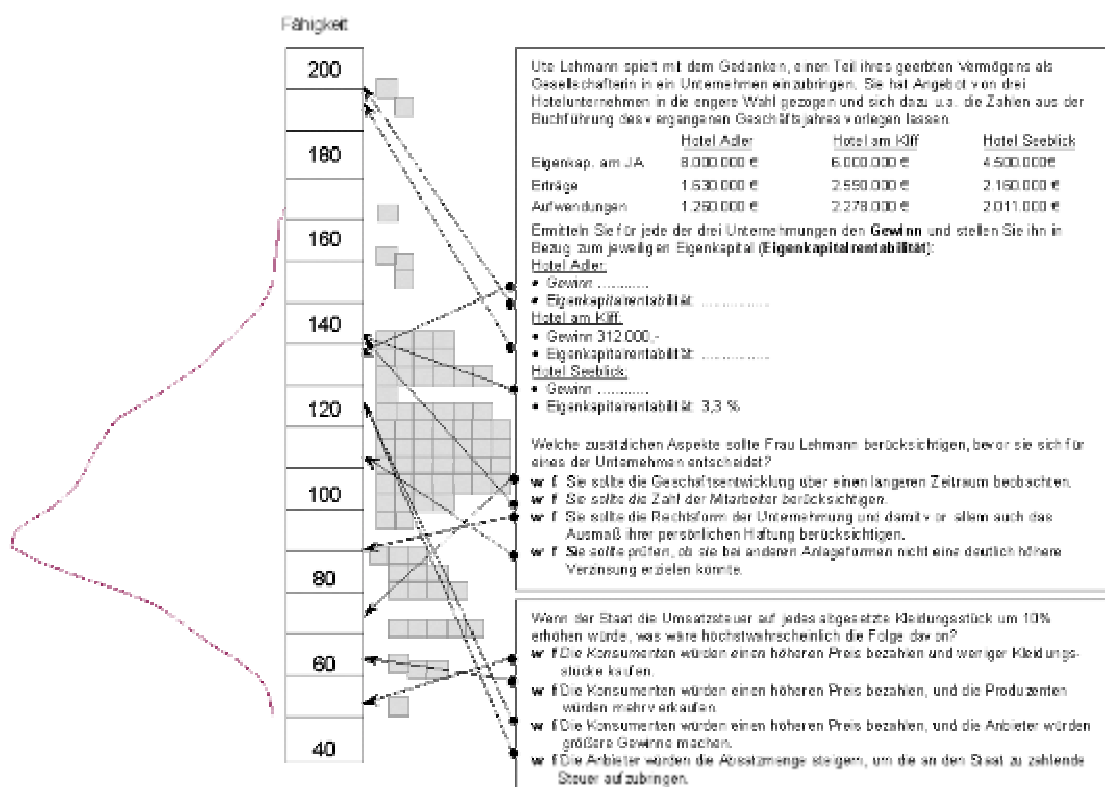


Abb. 1: Aufgabenschwierigkeiten und Verteilung der Personenfähigkeiten im Hamburger Fachleistungstests für teilqualifizierende Bildungsgänge an Handelsschulen

In Abbildung 1 sind die Verteilungen der Personenfähigkeiten- und Schwierigkeitsparameter grafisch dargestellt. Unmittelbar rechts der Skala sind die 80 Test-Items mit ihrem jeweiligen Schwierigkeitsparameter in Form einzelner Kästchen repräsentiert und so dem zugehörigen Skalenwert zugeordnet. Am rechten Rand sind exemplarisch zwei Aufgaben aus dem Test aufgeführt, für deren einzelne Items die Schwierigkeitskennwerte ausgewiesen sind. Die

Funktion links der Skala entspricht der Verteilung der Testleistungen in der Weise, dass aus den Flächeninhalten zwischen Kurve und Skalenbalken auf die Anteile der Zielgruppe geschlossen werden kann, die ein bestimmtes Kompetenzniveau erreicht oder überschritten – oder eben nicht erreicht – haben.

Aus der grafischen Zuordnung von Testaufgaben und Aufgabenschwierigkeit wird ersichtlich, dass sich selbst innerhalb eines Themenbereichs und bei identischer taxonomischer Zuordnung sehr unterschiedliche Anforderungsgehalte der einzelnen Aufgaben ergeben. Beispielsweise wurden in der Testkonzeption alle vier Teilaufgaben zur rechnerischen Ermittlung des Gewinns und der Eigenkapitalrentabilität (oberes Beispiel) ursprünglich als konzeptionelles Wissen eingestuft, das es zu reproduzieren gelte. Die empirisch geschätzten Schwierigkeitskennwerte machen jedoch deutlich, dass den Jugendlichen die Berechnung des Gewinns offensichtlich sehr viel leichter fällt als die Bestimmung der Eigenkapitalrentabilität. Die vier Wahr-Falsch-Paare dieser Aufgabe wurden a priori als konzeptionelles Wissen klassifiziert, das anzuwenden bzw. auf neue Situationen zu übertragen sei. Auch hier zeigt sich, dass die vier Items trotz identischer Klassifikation evident unterschiedlich hohe Anforderungen an die kognitiven Verarbeitungsprozesse stellen.

Ein ähnliches Bild zeichnet sich für die zweite hier ausgewählte Testaufgabe ab, bei der alle vier Wahr-Falsch-Paare ursprünglich als konzeptionelles Wissen klassifiziert worden waren, das es auf neue Situationen zu übertragen gilt. Hier variieren die empirisch geschätzten Schwierigkeitsparameter zwischen 54 und 124 Skalenpunkten, also sehr beträchtlich.

Obwohl sich bereits in diesem Stadium der Analysen abzeichnete, dass die *a priori*-Klassifikationen die Aufgabenanforderungen vermutlich nicht hinreichend erklären würden, wurde quantitativ überprüft, inwiefern mit den Aufgabenmerkmalen, die der ursprünglichen Klassifikation zu Grunde gelegt worden waren, auch eine Vorhersage der Aufgabenschwierigkeiten möglich ist.

Wie oben dargestellt, war über die Rasch-Skalierung für jedes Test-Item  $I_j$  ein Schwierigkeitsparameter  $\sigma_j$  ermittelt worden, der die Verortung der Aufgabe auf der Kompetenzskala festlegt (vgl. Abbildung 1). Regressionsanalytisch wurde bestimmt, in welchem Grade die a priori-Klassifikation der Testaufgaben als Prädiktor für die ermittelten Aufgabenschwierigkeiten geeignet ist. Hierfür wurden die 80 Test-Items im Datensatz als ‘Fälle’ behandelt. Die fünf auftretenden Kombinationen aus Wissensart und kognitiver Leistung – ‘Faktenwissen reproduzieren’, ‘konzeptionelles Wissen reproduzieren’ sowie ‘Faktenwissen anwenden’, ‘konzeptionelles Wissen anwenden’ und ‘prozedurales Wissen anwenden’ – wurden über eine Dummy-Codierung als unabhängige Variablen und die Aufgabenschwierigkeit als abhängige Variable definiert. In der nachfolgenden Tabelle 1 sind die unstandardisierten und standardisierten Regressionskoeffizienten sowie die Irrtumswahrscheinlichkeit für den jeweiligen  $\alpha$ -Fehler ausgewiesen.

Tabelle 1: **Ergebnisse der Regressionsanalyse zur Vorhersage der Aufgabenschwierigkeiten im wirtschaftsberuflichen Test (N<sub>k</sub>=80 Items)**

Modell	Nicht standardisierte Koeffizienten B	Standardisierte Koeffizienten Beta	Signifikanz
Konstante	121,101		0,000
Konzeptionelles Wissen reproduzieren	-4,962	-0,040	0,718
Faktenwissen anwenden	-33,168	-0,230	0,035
Konzeptionelles Wissen anwenden	-13,137	-0,233	0,071
Prozedurales Wissen anwenden	29,490	0,323	0,007

*Referenzkategorie:* Faktenwissen reproduzieren; *Abhängige Variable:* Rasch-skalierte Aufgabenschwierigkeit im wirtschaftsberuflichen Test;  $R^2=.26$

Die Analyse zeigt, dass die mit einer niedrigeren kognitiven Leistung verbundenen Aufgaben, bei denen ein einfaches Reproduzieren von ökonomischen Begriffen, Sachverhalten und Strukturen gefordert ist, in der gegebenen Item-Auswahl im Durchschnitt offenbar höhere Anforderungen an die Jugendlichen stellen als jene Aufgaben, die beispielsweise Vergleiche, Begründungszusammenhänge und die Übertragung erlernter Schemata auf neue Situationen erfordern. Da im Test nur vereinzelt innerhalb bestimmter thematischer Blöcke unterschiedliche Kombinationen aus Wissensart und kognitiver Leistung auftreten, kann nicht geprüft werden, ob es sich hier um ein generelles Problem der Klassifikation handelt, oder ob die intensivere Beschäftigung mit bestimmten Themen, bei denen im Test der Anwendungsbezug abgefordert wird, im Vergleich zu möglicherweise weniger vertrauten Inhalten im Bereich des Reproduzierens zu diesem Ergebnis geführt hat. Diejenigen Aufgaben, für deren Lösung Heuristiken zu entwickeln und einzusetzen waren und die deshalb der Anwendung prozeduralen Wissens zugeordnet wurden, stellen jedoch in *diesem* Test ausweislich der Regressionskoeffizienten erheblich höhere kognitive Anforderungen dar, wie erwartet worden war und wie es sich auch zufallskritisch bestätigen ließ.

Insgesamt bleibt jedoch der Zusammenhang zwischen der a priori-Klassifikation und der empirisch gefundenen Aufgabenschwierigkeit vorläufig unbefriedigend ( $R^2=.26$ ). Für eine überzeugende Definition von Kompetenzniveaus wird man – in Anlehnung an den Kompetenzbegriff von WEINERT (2001, 27f.) – nach aussagekräftigeren Prädiktoren suchen müssen.

#### 4.2 Überlegungen zur Identifikation schwierigkeitsbestimmender Aufgabenmerkmale und zur Bildung von Kompetenzniveaus

Ohne einer späteren Einschätzung bzw. Bestätigung durch Fachexperten vorgreifen zu wollen, lassen sich bereits an dieser Stelle einige schwierigkeitsbestimmende Aufgabenmerkmale identifizieren, mit denen im Rahmen einer vorläufigen Festlegung von Kompetenzniveaus gearbeitet werden kann. So erfordern beispielsweise Aufgaben mit niedrigen Schwierigkeits-

parametern zumeist alltagsnahe Überlegungen, für deren Lösung nicht unbedingt explizites ökonomisches Fachwissen aktiviert und eingesetzt werden muss.

Zu diesen Aufgabentypen gehören z.B. die ersten beiden Wahr-Falsch-Paare des Themenbereichs „Folgen einer Mehrwertsteuererhöhung auf Kleidungsstücke um 10 Prozent“:

- (1) „Die Konsumenten würden einen höheren Preis bezahlen und weniger Kleidungsstücke kaufen“,
- (2) „Die Konsumenten würden einen höheren Preis bezahlen, und die Produzenten würden mehr verkaufen“.

Im ersten Fall wäre der Begriff der Mehrwertsteuer in seiner mathematischen Struktur ins Bewusstsein zu rufen, um sodann die Folgen, die *Implikationen*, einer Erhöhung für die Konsumenten, insbesondere die Konsequenzen für deren Kaufverhalten abzuwägen; es geht also um die Beziehung zwischen vertrauten Variablen. Im zweiten Fall wären die Konsequenzen einer Steuererhöhung aus zwei verschiedenen Perspektiven zu betrachten, nämlich aus Konsumenten- und Produzentensicht. Dieser Perspektivenwechsel stellt sich als etwas anspruchsvoller im Vergleich zur ersten Aufgabe dar. Aus fehler- bzw. förderdiagnostischer Sicht wäre jedoch die Vermeidung solcher ‘*double-barreled questions*’ zu empfehlen, bei der die Gründe für eine falsche Antwort unklar bleiben. Dennoch gehört auch dieses zweite Item – trotz des erforderlichen Perspektivenwechsels – zu den leichteren Aufgaben.

Insgesamt betrachtet sind die Aufgaben im unteren Schwierigkeitsbereich anhand plausibler Überlegungen, grundlegender Erfahrungen in wirtschaftlichen Kontexten und alltagsnaher Schlussfolgerungen lösbar. Es müssen intern nur relativ einfache Handlungssituationen oder Ereignissequenzen modelliert werden, bei denen wenige und weitgehend vertraute Parameter sowie deren Wechselwirkung zu berücksichtigen sind.

Für Aufgaben im *mittleren Schwierigkeitsbereich* ist es hingegen erforderlich, ökonomische Begriffe und Zusammenhänge explizit zu berücksichtigen. Dabei handelt es sich um Aufgaben, die die Aktivierung eines oder mehrerer Begriffe und der dahinter stehenden ökonomischen Zusammenhänge verlangen, wie beispielsweise die Abwägung, ob der Effektivzins eine aussagekräftige Kennzahl für den Vergleich von Kreditangeboten darstellt. Ähnliche Schwierigkeiten weisen Aufgaben auf, die den Zusammenhang zwischen dem Strukturwandel in der Volkswirtschaft und der Arbeitslosigkeit oder die Wechselwirkung zwischen einer Vergrößerung der Lagerkapazität und der wirtschaftlichen Abhängigkeit von Lieferanten thematisieren. Es sind bei diesen Aufgaben für die korrekte Lösung in der Regel mehrere ökonomische Fachbegriffe im Arbeitsgedächtnis präsent zu halten, um deren wirtschaftliche Zusammenhänge modellieren und prüfen zu können.

Aufgaben, die für die getesteten Jugendlichen ein sehr *hohes Anspruchsniveau* markieren, erfordern die Verknüpfung von Verständnis- und Begründungswissen zu bestimmten ökonomischen Sachverhalten. Diese Elemente – sozusagen die ‘Bestimmungsstücke’ – sind anschließend zueinander in Beziehung zu setzen. Für die Lösung einiger Aufgaben müssen zusätzlich



Wissensbestandteile und Fähigkeiten aus anderen Fachgebieten eingesetzt werden. Darüber hinaus können schwierige Aufgaben auch dadurch charakterisiert sein, dass komplexe Handlungsabläufe und Schrittfolgen zu erarbeiten und miteinander in Verbindung zu bringen sind.

So gehören die Aufgaben zur Ermittlung der Eigenkapitalrentabilität oder zur Ermittlung der Umsatzsteuerzahllast zu den anspruchsvollsten Aufgaben. Dies mag zum einen am offenen Antwortformat liegen, vor allem aber vermutlich an der Anzahl durchzuführender Formaloperationen und zu leistender Verknüpfungen zwischen unterschiedlichen Inhaltsbereichen. Zu den eher anspruchsvolleren Aufgaben gehören auch jene, bei denen mehr als zwei ökonomische Aspekte miteinander verknüpft werden müssen, wie dies in einer Aufgabe über den Zusammenhang zwischen der Senkung von Lohnnebenkosten, Produktionskosten und Wettbewerbsfähigkeit der Fall ist. Um hier eine sinnvolle Entscheidung zum Wahrheitsgehalt der Aussagen treffen zu können, müssen präzise Vorstellungen von den wirtschaftlichen Begriffen und Begriffshierarchien vorhanden sein sowie ein zuverlässiges Verständnis der Wechselwirkungen zwischen mehreren ökonomischen Parametern, die im konkreten Beispiel auf unterschiedlichen Ebenen angesiedelt sind.

Will man nun in einem nächsten Schritt Kompetenzniveaus ermitteln und die dahinter stehenden Annahmen empirisch prüfen, so ist eine multiple Klassifikation der Testaufgaben erforderlich, in deren Rahmen Aspekte wie Vertrautheit der Situation und der zu berücksichtigenden Variablen, Komplexität der zu modellierenden Handlungssituation, Mehrschrittigkeit des Lösungsprozesses selbst, Abstraktheit der Informationen, Integration von Wissen und Fähigkeiten aus unterschiedlichen ökonomischen Inhaltsbereichen und/oder anderen Fachgebieten etc. berücksichtigt werden.

Prinzipiell sind für die Bestimmung von Kompetenzniveaus methodisch zwei verschiedene Ansätze geeignet: zum einen eine *post hoc* durchzuführende Analyse der Anforderungen der Aufgaben, und zum anderen eine Klassifikation von Aufgabenmerkmalen *a priori*. Diese beiden prädiktiven Elemente können unter Nutzung von linearen oder logistischen regressionsanalytischen Verfahren zur 'Vorhersage' der Aufgabenschwierigkeiten eingesetzt werden (vgl. dazu einen Überblick in HARTIG 2004). Die Klassifikation von Testaufgaben *a priori* bietet den Vorteil, eine inhaltliche Interpretation der Befunde zu den Fähigkeiten und Leistungen der Schüler vor dem Hintergrund fachdidaktisch begründeter Leistungsanforderungen und vorab formulierter Kompetenzmodelle zu ermöglichen (zu diesem Vorgehen vgl. BECK & KLIEME 2003). Da jedoch eine solche Definition von Leistungsanforderungen und Kompetenzniveaus *a priori* nicht immer gelingt, hat auch das in seinen Grundzügen von BEATON & ALLEN (1992) entwickelte und zur Klassifikation der TIMSS-Aufgaben genutzte Verfahren der *post hoc*-Analyse von Aufgabeninhalten nach wie vor seine Berechtigung. Bei diesem Vorgehen werden Ankerpunkte auf der Kompetenzskala gesetzt und abschließend diejenigen Aufgaben als 'Markieritems' betrachtet, die auch nach formalen Kriterien (etwa dem der relativen Lösungsquoten) zur Beschreibung der jeweils charakteristischen Anforderungen geeignet erscheinen (vgl. dazu den Beitrag von IVANOV & LEHMANN in dieser Ausgabe).

## 5 Erklärungsansätze zu den berufsbezogenen Fachleistungen

### 5.1 Methodisches Vorgehen

Werden nun die Testleistungen auf der Ebene der Schulklassen betrachtet, so zeichnen sich deutliche Unterschiede hinsichtlich der erreichten Mittelwerte, aber auch hinsichtlich der beobachteten Streuungen ab (vgl. Abbildung 2). Somit kann man die analytisch, aber auch pädagogisch bedeutsame Frage stellen, wodurch solche unterschiedlichen Lernstände am Ende der teilqualifizierenden Ausbildung bedingt sind.

Als Erklärung für einen günstigen Klassenmittelwert im beruflichen Fachleistungstest beispielsweise kommen einerseits günstige kumulierte Individualmerkmale in Frage – etwa ein hoher CFT-Wert bei vielen Jugendlichen in der Klasse bei konstantem Zusammenhang zwischen kognitivem Lernpotenzial und Testergebnis. Andererseits können aber Charakteristika der Lerngruppe – wiederum z. B. ein hoher durchschnittlicher CFT-Wert – die Zusammenhänge zwischen der abhängigen Variablen und den Prädiktoren selbst beeinflussen, ein Effekt, den man als *‘cross-level interaction’* bezeichnet. In einem solchen Falle also würde sich der positive Einfluss einer günstigen Ausprägung der allgemeinen kognitiven Fähigkeiten gleichsam doppelt bemerkbar machen: auf individueller Ebene als förderliche persönliche Lernvoraussetzung und zugleich auf der *‘aggregierten Ebene’* (hier: der Ebene der Schulklasse) als ein Faktor, der über die Individualeffekte hinaus etwa das Lernanspruchsniveau erhöht.

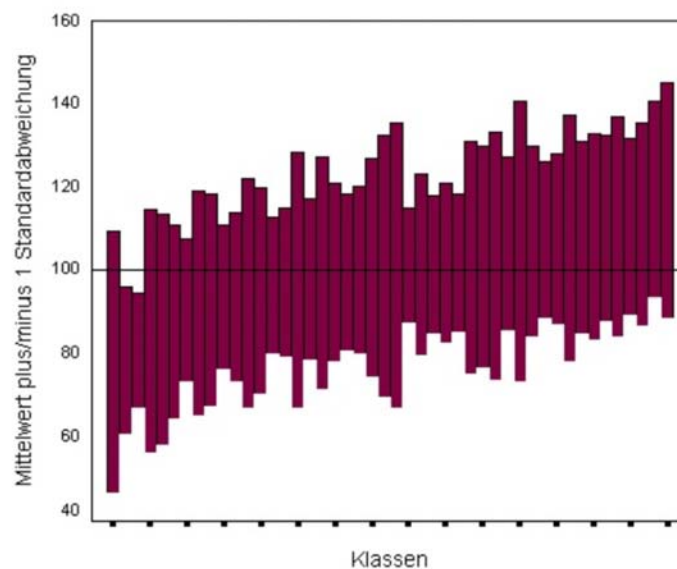


Abb. 2: Mittelwert im wirtschaftsberuflichen Test plus/minus eine Standardabweichung nach Klassen

In der üblichen Betrachtungsweise werden individuelle Schulleistungen in ihrer Abhängigkeit von bestimmten Prädiktoren betrachtet, etwa den kognitiven Lernvoraussetzungen oder bestimmten Fachkompetenzen. Im vorliegenden Fall wurde zur Aufklärung der Hintergründe

individueller Erfolge in der Bearbeitung des wirtschaftsberuflichen Fachleistungstests zum Untersuchungszeitpunkt 2 der nachstehende Ansatz gewählt:

$$WUVU2_i = \beta_0 + \beta_1 (CGU_i) + \beta_2 (MA2U2_i) + \beta_3 (MSPRA_i) + \beta_4 (FLIX1_i) + r_i$$

mit

WUVU2 = Fachleistungstest Wirtschaft und Verwaltung (Untersuchungszeitpunkt 2)

CGU = Testpunkte im nonverbalen Intelligenztest CFT 20 (Untersuchungszeitpunkt 1)

MA2U2 = Testpunkte im Fachleistungstest Mathematik II (Untersuchungszeitpunkt 2)

MSPRA = deutsche Muttersprache (dichotom kodiert)

FLIX1 = Index Allgemeine Fachleistung (Untersuchungszeitpunkt 1).

Der Betrag der standardisierten Regressionskoeffizienten  $\beta_j$  ist für das Weitere zunächst ohne Belang. Konkret wird hier also folgenden Fragen nachgegangen:

- Welchen Erklärungsbeitrag liefern Faktoren wie die allgemeinen Fachleistungen und die Fähigkeiten zum schlussfolgernden Denken sowie sozio-kulturelle Hintergrundmerkmale des Individuums bei Eintritt in die Bildungsmaßnahme zur Entwicklung berufsbezogener Kompetenzen?
- Wirken über diese individuellen Merkmale hinaus weitere Einflussgrößen auf der Ebene der Schulklassen auf die Entwicklung die Lernstände im berufsbezogenen Bereich?

Um die damit angesprochene hierarchische Struktur der Daten angemessen berücksichtigen zu können, wurde ein statistisches Verfahren gewählt, das als *'Mehrebenenanalyse'* bezeichnet wird und hier unter Nutzung des Programms HLM 5 (RAUDENBUSCH, BRY, CHEONG & CONGDON 2001) realisiert wurde.

## 5.2 Ergebnisse

Den Werten aus Tabelle 2 ist zu entnehmen, dass die Zusammenhänge zwischen Prädiktoren und der berufsbezogenen Fachleistung als abhängiger Variabler mit einer Ausnahme in allen Schulklassen hinreichend ähnlich sind: Ausweislich der Irrtumswahrscheinlichkeiten  $p$  kann hier von einem konstanten, daher als *'fixed'* bezeichneten Zusammenhang ausgegangen werden. Nur im Falle des Einflusses der mit dem Mathematiktest II erhobenen Kompetenzen kann angenommen werden, dass zwischen den Schulklassen systematische, u. U. durch die vorliegenden Daten erklärbare Variationen – so genannte *'random effects'* –, aufgetreten sind ( $p < 0,05$  s.).

Tabelle 2: **Regressionsgleichung der Ebene 1 (Individualebene)**

Variable / Koeffizient	SD	Varianzkomponente	Chi <sup>2</sup>	p
WUVU2 $\beta_0, u_0$	10,96	120,20	36,36	>0,500
CGU $\beta_1, u_1$	0,26	0,07	34,81	>0,500
MA2U2 $\beta_2, u_2$	0,24	0,06	64,43	0,019
MUSPRA $\beta_3, u_3$	2,52	6,37	53,63	0,152
FLIX1 $\beta_4, u_4$	0,27	0,07	49,20	0,272
Individualebene, R	20,60	424,56		

Obzwar bislang vergleichsweise wenig übergreifende Studien zum Einfluss von individuellen Persönlichkeitsmerkmalen auf die Entwicklung berufsbezogener Kompetenzen vorliegen, so werden doch mit den hier referierten Befunden weitgehend Erkenntnisse der allgemeinen Schulleistungsforschung bestätigt, in der Weise nämlich, dass jeder der berücksichtigten Prädiktoren einen eigenständigen Beitrag zur Erklärung der Fachleistung liefert:

- Je günstiger der allgemeine Fachleistungsindex (FLIX1) bei Eintritt in die Maßnahme, desto höhere Fachleistungen wurden im wirtschaftsberuflichen Test am Ende der teilqualifizierenden Ausbildung beobachtet. [Angemerkt sei, dass dieser allgemeine Fachleistungsindex über z-standardisierte Testwerte in den Untertests Mathematik I und II, Deutsch-Leseverständnis und Deutsch-Rechtschreibung, Englisch-Leseverständnis und Englisch-Wortergänzung (C-Test) gebildet wurde].
- Hohe Werte in der Fähigkeit zum schlussfolgernden Denken, gemessen mit dem sprachfreien CFT 20 (CGU), gehen mit günstigeren Lernständen im wirtschaftsberuflichen Test einher.
- Eine positive Entwicklung der Fachleistungen in Mathematik (MA2U2) während der teilqualifizierenden Bildungsmaßnahme, hier berücksichtigt über den Testwert in Mathematik II am *Ende* der Berufsfachschule, spiegelt sich in höheren wirtschaftsberuflichen Testleistungen wider.
- Eine deutsche Herkunftssprache (MSPRA) begünstigt höhere Fachleistungen im berufsbezogenen Kompetenztest. Ein Migrationshintergrund hingegen, erkennbar an einer nicht deutschen Herkunftssprache, zeigt sich auch im berufsbezogenen Test in der Regel an niedrigeren Punktwerten.

Nach den Ergebnissen, die auf 792 Schülern in 45 Klassen beruhen, ist also davon auszugehen, dass nur die Entwicklung der Fachleistung in Mathematik während der Ausbildung einen – je nach Klasse – differenziellen Einfluss auf die Ergebnisse im berufsbezogenen

Fachleistungstest ausübt. Bei allen übrigen Prädiktoren ist es nicht gerechtfertigt, von solchen differenziellen Einflüssen auszugehen.

In Mehrebenenanalysen können Merkmale der Aggregateinheiten sowohl aus der Aggregation individueller Merkmale gewonnen werden als auch direkt auf der Aggregatebene erfasst worden sein, wie etwa die Klassengröße oder zusammengefasste Daten aus Unterrichtsbeobachtungen (DITTON 1998, 104ff.). Im konkreten Fall wurden aggregierte Merkmale der Individualebene in das Modell einbezogen.

Tabelle 3: **Variablenstruktur nach Aggregationsebene**

<b>Aggregatebene</b>	<b>Individualebene</b>		<b>Klassenebene</b>
<b>Variable / Koeffizient</b>	<b>Koeffizient</b>	<b>Zentrierung</b>	<b>Koeffizient</b>
WUVU2	Konstante $\beta_0$	---	Konstante $\gamma_{00}$
			FLIX1_1 $\gamma_{01}$
CGU	Steigung $\beta_1$	unzentriert	Konstante $\gamma_{10}$
MA2U2	Steigung $\beta_2$	klassenweise	Konstante $\gamma_{20}$
			FLIX1_1 $\gamma_{21}$
MSPRA	Steigung $\beta_3$	unzentriert	Konstante $\gamma_{30}$
FLIX1	Steigung $\beta_4$	klassenweise	Konstante $\gamma_{40}$

In dieser Spezifikation sorgt die klassenweise ‘Zentrierung’, d. h. die Betrachtung der Abweichungen vom entsprechenden Klassenmittelwert, dafür, dass die relative Leistungsposition der bzw. des Einzelnen als Prädiktor fungiert und dass zugleich die berufsbezogene Fachleistung in den Klassen vergleichbar wird. Unterschiede im allgemeinen Fachleistungsdurchschnitt werden dann durch Zentrierung der Klassenmittelwerte um den Gesamtdurchschnitt angemessen berücksichtigt. Tabelle 4 enthält die Ergebnisse der entsprechenden Berechnungen.

Tabelle 4: **Regressionsgleichung der aggregierten Ebene 2 (Klassenebene)**

Effekt	Unstandardisierter Regressionskoeffizient	Standardfehler	t-Statistik	P
Konstante $\beta_0, \gamma_{00}$	87,89	3,97	22,16	0,000
Konstante $\beta_0$ , Beitrag des Durchschnitts Allgemeine Fachleistung (FLIX1_1) $\gamma_{01}$	1,03	0,35	2,91	0,006
Steigung CGU $\beta_1$ , Konstante $\gamma_{10}$	0,28	0,12	2,26	0,029
Steigung MA2U2 $\beta_2$ , Konstante $\gamma_{20}$	0,38	0,06	6,00	0,000
Steigung $\beta_2$ , Beitrag des Durchschnitts Allgemeine Fachleistung (FLIX1_1) $\gamma_{21}$	0,04	0,02	2,38	0,022
Steigung MSPRA $\beta_3$ , Konstante $\gamma_{30}$	8,80	1,56	5,64	0,000
Steigung FLIX1 $\beta_4$ , Konstante $\gamma_{40}$	0,53	0,10	5,10	0,000

Oben war bereits festgestellt worden, dass die Variablen „Fähigkeit zum schlussfolgernden Denken“ (CGU), „Deutsch als Herkunftssprache“ (MSPRA) und „Index Allgemeine Fachleistung“ (FLIX1) in guter Näherung in den einzelnen Schulklassen gleich aussagekräftige Prädiktoren für den Erfolg im berufsbezogenen Fachleistungstest sind: Die Steigungskoeffizienten der Regressionsgleichungen streuen jeweils eng um einen gemeinsamen Wert. Der Tabelle 4 lässt sich nun entnehmen, dass Jugendliche nichtdeutscher Herkunftssprache (MSPRA) unter sonst gleichen Bedingungen mit nahezu neun Punkten oder 0,35 SD hinter den anderen zurückstehen oder dass jedem zusätzlichen Punkt auf dem Index Allgemeine Fachleistung (FLIX1) etwas mehr als ein halber Punkt zusätzlich im Fachleistungstest entspricht. Auch die unabhängigen Erklärungsbeiträge der allgemeinen kognitiven Lernvoraussetzungen (CGU) und des abschließend erreichten Lernstands im Untertest Mathematik II (MA2U2) wurden bereits erwähnt. Prinzipiell ließen sich diese Koeffizienten auch so berechnen, dass sie wie standardisierte Regressionskoeffizienten allgemein vergleichbar sind.

Für den vorliegenden Zusammenhang sollen jedoch zwei Ergebnisse hervorgehoben werden, die ohne die mehrebenenanalytische Betrachtungsweise unentdeckt geblieben wären. Beide beziehen sich auf Effekte in der Wechselwirkung zwischen den beiden Ebenen, den *cross-level interaction effects*, die mit der durchschnittlichen allgemeinen Fachleistung verbunden sind: Je höher dieser durchschnittliche allgemeine Lernstand in der Klasse insgesamt, desto günstiger sind die Ergebnisse im berufsbezogenen Fachleistungstest ausgefallen, und zwar über die Kumulation der individuellen Effekte hinaus. Man wird darin am ehesten ein Indiz für ein anregendes und produktives Lernmilieu sehen, in dem auch besonders ausgeprägte berufsbezogene Lernfortschritte möglich geworden sind. In eine ähnliche Richtung verweist der Umstand, dass in solchen allgemein besser vorbereiteten und daher offenbar lernstärkeren

Klassen der Zusammenhang zwischen den mathematischen Grundqualifikationen (d. h. den Ergebnissen im Untertest Mathematik II) und den Erfolgen im berufsbezogenen Fachleistungstest *enger* ist als in anderen Lerngruppen. Hier gelingt es anscheinend besser, über die individuellen Beiträge hinaus das kollektive mathematische Niveau systematisch für die berufsbezogene Arbeit zu nutzen.

In diesen Befunden wird man keine Bestätigung der These sehen, dass sich die Genese berufsbezogener Kompetenzen im hier untersuchten Feld substitutiv zu den hier erfassten Grundqualifikationen verhält. Vielmehr liegt der Schluss nahe, dass eine solide Grundbildung als notwendige Voraussetzung berufsbezogenen Lernens für die hier betrachtete Gruppe funktiert.

## Literatur

BAETHGE, M., BUSS, K.-P. & LANFER, C. (2003): Konzeptionelle Grundlagen für einen Nationalen Berufsbildungsbericht: Berufliche Bildung und Weiterbildung/Lebenslanges Lernen. Bildungsreform Band 7. Bundesministerium für Bildung und Forschung. Bonn.

BAUMERT, J., BOS, W. & LEHMANN, R. H. (Hrsg.) (2000): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Opladen: Leske + Budrich.

BAUMERT, J., KÖLLER, O., LEHRKE, M. & BROCKMANN, J. (2000). Anlage und Durchführung der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur Sekundarstufe II (TIMSS/III – Technische Grundlagen. In: BAUMERT, J., BOS, W. & LEHMANN, R. (Hrsg.): Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Grundbildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Opladen: Leske + Budrich, 31-84.

BEATON, A. E. & ALLEN, N. L. (1992): Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17 (2), 191-204.

BECK, B. & KLIEME, E. (2003): DESI – Eine Längsschnittstudie zur Untersuchung des Sprachunterrichts in deutschen Schulen. In: *Empirische Pädagogik*, 17 (3), Landau: Verlag Empirische Pädagogik, 380-395.

BLUM, W., NEUBRAND, M., EHMKE, T., SENKBEIL, M., JORDAN, A., ULFIG, F. & CARSTENSEN, C. H. (2004): Mathematische Kompetenz. In: PISA-KONSORTIUM DEUTSCHLAND (Hrsg.): PISA 2003. Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. Münster u.a.: Waxmann, 47-92.

DITTON, H. (1998): Mehrebenenanalyse. Grundlagen und Anwendung des Hierarchisch Linearen Modells. Weinheim und München: Juventa.

FISCHER, G. H. & MOLENAAR, I. W. (1995): Rasch models – Foundations, recent developments, and applications. New York: Springer.

HARTIG, J. (2004): Methoden der Skalierung und Definition von Kompetenzniveaus. Vortrag auf der DESI-Fachtagung "Konzeptualisierung und Messung sprachlicher Kompetenzen", 09. und 10. September 2004, DIPF.

KLIEME, E., BAUMERT, J., KÖLLER, O. & BOS, W. ( 2000): Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In: BAUMERT, J., BOS, W. & LEHMANN, R. (Hrsg.): Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Grundbildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Opladen: Leske + Budrich, 85-133.

LEHMANN, R. H., IVANOV, S., HUNGER, S. & GÄNSFUSS, R. (2005): ULME I. Untersuchung der Leistungen, Motivationen und Einstellungen zu Beginn der beruflichen Ausbildung. Behörde für Bildung und Sport, Amt für Berufliche Bildung und Weiterbildung. Hamburg.

PRENZEL, M., DRECHSEL, B., CARSTENSEN, C. H. & RAMM, G. (2004): PISA 2003 – eine Einführung. In: PISA-KONSORTIUM DEUTSCHLAND (Hrsg.): PISA 2003. Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. Münster u.a.: Waxmann, 13-46.

RAUDENBUSCH, S. W., BRYK, A. S., CHEONG, Y. F. & CONGDON, C. (2001): HLM 5. Hierarchical Linear and Nonlinear Modeling. Lincolnwood, IL: Scientific Software International Inc. 1996

ROST, J. (2004): Lehrbuch Testtheorie – Testkonstruktion, Zweite, vollständig überarbeitete und erweiterte Auflage. Bern, Göttingen: Hans Huber.

WEINERT, F. E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: WEINERT, F. E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim und Basel: Beltz, 17-31.