

---

## Mathematische Grundqualifikationen zu Beginn der beruflichen Ausbildung

---

### 1 Einführung

Seit der Veröffentlichung der Ergebnisse der internationalen Vergleichsstudien *Third International Mathematics and Science Study, Population III* (TIMSS-III: BAUMERT, BOS & LEHMANN 2000) und *Programme for International Student Assessment* (PISA 2000: DEUTSCHES PISA-KONSORTIUM 2000) ist nicht allein die Frage von Bedeutung, wie gut Jugendliche ein bestimmtes schulisches Curriculum beherrschen. Im Mittelpunkt des Interesses der allgemeinen und fachlichen Öffentlichkeit steht vielmehr, wie gut sie ihre Fähigkeiten, ihre Kenntnisse und Fertigkeiten zur Bewältigung realitätsnaher – ‘authentischer’ – Herausforderungen einsetzen können. Das Hauptaugenmerk gilt also dem Verständnis von Zusammenhängen und Prozessen sowie der Fähigkeit, vorhandenes Wissen auf Alltagssituationen zu übertragen und anzuwenden.

Mathematische Grundqualifikationen sind nach allgemeiner Überzeugung für eine solide berufliche Ausbildung und längerfristig für eine erfolgreiche Berufsausübung unerlässlich. Deshalb gehörte ihre Erfassung und Analyse zu den wichtigsten Aufgaben der Hamburger „Untersuchung von Leistungen, Motivation und Einstellungen zu Beginn der beruflichen Ausbildung“ (ULME I; vgl. LEHMANN, IVANOV, HUNGER & GÄNSFUSS 2005).

ULME I ist die erste wissenschaftliche Untersuchung in Deutschland, die in einem Bundesland den Leistungsstand eines gesamten Jahrgangs zu Beginn der beruflichen Ausbildung, differenziert nach Bildungsgängen, erfasst. Sie wurde im Herbst 2002, parallel zur vierten Erhebungswelle der Längsschnittuntersuchung an den allgemein bildenden Schulen in Hamburg „Aspekte der Lernausgangslage und Lernentwicklung“ (LAU 11; vgl. LEHMANN, HUNGER, IVANOV & GÄNSFUSS 2004), durchgeführt. Im Frühjahr 2004 wurde sie ergänzt durch ULME II, eine Untersuchung, die ausschließlich die Schüler der Abschlussklassen der teilqualifizierenden Berufsfachschulen erfasste (vgl. SEEBER in dieser Ausgabe).

In der Vollerhebung ULME I konnten insgesamt 13.048 Schülerinnen und Schüler in 588 Klassen an 48 beruflichen Schulen erfasst werden (vgl. Tabelle 1).

Für die drei hier untersuchten Bildungsgänge an Hamburger berufsbildenden Schulen werden nachstehend die folgenden Abkürzungen verwendet:

- BFStq: Berufsfachschule, teilqualifizierend
- BFSvq: Berufsfachschule, vollqualifizierend
- BS: Berufsschule

Tabelle 1: **Datenstruktur: Anzahl der Jugendlichen nach Schulabschlussniveau und Bildungsgang an berufsbildenden Schulen**

Schulformen in den beruflichen Schulen	Schulabschlussniveau			
	Hauptschule	Realschule	Gymnasium	insgesamt
Berufsfachschule – teilqualifizierend	2.983	55		3.038
Berufsfachschule – vollqualifizierend	97	1.802	260	2.159
Berufsschule	1.730	3.609	2.512	7.851
insgesamt	4.810	5.466	2.772	13.048

Die Lernstände der in der ULME I berücksichtigten Jugendlichen wurden mit dem Hamburger Schulleistungstest für zehnte und elfte Klassen – SL-HAM 10/11 erfasst, der aus den folgenden Untertests besteht:

- Deutsch-Leseverständnis,
- Mathematik I,
- Mathematik II,
- Englisch: C-Test,
- Englisch: Leseverständnis.

Die Untertests lagen in drei Testversionen (Schwierigkeitsstufen) vor:

- Hauptschulversion,
- Realschulversion und
- Gymnasialversion.

Lediglich beim Untertest Mathematik II wurden nur zwei Testversionen verwendet: eine etwas kürzere für die ehemaligen Hauptschüler und eine um fünf Items erweiterte Fassung für die übrigen Probanden. Die Schülerinnen und Schüler bearbeiteten diejenige Testversion, die ihrem zuletzt erreichten Schulabschluss entsprach.

Um eine Vergleichbarkeit der Lernstände über die ca. 200 Lerngruppen mit zum Teil sehr unterschiedlichen Ausbildungszielen hinweg zu ermöglichen, wurden insgesamt 134 ‘Auswertungsgruppen’ (AWG) gebildet. Davon wird im Rahmen dieses Beitrags jedoch abgesehen; es wird lediglich zwischen den drei oben genannten Schulformen bzw. Bildungsgängen unterschieden.

## 2 Erhebungsinstrument „Mathematik II“

Für die Erfassung der mathematischen Grundqualifikation der Jugendlichen zu Beginn der beruflichen Ausbildung wurde in ULME I der Mathematik-II-Test eingesetzt. Um insgesamt den ‘authentischen’, nicht unbedingt streng curricular gebundenen Charakter der Untersuchung zu stärken, wurden bewährte Items aus der Komponente ‘Mathematische Grundbildung’ der TIMSS-III übernommen. Um den zu erwartenden unterschiedlichen Bearbeitungsgeschwindigkeiten Rechnung zu tragen, wurden den (ehemaligen) Hauptschülerinnen und Hauptschülern 12 Aufgaben vorgelegt; die (ehemaligen) Realschülerinnen und Realschüler sowie die Schülerinnen und Schüler mit Fachhochschulreife oder Abitur erhielten weitere 5 Aufgaben zur zusätzlichen Bearbeitung.

Unter Aufnahme von Aspekten des Testkonzepts von PISA wurden in den Mathematik-II-Test außer den genannten TIMSS-III Aufgaben zur mathematischen Grundbildung eine Grafik zur Arbeitslosenstatistik sowie Aufgaben zum Umgang mit Fahrplänen einbezogen. Auf diese beiden ‘diskontinuierlichen Texte’ bezogen sich zusammen 11 Aufgaben, die die Fähigkeit zum Umgang mit entsprechend kodierter Information erforderten (vgl. dazu ausführlicher LEHMANN, IVANOV, HUNGER & GÄNSFUSS 2005, 12).

## 3 Skalierung

Die Auswertung des Untertests Mathematik II sowie der restlichen Untertests des SL-HAM 10/11 basierte – wie beispielsweise in TIMSS, PISA und LAU auch – auf statistischen Modellen der probabilistischen Testtheorie (Item Response Theory; vgl. FISCHER & MOLENAAR 1995; HAMBLETON, SWAMINATHAN & ROGERS 1991), deren bekanntestes Auswertungsmodell das einparametrische Rasch-Modell ist. Eine gut nachvollziehbare Erklärung der Raschskalierung findet sich bei BAUMERT, BOS & LEHMANN (2000). Im Unterschied zu klassischen Methoden der Testauswertung wird bei diesem Verfahren nicht vorausgesetzt, dass alle gestellten Aufgaben gleich schwierig sind. Vielmehr wird das Leistungsniveau (die ‘Fähigkeit’) einer Schülerin oder eines Schülers unter Berücksichtigung der Schwierigkeit der gelösten Aufgaben ermittelt. Das Verfahren erlaubt es,

- die Einschlägigkeit von Testaufgaben für die theoretisch bestimmte Fähigkeitsdimension empirisch zu prüfen und ungeeignete Items auszusondern,
- den Schwierigkeitsgrad von Testaufgaben und die Fähigkeit von Personen auf demselben Maßstab abzubilden,
- die Fähigkeit einer Person zuverlässig zu schätzen, auch wenn nur eine Teilmenge der Aufgaben bearbeitet wurde, sowie
- Personenfähigkeiten und Aufgabenschwierigkeiten auf die Anforderungsniveaus hin zu analysieren (zur Bildung von Kompetenzstufen – zum so genannten ‘Proficiency-Scaling’ – im Rahmen von TIMSS vgl. KLIEME, BAUMERT, KÖLLER & BOS 2000,

116ff.; für PISA 2003 vgl. BLUM, NEUBRAND, EHMKE, SENKBEIL, JORDAN, ULFIG & CASTENSEN 2004, 55ff.).

Bei dem hier beschriebenen Skalierungsverfahren wird die Fähigkeit einer Schülerin bzw. eines Schülers über Wahrscheinlichkeiten definiert. Dabei steigt die geschätzte Fähigkeit mit der Gesamtzahl der jeweils richtigen Antworten; andererseits hängt die Festsetzung der Aufgabenschwierigkeiten von der Anzahl richtiger Lösungen in der Stichprobe ab. Wenn sich nun die Antwortmuster der Schülerinnen und Schüler in guter Näherung so darstellen lassen, dass die einzelnen Schülerinnen und Schüler die Aufgaben jeweils nur bis zu einer für sie charakteristischen Schwierigkeit lösen, so können für alle Aufgaben und Fähigkeitsgruppen die Lösungswahrscheinlichkeiten nach diesem Modell berechnet werden. Bei der konventionellen Festlegung der Lösungswahrscheinlichkeit für die Entsprechung von Personenfähigkeit und Aufgabenschwierigkeit auf  $p = 0,65$  geht man davon aus, dass ein in diesem Sinne kompetenter Proband 'auf lange Sicht' etwa bei jedem dritten Versuch an einer Aufgabe *seines* Fähigkeits- und Schwierigkeitsgrades scheitern wird (vgl. LEHMANN & NIKOLOVA 2005).

Übersteigt die auf solcher Grundlage geschätzte Fähigkeit einer Schülerin oder eines Schülers die Schwierigkeit einer Aufgabe, so wird sie bzw. er diese umso eher lösen, je größer die Differenz zwischen Fähigkeits- und Schwierigkeitsparameter ist, mindestens aber mit der genannten Wahrscheinlichkeit von  $p = 0,65$ . Unterschreitet umgekehrt die geschätzte Fähigkeit des Schülers oder der Schülerin die Aufgabenschwierigkeit, so wird er oder sie umso eher scheitern, je größer die Differenz ist.

Da der Untertest Mathematik II in wesentlichen Teilen auf dem Grundqualifikationstest von TIMSS III beruht, erschien ein Versuch angeraten, die dort definierten Kompetenzstufen auf den Hamburger Test zu übertragen. Wegen der unterschiedlichen Metriken – TIMSS war auf einen Mittelwert von 500 und eine Standardabweichung von 100 Punkten *für die internationale Gesamtpopulation* zugeschnitten, der Untertest Mathematik II dagegen auf einen Mittelwert 100 und eine Standardabweichung von 25 Punkten *für die Teilpopulation der Schülerschaft zu Beginn beruflicher Ausbildungsgänge in Hamburg*, bei unterschiedlichen Ankerpunkten und sogar abweichenden messtheoretischen Grundannahmen – musste ein Weg gefunden werden, wenigstens näherungsweise zu einer Entsprechung der beiden Skalen zu gelangen.

Für die gesuchte Approximation wurden die aus TIMSS übernommenen Items in einer einfachen Regressionsanalyse als 'Fälle' behandelt, wobei die Schwierigkeitsparameter der TIMSS-Items auf der neu gebildeten ULME-Skala als abhängige Variable und die internationalen Schwierigkeitsparameter der TIMSS-Skala als unabhängige Variable verwendet wurden. So konnten mithilfe der so ermittelten Regressionskoeffizienten die gegebenen Schwellen zwischen den für TIMSS definierten Kompetenzstufen näherungsweise bestimmt bzw. auf die neu entwickelte Skala projiziert werden. Die entsprechende Transformationsgleichung lautet für alle Schwierigkeitsparameter  $\theta_j$  der beiden Tests gemeinsamen Anker-Items:

$$\theta_j (\text{ULME}) = 40,88 + 0,17 * \theta_j (\text{TIMSS}).$$

Obwohl es sich hierbei nur um ein verhältnismäßig grobes Näherungsverfahren handelt, konnte dennoch auf diesem Wege gezeigt werden, dass einige leichte Veränderungen an den Items nicht mit bedeutenden systematischen Verschiebungen verknüpft waren: Sieben Items waren bei der Übernahme aus TIMSS um einen Distraktor verkürzt worden; die Anwendung der o.g. Transformationsformel lieferte deshalb geringfügig, nämlich um durchschnittlich 2,5 Skalenpunkte, verminderte Schwierigkeitswerte auf der ULME-Skala, verglichen mit einer scheinbaren Schwierigkeitssteigerung um durchschnittlich 1,7 Punkte für die acht Items, die aus der Originalversion übernommen worden waren.

#### **4 Anforderungsstruktur des ULME-Subtests Mathematik II**

Um einen Einblick in die Anforderungsstruktur des Untertests Mathematik II des SL-HAM 10/11 zu geben, ist in Abbildung 1 die nach Maßgabe der Rasch-Skalierung erfolgte wechselseitige Projektion der Aufgabenschwierigkeiten und der bei den Schülerinnen und Schülern ermittelten mathematischen Fähigkeiten dargestellt. Schon in dieser Darstellung zeigt sich an der Verschiebung der beiden Hälften der Grafik gegeneinander, dass die schülerseitige Kompetenzverteilung in ihrem Schwerpunkt hinter den Anforderungen der hier ausgewählten Aufgaben zurückbleibt. Noch aussagekräftiger ist aber die Einteilung des Skalenbereichs und damit auch der Leistungsverteilung in die mit römischen Zahlen gekennzeichneten Kompetenzstufen, die nach dem oben geschilderten Verfahren aus der Kompetenzstufenhierarchie der TIMSS-III-Aufgaben näherungsweise abgeleitet werden konnte. Damit erscheint es auch als legitim, die inhaltlichen Kompetenzbeschreibungen, die auf der Grundlage der TIMSS-Daten von Expertengruppen erarbeitet worden sind, für eine Charakterisierung der fachlichen Anforderungen des Untertests Mathematik II der ULME zumindest probeweise zu übernehmen. Um diese Interpretationen zu konkretisieren, ist jeder Kompetenzstufe ein charakteristisches Item zugeordnet, das von mindestens 65 Prozent der Schülerinnen und Schülern des fraglichen Niveaus erfolgreich bearbeitet worden ist, von der nächst niedrigeren Stufe jedoch von weniger als 50 Prozent (vgl. einzige Ausnahme: das Beispiel für Kompetenzstufe I wurde in der untersten Fähigkeitsgruppe von 57 Prozent korrekt gelöst).

Wenn auch mit diesen Entscheidungsregeln die strengeren Kriterien von BEATON & ALLEN (1992), die einen größeren Abstand zwischen den Kompetenzstufen vorsehen, nicht ganz haben eingehalten werden können, so folgen die exemplarischen Aufgaben doch in wesentlichen Punkten den von KLIEME et al. (2000) vorgeschlagenen Interpretationen.

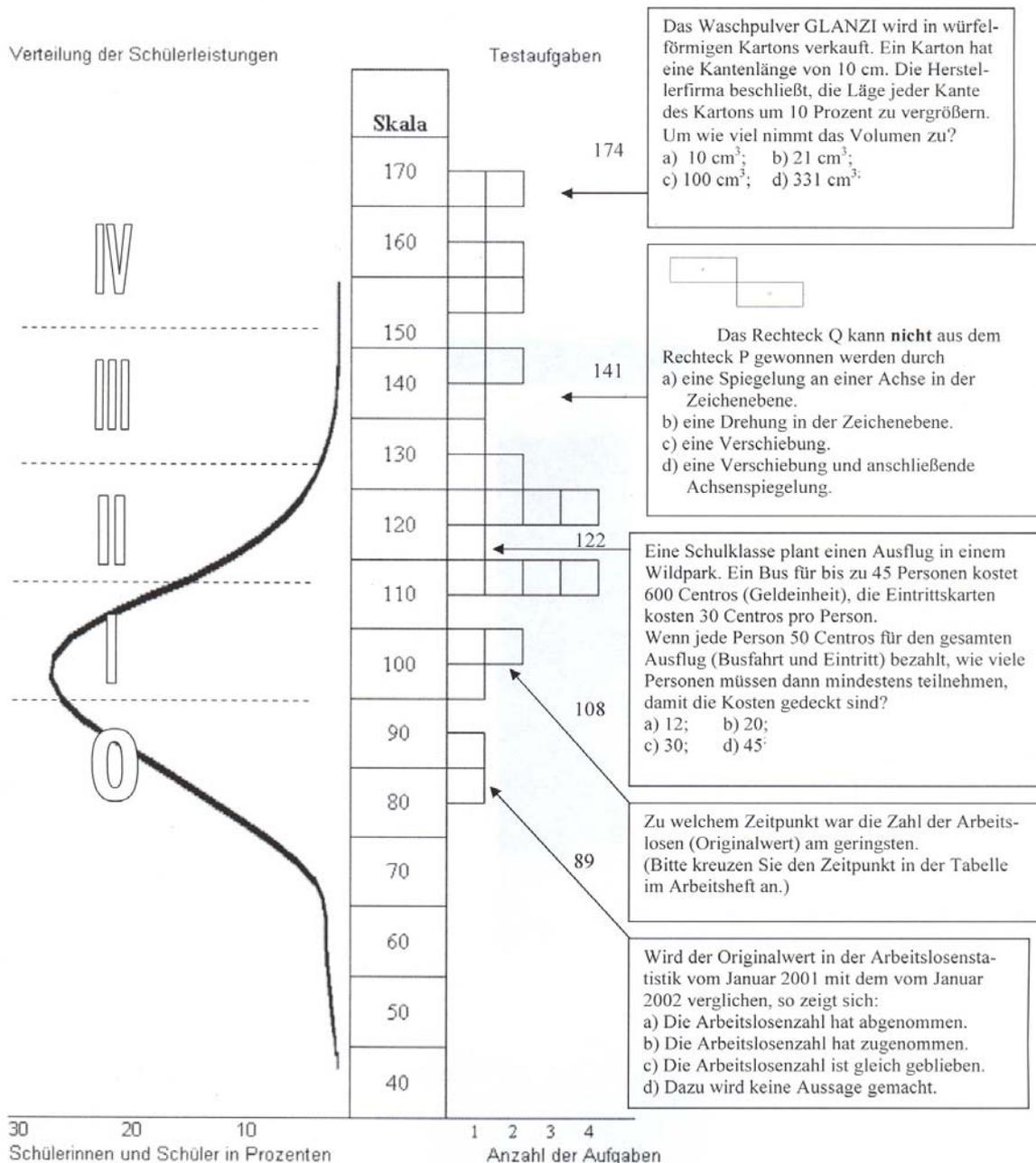


Abb. 1: Verteilung der Schülerfähigkeiten und Aufgabenschwierigkeiten im Mathematik-II-Test

1. *Kompetenzstufe I* ist in TIMSS durch eine Aufgabe repräsentiert, die keine expliziten mathematischen Operationen verlangt, sondern lediglich intuitive, alltagsnahe Überlegungen. In ähnlicher Weise kann man auch im ULME-Subtest Mathematik II die zweiteinfachste Beispielaufgabe mit dem Schwierigkeitsparameter 108 lösen, indem mental die lebensweltlich vertraute Zuordnung von geometrischen zu numerischen Relationen vorgenommen wird. KLIEME et al. (2000, 123) schlagen für die Anforderung

derungen dieses Kompetenzniveaus die Bezeichnung „*alltagsbezogene Schlussfolgerungen*“ vor.

2. *Kompetenzstufe II* ist dagegen in TIMSS-III durch Items vertreten, die die *Anwendung von einfachen Routinen* erfordern (a.a.O.). Zu erwarten ist also, dass die für den ULME-Test ausgewählte charakteristische Aufgabe mit dem Schwierigkeitsparameter 122 ebenfalls solche Kompetenzen anspricht. Es ist leicht zu rekonstruieren, dass die hier verlangte Fähigkeit zur Anwendung basaler Routinen lediglich die Kombination einer einfachen Subtraktion mit einer einfachen Division [ $X = 600/(50-30)$ ] umfasst. Allerdings sind erste *Ansätze* zum mathematischen Modellieren erforderlich, wie sie in TIMSS sonst eher für die Kompetenzstufe III typisch sind.
3. *Kompetenzstufe III* wird in TIMSS so charakterisiert, dass es dabei um *mathematisches Modellieren auf einfachem Niveau* gehe (a.a.O., 125). Das dort gegebene Beispiel kombiniert die Berechnung von Volumina mit der Verhältnisrechnung. Dagegen setzt das ULME-Beispiel mit dem Schwierigkeitsparameter 141 die Beherrschung der geometrischen Fachterminologie und ein gutes räumliches Vorstellungsvermögen voraus; zusätzlich wird es durch die **negative** Frageformulierung erschwert. Insoweit ist die Vermutung äquivalenter Schwierigkeiten zwischen diesen in TIMSS und ULME festgelegten Kompetenzstufen nicht unplausibel, bedarf aber der Überprüfung an weiteren Testaufgaben.
4. *Kompetenzstufe IV* schließlich, für die keine obere Grenze existiert, wird in ULME durch ein Beispielitem vertreten, das in seinen formalen Anforderungen stark demjenigen des charakteristischen TIMSS-Items für Stufe III (B 17, vgl. a.a.O., 124) ähnelt. Zusätzliche Schwierigkeit generierende Faktoren sind in ULME die abstraktere Form der Frage – so fehlt eine die Modellierung unterstützende Grafik – und das lediglich implizierte Potenzieren, das an die Stelle der expliziten Multiplikation tritt. Insoweit erscheint die für TIMSS (a.a.O., 125) gewählte Schwierigkeitsumschreibung – „*mathematisches Argumentieren, insbesondere anhand graphischer Darstellungen*“ – hier kaum angemessen. Sie beruht vielleicht auch innerhalb des TIMSS-Bezugsrahmens auf einer zu schmalen und wenig generalisierbaren Datenbasis. Unter Bezug auf die vorliegenden Daten wird man bei dieser Stufe wohl eher von der *Anwendung komplexer Routinen im Rahmen eigenständiger Modellierungen* sprechen.

Der inhaltliche Abgleich dieses ersten Ansatzes zur Definition von Kompetenzstufen im Rahmen des Untertests Mathematik II der ULME mit der für TIMSS vorgeschlagenen Hierarchie mathematischer Grundqualifikationen zeigt die erwarteten Struktur analogien. Er legt aber zugleich die Vermutung nahe, dass die Stufenabgrenzungen im oberen Teil des Leistungsspektrums im Falle der ULME tendenziell niedrigere Anforderungen spezifizieren, als dies in der TIMSS der Fall war. Fragt man also auf dieser Basis nach den Grundqualifikationen der Hamburger Jugendlichen zu Beginn der beruflichen Ausbildung, so wird man unmittelbare Vergleiche mit Prozentzahlen, die sich auf TIMSS beziehen, eher vermeiden. Gleichzeitig bleibt das beschriebene Verfahren durchaus von heuristischem Wert, insbesondere dann,

wenn für die ULME-Skala und die darauf bezogenen Kompetenzstufen gezeigt werden kann, dass es sich dabei um reliable und valide Maße handelt.

## 5 Leistungen im ULME-Subtest Mathematik II, differenziert nach Bildungsgängen

Die in Abbildung 1 gezeigte Leistungsverteilung enthielt keinerlei Bezug zu den Schulformen, deren Unterschiede bereits durch die Erstausswertung der Daten (LEHMANN et al. 2005, 42 ff.) belegt worden waren. Es liegt nun nahe, an die Stelle einer grafischen Darstellung mit schwer interpretierbaren Überschneidungsbereichen eine Aufschlüsselung der Testergebnisse nach Schulform und erreichtem Kompetenzniveau zu setzen. Die Leistungsunterschiede zwischen den drei hier untersuchten beruflichen Schulformen können auf diese Weise in konkretem Bezug zu den oben bestimmten Kompetenzstufen veranschaulicht werden. Tabelle 2 zeigt, wie sich die Schülerschaft jeweils auf diese Niveaus verteilt.

Tabelle 2: **Mathematik II: Verteilung der Schülerinnen und Schüler in den Kompetenzstufen nach Schulformen (in Prozent)**

Schulform	Kompetenzstufen					N
	0	I	II	III	IV	
BFStq	79,1	16,4	4,0	0,5	0,1	2.836
BFSvq	45,3	33,7	15,8	4,5	0,7	2.076
BS	37,6	28,8	20,3	10,5	2,8	7.670
<i>insgesamt</i>	48,2	26,8	15,2	7,3	1,8	12.582

An diesen Prozentwerten zeigt sich einerseits – in der Summe über alle Schulformen –, dass knapp die Hälfte der Jugendlichen die erste Kompetenzstufe, d. h. das Niveau „alltagsbezogener (sc. mathematischer) Schlussfolgerungen“, zu Beginn der beruflichen Ausbildung (noch) nicht erreicht haben. Hiervon sind insbesondere die teilqualifizierenden Bildungsgänge betroffen, in denen nur etwa jeder Fünfte mindestens die Stufe I bewältigt, jedoch weniger als 5 Prozent die Stufe II oder höher. In den vollqualifizierenden Bildungsgängen der Berufsfachschulen ist die Situation zu Beginn der Ausbildung bereits wesentlich günstiger, doch auch hier liefert der Umstand, dass sich nur jeder Fünfte zu Beginn der Ausbildung wenigstens auf Kompetenzstufe II, d. h. auf der Ebene der „Anwendung von einfachen (mathematischen) Routinen“ einigermaßen sicher bewegen kann, durchaus Anlass zur Sorge. An den Berufsschulen schließlich sind immerhin bei etwa einem Drittel der Jugendlichen mathematische Grundqualifikationen auf dem Niveau der einfachen mathematischen Routinen oder darüber vorauszusetzen (vgl. Abbildung 2).

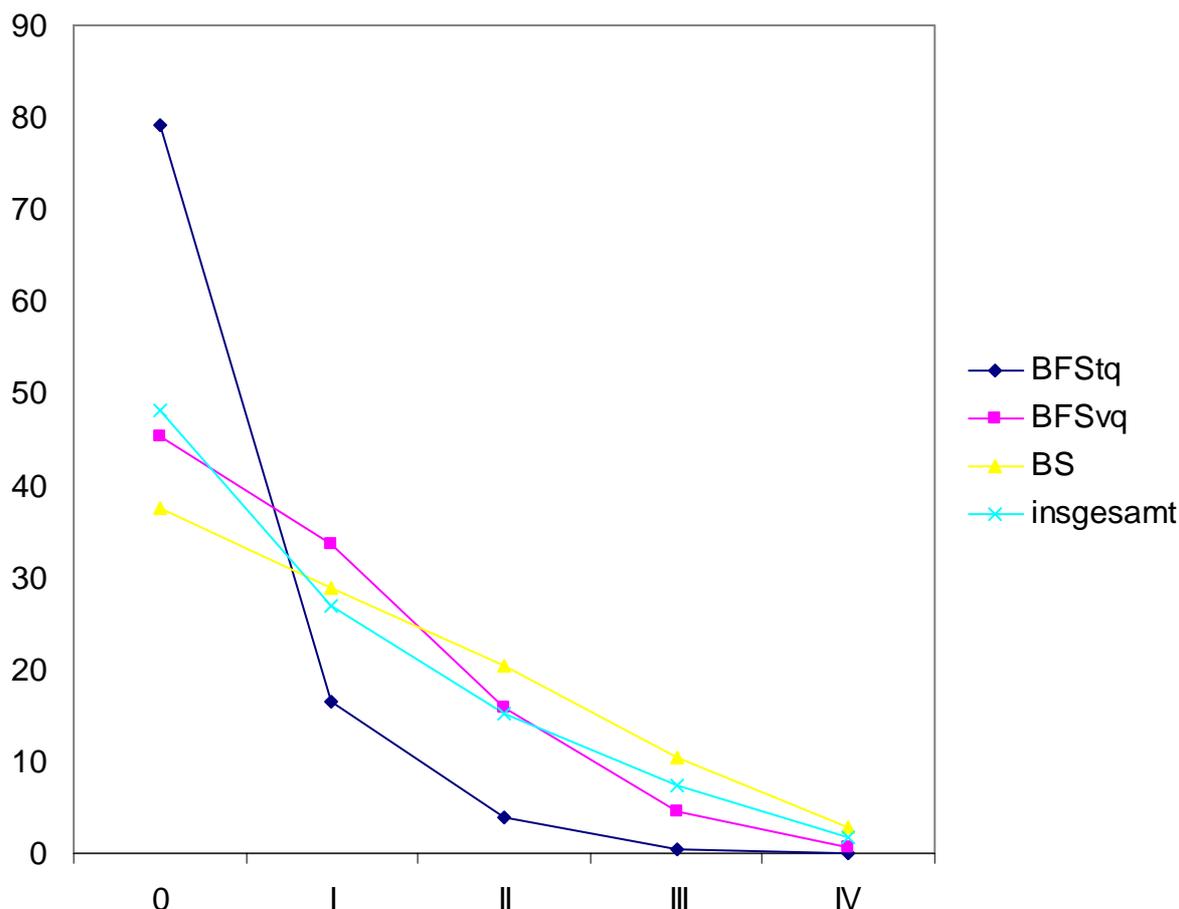


Abb. 2: Mathematik II: Verteilung der Schülerinnen und Schüler in den Kompetenzstufen nach Schulformen (in Prozent)

Ähnliche Leistungsunterschiede sind sichtbar, wenn man zwischen den Jugendlichen nach Art ihres bisher erreichten Schulabschlusses unterscheidet, jetzt aber die Messwerte unmittelbar zum Vergleich heranzieht (vgl. Tabelle 3.). So beträgt z. B. die Leistungsunterschied zwischen den Jugendlichen mit Realschulabschluss und jenen mit Abitur oder einer Versetzung in die gymnasiale Oberstufe knapp eine (Gesamt-) Standardabweichung zugunsten der letzteren; d. h. sie entspricht dem durchschnittlichen Lernzuwachs in der Mittelstufe von ca. drei Lernjahren. Die Jugendlichen mit Hauptschulabschluss haben einen durchschnittlichen Lernstand erreicht, der nochmals um den Gegenwert von mehr als zwei Lernjahren unter dem der ehemaligen Realschüler steht. Diese Angaben mögen genügen, um anzuzeigen, mit welcher Heterogenität im Fach Mathematik die berufsbildende Sekundarstufe II zu Beginn ihrer Ausbildungsgänge konfrontiert ist.

Tabelle 3: **Mathematik II: Ergebnisse zu Beginn der beruflichen Ausbildung: Mittelwerte, Standardabweichungen und Effektstärken, insgesamt und getrennt nach Schulabschlüssen**

Schulabschluss	Mittelwert	Standardabweichung	Effektstärke $d$	$N$
Hauptschüler	83,8	21,8	-0,65	4.549
Realschüler	102,5	19,2	0,10	5.315
Schüler mit Abitur oder Versetzung in die gymnasiale Oberstufe	122,2	20,8	0,89	2.718
<i>insgesamt</i>	100,0	25,0		12.582

## 6 Zur Validität des ULME-Subtests Mathematik II

Die konzeptionellen Vorgaben für den Subtest Mathematik II sahen keine direkte curriculare Verankerung dieses Instruments vor; vielmehr ging es ähnlich wie bei TIMSS und vor allem dann PISA um *Grundqualifikationen*, die zwar in Zusammenhang mit der ‘Schulmathematik’ stehen mögen, die sich jedoch vor allem in den biographischen Phasen danach, allen voran in der beruflichen Ausbildung, zu bewähren haben.

Bereits in der ersten Berichtslegung zu den Testergebnissen (LEHMANN et al. 2005, 46) war dargestellt worden, dass der strenger curricular begründete Untertest Mathematik I und der an Grundqualifikationen orientierte Untertest Mathematik II mit  $r = 0,73$  relativ hoch untereinander korrelieren. Zum damaligen Zeitpunkt lagen allerdings noch keine skalierten Daten für den berufsbezogenen Untertest Wirtschaft und Verwaltung aus ULME II vor (vgl. hierzu SEEBER in dieser Ausgabe). Es liegt deshalb nahe, die damit gegebene Möglichkeit zur Prüfung der Übereinstimmung zwischen diesem berufsbezogenen Test und den Testergebnissen Mathematik II zu nutzen und so einen Anhaltspunkt für dessen prognostische Validität zu gewinnen.

Einschränkend ist an dieser Stelle zu sagen, dass von den ursprünglich 1.403 Jugendlichen, die sich im Jahre 2002 (ULME I) in kaufmännischen Bildungsgängen befanden und denen dort der Untertest Mathematik II vorgelegt worden war, im Jahre 2004 (ULME II) nur mehr 792 (also 56,5 Prozent) die beiden hier interessierenden Tests bearbeitet haben. Ursachen hierfür sind vor allem in den Ausbildungsabbrüchen zu finden, deren Hintergründe an anderer Stelle zu untersuchen sind. Jedenfalls aber ist davon auszugehen, dass damit Abschwächungen der Korrelationen einhergegangen sind, die tendenziell zu einer Unterschätzung der jetzt durchgeführten Validitätsschätzungen geführt haben. Immerhin darf es als Bestätigung der getroffenen Annahmen gelten, dass die Testleistungen Mathematik II im Jahre 2002 mit den

erzielten Testergebnissen im Test Wirtschaft und Verwaltung 2004 in Höhe von  $r = 0,28$  korreliert sind und die Testleistungen Mathematik II im Jahre 2004 sogar in Höhe von  $r = 0,36$ , wobei die Wiederholungszuverlässigkeit für Mathematik II allein über zwei Jahre hinweg mit  $r = 0,34$  sogar noch etwas geringer ausgefallen ist.

Demnach kann also davon ausgegangen werden, dass mit dem Untertest Mathematik II der ULME nicht nur ein in sich aussagekräftiges Instrument zur Messung mathematischer Grundqualifikationen vorliegt, sondern darüber hinaus eine wichtige Erkenntnisquelle für die Optimierung der vorberuflichen und beruflichen Ausbildung, namentlich im Bereich eher schwach ausgeprägter Kompetenzen.

## Literatur

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.

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

BEHÖRDE FÜR BILDUNG UND SPORT, AMT FÜR SCHULE, HAMBURG (Hrsg.) (2002): Testentwicklung: Zentrum für empirische pädagogische Forschung (ZEPF) der Universität Koblenz-Landau, Abteilung Empirische Bildungsforschung und Methodenlehre der Humboldt-Universität zu Berlin, Behörde für Bildung und Sport in Hamburg.

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.

DEUTSCHES PISA-KONSORTIUM (Hrsg.). (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.

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

HAMBLETON, R. K., SWAMINATHAN, H. & ROGERS, H. J. (Eds.). (1991): Fundamentals of item response theory. Newbury Park, CA: Sage.

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. & NIKOLOVA, R.(2005): ELEMENT. Erhebungen zum Lese- und Mathematikverständnis. Entwicklungen in den Klassenstufen 4 bis 6 in Berlin. Bericht über die Untersuchung 2003 an Berliner Grundschulen und grundständigen Gymnasien. Berlin: Senatsverwaltung für Bildung, Jugend und Sport.

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

LEHMANN, R.H., HUNGER, S., IVANOV, S., GÄNSFUß, R. (2004). LAU 11. Aspekte der Lernausgangslage und Lernentwicklung, Klassenstufe 11. Ergebnisse einer Längsschnittstudie. Hamburg: Behörde für Bildung und Sport.