[Eingereichte Fassung; Zitationen erfolgen bitte nach dem Original: Blömeke, S., Schwarz, B., Kaiser, G., Seeber, S. & Lehmann, R. (2009). Untersuchungen zum mathematischen und mathematikdidaktischen Wissen angehender GHR- und Gymnasiallehrkräfte. *Journal für Mathematik-Didaktik, 30(3-4),* 232-255.]

Sigrid Blömeke, Björn Schwarz, Gabriele Kaiser, Susan Seeber und Rainer Lehmann

Untersuchungen zum mathematischen und mathematikdidaktischen Wissen angehender GHR- und Gymnasiallehrkräfte

Kurzfassung

Die international-vergleichenden Studie zur Leistungsfähigkeit der Lehrerausbildung „Mathematics Teaching in the 21st Century“ (*MT21*) zeigt für die deutsche Stichprobe auf, dass das mathematische und mathematikdidaktische Wissen angehender Mathematiklehrkräfte für das Gymnasium im Mittel deutlich umfangreicher ist als das angehender Lehrkräfte für Grund-, Haupt- und Realschulen. Nichtsdestotrotz werden einige Items von der Gruppe angehender deutscher Lehrkräfte für Grund-, Haupt- und Realschulen relativ häufiger gelöst als von der Gruppe der angehenden Gymnasiallehrkräfte, andere fallen ihnen in Relation zu den übrigen Items des Itempools deutlich leichter. Umgekehrt lassen sich Items finden, bei denen der Leistungsvorsprung angehender Gymnasiallehrkräfte besonders groß ist. Im vorliegenden Beitrag wird untersucht, was die beiden Item-Sets auszeichnet, die entsprechende differenzielle Itemfunktionen aufweisen. Auf diese Weise kann das besondere Leistungspotenzial sowohl angehender angehender Lehrkräfte für Grund-, Haupt- und Realschulen als auch angehender Gymnasiallehrkräfte herausgearbeitet werden.

Abstract

Results of the study „Mathematics Teaching in the 21st Century“ (*MT21)*, a comparative study on teacher education across six countries, showed for the German sample that future high-school teachers on average developed higher knowledge than future primary and lower-secondary teachers. However, the latter performed relatively well on some items, other items were less difficult for them in relation to the whole item-pool. In contrast, there were items on which future high-school teachers achieved test scores way above the average difference. In this paper, we analyze the two sets of items that showed differential item functioning in order to work out the specific strengths of future primary and lower-secondary teachers as well as of future high-school teachers.

1 Einleitung

In der 2006 durchgeführten international-vergleichenden Studie „Mathematics Teaching in the 21st Century (*MT21*)“ wurde erstmals die professionelle Kompetenz angehender Sekundarstufen-I-Lehrkräfte am Ende der Lehrerausbildung in standardisierter Form getestet.[1] Die Studie wurde in Bulgarien, Deutschland, Südkorea, Taiwan, Mexiko und den USA durchgeführt. In Analysen zum mathematischen und mathematikdidaktischen Wissen der deutschen Stichprobe wurde deutlich, dass jene Lehrkräfte, die sich in einem Ausbildungsgang für das Gymnasium und die Gesamtschule befinden, deutliche Leistungsvorsprünge gegenüber jenen aufweisen, die sich in einem Ausbildungsgang für Grund-, Haupt- und Realschulen befinden (Blömeke et al., 2008a). Wir verzichten aus Platzgründen auf eine Darstellung der Ergebnisse der Studie und verweisen stattdessen auf bereits vorliegende umfangreiche Publikationen (u.a. Blömeke, Kaiser & Lehmann, 2008). In diesem Beitrag gehen wir der Frage nach, wie Unterschiede zwischen den beiden untersuchten deutschen Gruppen mit den in den Leistungstests verwendeten Items zusammenhängen – kurz gesagt, bei welchen Items zukünftige GHR-Lehrkräfte bzw. Gymnasiallehrkräfte substanziell bessere bzw. schlechtere Leistungen erzielt haben als aufgrund der durchschnittlichen Leistungsunterschiede zu erwarten gewesen wäre.

1.1 Ziel des vorliegenden Beitrags

In der deutschen Stichprobe zu *MT21* wurde ein deutlicher Leistungsvorsprung angehender Gymnasial- und Gesamtschullehrkräfte gegenüber angehenden Grund-, Haupt- und Realschullehrkräften festgestellt. Dieses Ergebnis trifft weitgehend unabhängig davon zu, welche Dimension des fachbezogenen Wissens untersucht wurde. Angehende Gymnasial- und Gesamtschullehrkräfte erreichen signifikant bessere Leistungswerte in den beiden großen Ausbildungskomponenten Mathematik und Mathematikdidaktik, aber auch in den jeweils näher untersuchten fünf Bereichen Arithmetik, Algebra, Funktionen, Geometrie und Stochastik sowie den drei kognitiven Dimensionen Algorithmisieren, Problemlösen und Modellieren. Die dabei jeweils erreichten Effektstärken erreichen zum Teil beachtliche Ausmaße und deuten eine hohe praktische Relevanz der Leistungsunterschiede an.

Dieses Ergebnis bedeutet allerdings nicht, dass angehende GyGS-Lehrkräfte *jedes* Item der untersuchten Dimensionen zu einem größeren Anteil lösen als angehende GHR-Lehrkräfte oder dass den beiden Gruppen dieselben Items besonders leicht oder schwer fallen. Im Gegenteil ist festzustellen, dass einzelne Items von der GHR-Gruppe häufiger gelöst werden als von der GyGS-Gruppe. Andere Items werden von angehenden GHR-Lehrkräften zwar nicht häufiger gelöst, sie fallen ihnen aber in Relation zu den übrigen Items des Itempools deutlich leichter als angehenden GyGS-Lehrkräften. Umgekehrt gibt es Items, auf denen Letztere einen besonders starken Leistungsvorsprung zeigen, der die mittlere Differenz zwischen den beiden Gruppen übersteigt.

[1] Eine Förderung der Studie erfolgt durch die National Science Foundation (REC-0231886), durch die Alexander von Humboldt-Stiftung, durch die Humboldt-Universtität zu Berlin und durch die Michigan State University.

Diese gruppenspezifischen Unterschiede in der Itemschwierigkeit sind Ausdruck davon, dass ein theoretisch entwickeltes Testmodell keine perfekte empirische Anpassung aufweisen kann. Auch wenn sich alle Modellierungen des fachbezogenen Wissens angehender Mathematiklehrkräfte in *MT21* als sehr gut passend erwiesen haben (Blömeke et al., 2008c), zeigen sich Abweichungen in den Daten. Dies liegt darin begründet, dass Ziel unseres Tests war, die zu erfassenden Konstrukte möglichst breit abzudecken. Damit ist gleichzeitig vorgezeichnet, dass – über zufällige Schätzfehler hinaus – Teile des Tests ggf. nicht von allen Personen gleich häufig gelöst werden, da unterschiedliche Fähigkeiten zum Tragen kommen. Lassen sich systematische Abweichungen für ganze Item-Sets bei Gruppen an Personen feststellen, spricht man von „differentiellen Itemfunktionen" (Budgell, Namburty & Douglas, 1995). Unabhängig von generellen Unterschieden in der mittleren Testleistung weisen Items in diesem Falle unterschiedliche Messeigenschaften auf, in Bezug auf den vorliegenden Beitrag insbesondere unterschiedliche Schwierigkeitsparameter, aber es kann sich auch um unterschiedliche Trennschärfen handeln.

Ziel des vorliegenden Beitrags ist es, vor diesem Hintergrund zu untersuchen, was Items auszeichnet, die besonders gut von angehenden GHR- bzw. GyGS-Lehrkräften gelöst werden, die also lehramtsspezifische differentielle Itemfunktionen hinsichtlich der Schwierigkeitsparameter aufweisen. Auf diese Weise können die mathematischen und mathematikdidaktischen Stärken der beiden Gruppen im Detail herausgearbeitet werden, um daraus Schlüsse für die Gestaltung der Lehrerausbildung zu ziehen. Beide Ausbildungsgänge haben in der Vergangenheit Kritik auf sich gezogen. Während in Bezug auf die Gymnasiallehrerausbildung die Relation von fachlicher und fachdidaktischer Ausbildung in Frage gestellt wird, wird die deutlich kürzere und weniger fachlich fokussierte GHR-Ausbildung häufig sogar grundsätzlich in Frage gestellt. Uns geht es vor allem darum, den Blick einmal abzuwenden von Defiziten und stattdessen zu analysieren, in welchen Bereichen *besonders gute* Leistungen erbracht werden – nicht dass eine Reform der Lehrerausbildung bisherige Stärken aus Unkenntnis unterminiert.

Zum besseren Verständnis der Studie wird im Folgenden zunächst der theoretische Rahmen von *MT21* dargestellt und die dort vorgenommene Konzeptualisierung wird in Beziehung zu anderen Studien gesetzt, bevor die untersuchten Hypothesen entwickelt werden.

1.2 Theoretischer Rahmen von *MT21*

Ausgangspunkt der Konzeptualisierung von *MT21* sind Überlegungen zu den zentralen analytischen Subdimensionen professionellen Wissens von Lehrkräften, wie sie grundlegend von Shulman (1986) formuliert und u.a. von Bromme (1992) weiterentwickelt und ausdifferenziert wurden (für die Diskussion zu professionstheoretischen Grundlagen siehe Blömeke, 2002). *MT21* orientiert sich darüber hinaus an der Weinertschen Grundkonzeption von Kompetenz, wonach diese anforderungsbezogen zu formulieren ist, d.h. in diesem Falle bezogen auf die beruflichen Aufgaben zukünftiger Mathematiklehrkräfte wie Lehren und Diagnostizieren, und wonach neben kognitiven Subdimensionen zudem affektiv-motivationale Subdimensionen für eine erfolgreiche Bewältigung beruflicher Anforderungen zu berücksichtigen sind.

Nach Shulman (1986) kann Lehrerwissen in die drei folgenden großen Dimensionen ausdifferenziert werden, die wir in *MT21* weiter operationalisiert haben:

(1) mathematisches Wissen (*content knowledge*), unterteilt nach

- den geforderten kognitiven Aktivitäten der angehenden Mathematiklehrkräfte, worunter – orientiert an den fundamentalen Ideen der Mathematik – Algorithmisieren, Problemlösen/Begründen und Modellieren verstanden wird (im Anschluss an Überlegungen von Tietze, Klika & Wolpers, 1997);
- mathematischen Inhaltsbereichen, nämlich Arithmetik, Algebra, Funktionen, Geometrie und Stochastik;
- Niveaustufen, d.h. Mathematik der Sekundarstufe I, Mathematik der Sekundarstufe II, Schulmathematik vom höheren Standpunkt und universitäre Mathematik (im Anschluss an Vorstellungen von Klein, 1933; Kirsch, 1987).

(2) mathematikdidaktisches Wissen (*pedagogical content knowledge*):

- mathematische Inhaltsbereiche wie bei (1);
- lehrbezogene Anforderungen curricularer und unterrichtsplanerischer Art, die bereits vor dem Unterricht anfallen wie die Auswahl, Begründung und angemessene Vereinfachung der fachlichen Inhalte für die Schülerinnen und Schüler sowie Fragen des Aufbaus mathematischer Kompetenz über die Schuljahre hinweg (vgl. u.a. Vollrath, 2001);
- lernprozessbezogene Anforderungen betreffend das unterrichtliche Handeln von Lehrperson in der unmittelbaren Interaktion mit Schülerinnen und Schülern, d.h. Antworten – seien sie verbaler oder schriftlicher Art als Reaktion auf Aufgaben oder Fragen – bezüglich kognitiver Niveaus, Komplexität der Struktur sowie eventueller Fehler und Fehlermuster einzuordnen, Rückmeldungen zu geben und angemessen mit Interventionsstrategien darauf zu reagieren (ebd.).

(3) pädagogisch-psychologisches Wissen (*general pedagogical knowledge*), fokussiert auf lehr- und diagnosebezogenen Fragen.

Im Anschluss an Bromme (1992) wird dabei mathematikdidaktisches Wissen als derjenige zentrale Bereich aufgefasst, in dem mathematisches Wissen, allgemeine Vorstellungen von Mathematik, Wissen über curriculare Konzeptionen zum Mathematikunterricht und unterrichtspraktische Aspekte sowie das Wissen über Schülervorstellungen aufeinander bezogen werden.

Professionelle Kompetenz umfasst neben kognitiv orientierten Wissensdimensionen auch affektiv-wertorientierte Aspekte, die in *MT21* ebenfalls erfasst werden. Im Hinblick auf Überzeugungen wird zwischen

- epistemologischen Überzeugungen zur Mathematik,
- Überzeugungen zum Lehren und Lernen von Mathematik sowie
- schul- und professionstheoretischen Überzeugungen

unterschieden. Aus Platzgründen können wir nicht ins Detail gehen, weisen jedoch darauf hin, dass sich die Ansätze der deutschen Gruppe, auf die sich im Folgenden bezogen wird, in spezifische Traditionen der deutschen Mathematikdidaktik eingebettet sind. Diese Ansätze haben starke philosophische Wurzeln, sind aber nicht unabhängig von empirischen Studien zu sehen. Das Fachwissen spielt in diesen stoffdidaktisch geprägten Analysen traditionell eine große Rolle (Pepin, 1999).

Ähnliche Konzeptualisierungen des Professionswissens von Mathematiklehrkräften werden auch in anderen Studien verwandt. So unterscheidet die COACTIV-Studie von Baumert, Blum und Neubrand, die sich mit Fragen der Konzeptualisierung und Messung des fachspezifischen Professionswissens von Mathematiklehrkräften und möglicher Bezüge zur Leistungsentwicklung von Schülerinnen und Schülern befasst, ebenfalls zwischen fachdidaktischem Wissen und mathematischem Fachwissen (siehe u.a. Krauss et al., 2004, Brunner et al., 2006 sowie den Übersichtsartikel von Baumert & Kunter, 2006). Dabei umfasst fachdidaktisches Wissen in der COACTIV-Studie die Wissensfacetten „Verständlichmachen von mathematischen Inhalten", „mathematikbezogene Schülerkognitionen" und „kognitives Potential von Mathematikaufgaben", was eine starke Nähe zu den in *MT21* vorgenommenen Konzeptualisierungen aufweist (siehe auch Krauss et al., 2008).

Die Unterscheidung zwischen didaktischem Wissen (*pedagogical content knowledge*) und Fachwissen (*content knowledge*) wird in der Mathematikdidaktik seit Jahren intensiv diskutiert, insbesondere unter dem Aspekt, ob eine Separierung überhaupt möglich ist und wenn ja, wie diese zu konzeptualisieren sei. In ihrer grundlegenden Auseinandersetzung mit dem Ansatz von Shulman zeigen Graeber und Tirosh (2008) auf, dass viele Konzeptionen in einem gewissen Maße immer noch schwer fassbar sind und häufig eher mit Listen von Beispielen als theoretischen Überlegungen arbeiten.

Einen Ansatz zur Weiterentwicklung der Konzeption von Shulman hat in den USA die sog. Michigan-Gruppe um Loewenberg Ball und Bass (2003) entwickelt, die eine sehr differenzierte Klassifikation von Lehrerprofessionswissen vorschlagen und dabei als Wissensdomäne mathematisches Wissen für Lehren (*mathematical knowledge for teaching*) als zentral ansehen. Als zentrale Aufgabe von Lehrpersonen sehen sie die Entfaltung ("unpack") von komprimierten abstrakten mathematischen Ideen im Lehrprozess an (Hill, Loewenberg Ball & Schilling, 2008). Sie unterscheiden weitere Wissensdomänen wie alltägliches Fachwissen (*common content knowledge*) und spezialisiertes Fachwissen (*specialised content knowledge*), Wissen über Inhalte und Lernende (*knowledge of content and students*) sowie Wissen über Inhalte und Lehren (*knowledge of content and teaching*).

Krauss, Baumert und Blum (2008) stellen daher in ihrer zusammenfassenden Gegenüberstellung der in *MT21,* COACTIV und der Michigan-Gruppe um Ball und Bass entwickelten theoretischen Ansätze zum Lehrerprofessionswissen fest, dass alle drei Studien fachdidaktischem Wissen eine hohe Bedeutung zuweisen, dass die fachwissenschaftliche Komponente von der Michigan-Gruppe jedoch deutlich anders konzeptualisiert wird als von den Studien COACTIV und *MT21*. Es kann jedoch als Konsens innerhalb der internationalen mathematikdidaktischen Diskussion angesehen werden, dass die drei Wissensdomänen Fachwissen, fachdidaktisches Wissen und pädagogisches Wissen im Laufe der Ausbildung und späteren Berufspraxis quasi in Form eines Geflechts immer stärker miteinander integriert werden (Liljedahl et al., 2009).

1.3 Hypothesen

Vermutungen zu den Ursachen besonderer Stärken angehender GHR-Lehrkräfte bzw. angehender Gymnasiallehrkräfte lassen sich insbesondere aus drei Zugängen des im Rahmen der *MT21*-Studie entwickelten Kompetenzmodells ableiten:

1. aus der Modellierung von Kompetenzstufen des fachbezogenen Wissens, wie sie in Blömeke et al. (2008b) erfolgt ist;
2. aus den Analysen zur dimensionalen Struktur des fachbezogenen Wissens der angehenden Lehrkräfte (Blömeke et al., 2008c), sowie
3. aus den mehrebenenanalytischen Analysen zum Zusammenhang von Lerngelegenheiten in der Mathematiklehrerausbildung und erworbenem Wissen (Blömeke et al., 2008b).

Das Kompetenzstufen-Modell von *MT21* macht deutlich, dass für die Gesamtgruppe angehender Mathematiklehrkräfte der Sekundarstufe I zwei Item-Merkmale als besonders schwierigkeitsbestimmend angesehen werden können: die Anforderung, über schulmathematisches Wissen hinaus universitäres Mathematikwissens anwenden zu müssen, und das Erbringen vielschichtiger Verknüpfungsleistungen, vor allem in Form der Verknüpfung von mathematischem und mathematikdidaktischem Wissen, aber auch in Form der Verknüpfung von verbalen und grafischen Aussagen (Blömeke et al., 2008b, S. 110f). Über *universitäres Mathematikwissen* verfügt vermutlich lediglich die Gruppe der angehenden GyGS-Lehrkräfte. Ihre Ausbildung ist durch ein langes fachwissenschaftliches Studium geprägt, das in den Fakultäten für Mathematik stattfindet, deren Lehrveranstaltungen sie gemeinsam mit angehenden Diplom-Mathematikerinnen und -Mathematikern besuchen. Angehende GHR-Mathematiklehrkräfte haben deutlich weniger fachwissenschaftliche Lerngelegenheiten. Ihr fachbezogenes Studium ist dagegen häufig durch eine *Verknüpfung des mathematischen Wissens mit unterrichtlichen Anforderungen* der zukünftigen Schulstufe verknüpft – und damit eher auf *elementare Mathematik* fokussiert.

Vor diesem Hintergrund ist anzunehmen, dass sich Stärken der GyGS-Gruppe immer dann bemerkbar machen, wenn universitäres Mathematikwissen zur Bearbeitung der *MT21*-Items herangezogen werden muss, während sich Stärken der angehenden GHR-Lehrkräfte immer dann bemerkbar machen können, wenn es um elementarmathematische Aufgaben geht. Dabei verstehen wir im Folgenden unter elementarmathematischen Aufgaben solche, die sich auf Grundlagen der Zahlentheorie inklusive Bruchrechnung, Grundlagen funktionaler Zusammenhänge inklusive der zugehörigen Graphen sowie grundlegende geometrische Begriffe und Zusammenhänge beziehen.

Eine ähnlich klare Annahme lässt sich zu den Verknüpfungsleistungen formulieren. Die GHR-Ausbildung ist in dieser Hinsicht häufig integrativer angelegt als die GyGS-Ausbildung, indem fachliche und fachdidaktische Ausbildungskomponenten meist von denselben Personen – den Mathematikdidaktikerinnen bzw. -didaktikern der Universitäten bzw. den Fachseminarleiterinnen und -leitern der zweiten Phase – vermittelt werden. In der GyGS-Ausbildung gilt dies eher nicht: In der ersten Phase stellen unterrichtsbezogene Einbettungen in der fachwissenschaftlichen Ausbildung die Ausnahme dar, in der zweiten Phase wird das universitäre Mathematikwissen dann in der Regel nicht mehr thematisiert. Insofern ist anzunehmen, dass sich Stärken der GHR-Gruppe immer dann zeigen können, wenn Verküpfungsleistungen zwischen mathematischem und mathematikdidaktischem Wissen und damit prospektiv zwischen Mathematik und ihrer professionellen Anwendungspraxis gefordert sind.

In Ergänzung zur Bedeutung des mathematischen Niveaus und der Verknüpfungsleistung haben die dimensionalen Analysen in *MT21* zur Struktur des fachbezogenen Wissens auf dessen *Inhaltsabhängigkeit* aufmerksam gemacht. Das fünfdimensionale Modell, das das Wissen der angehenden Mathematiklehrkräfte in die Inhaltsgebiete Arithmetik, Algebra, Funktionen, Geometrie und Stochastik ausdifferenziert, weist die mit Abstand beste Modellanpassung an die vorliegenden Daten auf (Blömeke et al., 2008c, S. 76). Das heißt, dass in Bezug auf die Identifikation von Stärken angehender GHR- und GyGS-Lehrkräfte vermutlich auch inhaltliche Zuordnungen berücksichtigt werden müssen.

Betrachtet man die Lerngelegenheiten in den beiden Ausbildungsgängen unter diesem Gesichtspunkt, so ist festzustellen, dass die curricularen Schwerpunktsetzungen der Mathematiklehrerausbildung ähnlich eindeutige Annahmen zulassen wie zuvor. Arithmetik stellt das bedeutendste Themengebiet der GHR-Ausbildung dar. Gegenüber diesem steht Algebra bereits deutlich zurück. Empirisch orientierte Untersuchungen weisen zudem darauf hin, dass der Übergang von der Arithmetik zur Algebra auf der Schülerebene mit hohen kognitiven Barrieren verbunden ist (für eine Beschreibung einschlägiger Schwierigkeiten siehe die umfangreichen empirischen Untersuchungen in Malle, 1993). Funktionen als ein zentraler Leitbegriff der Mathematik strukturieren insbesondere mathematische Inhalte der Sekundarstufe II (Tietze, Klika & Wolpers, 1997). Sie werden in den universitären Lehrveranstaltungen für Lehramtsstudierende der Sekundarstufe II deutlich ausführlicher und intensiver behandelt als in der GHR-Ausbildung, und zwar sowohl im Grund- als auch im Hauptstudium. Die Rolle der Geometrie im Mathematikunterricht der Sekundarstufen ist dagegen seit Jahrzehnten weltweit durch einen Bedeutungsverlust gekennzeichnet. Dies gilt besonders für die Sekundarstufe II, wo sie ein Schattendasein führt. Für die Grundschule und die Sekundarstufe I ist die Notwendigkeit der Behandlung geometrischer Inhalte dagegen unstrittig, insbesondere in den letzten Jahren hat sich hier auch eine deutliche Stärkung entsprechender Inhalte ergeben (Franke, 2000; Krauthausen & Scherer, 2007). Dieser schulischen Schwerpunktsetzung entsprechen die Lehrpläne in der Mathematiklehrerausbildung, in der Inhalte der Geometrie für GHR-Lehramtsstudierende in der Regel verpflichtend sind, während sie für Lehramtsstudierende der Sekundarstufe II häufig nur ein fakultatives Angebot darstellen, welches dann in der Regel axiomatisch geprägt ist und nur wenig Bezüge zur Schulmathematik beinhaltet. Stochastik stellt ein Inhaltsgebiet dar, dem aufgrund seiner hohen Anwendungsrelevanz in Alltag und Wissenschaft zunehmendes Gewicht eingeräumt wird (KMK, 2003, Biehler & Hartung, 2006). Dies ist bisher deutlich stärker in der GHR-Ausbildung als in der GyGS-Ausbildung angekommen.

Zusammenfassend können vor dem Hintergrund dieser Analysen die folgenden beiden Hypothesen aufgestellt werden:

> *H1:* *Drei Kriterien bestimmen, bei welchen Items sich die besonderen Stärken angehender GHR- und GyGS-Lehrkräfte zeigen: das zu bewältigende mathematische Niveau (in der Abstufung elementare vs. nicht-*

elementare Mathematik), die zu erbringenden Verknüpfungsleistungen (in der Abstufung rein mathematische Anforderungen vs. Verknüpfung von mathematischem und mathematikdidaktischem Wissen) sowie das Inhaltsgebiet, das zu bearbeiten ist (Arithmetik, Geometrie und Stochastik vs. Algebra und Funktionen).

H2: Angehende GHR-Lehrkräfte zeichnen sich durch Stärken bei elementarmathematischen Aufgaben, Aufgaben mit mathematikdidaktischen Verknüpfungsleistungen sowie arithmetischen, geometrischen bzw. stochastischen Aufgaben aus, während angehende GyGS-Lehrkräfte Stärken bei Aufgaben, die über elementare Mathematik hinausgehen, bei rein mathematischen Aufgaben ohne unmittelbaren Unterrichtsbezug sowie bei algebraischen und funktionenbezogen Aufgaben aufweisen.

2 Untersuchungsdesign

2.1 Stichprobenziehung

Der vorliegende Beitrag beruht auf den Daten der deutschen *MT21*-Stichprobe. Die internationale Vergleichsstudie zur Wirksamkeit der Mathematiklehrerausbildung *MT21* fand im Jahr 2006 in sechs Ländern statt: in Bulgarien, Deutschland, Mexiko, Südkorea, Taiwan und den USA. Die Stichprobenziehung beruhte auf einem von den Ländern gemeinsam entwickelten Set an Kriterien. Im Vordergrund stand, dass vier zentrale Strukturmerkmale der Lehrerausbildung für die Sekundarstufe I die Auswahl der teilnehmenden Institutionen leiteten: der Typ der Ausbildungsinstitution (Universität vs. Lehrerbildungshochschule), ihre Größe, ihre Selektivität und ihre regionale Verteilung. In Deutschland wurden vor diesem Hintergrund vier Ausbildungsregionen in Nord-, Ost- und Westdeutschland ausgewählt, in denen jeweils alle Institutionen der ersten (Universitäten) und zweiten Phase (Studienseminare) der Lehrerausbildung an der Studie teilnahmen. Süddeutschland mit seiner anders gelagerten institutionellen Struktur (erste Phase in Baden-Württemberg: Pädagogische Hochschulen, zweite Phase in Bayern: Ausbildungsschulen) blieb außen vor, um die äußere Varianz nicht zu groß werden zu lassen und die Ausbildung auf der Aggregatebene noch adäquat beschreiben zu können.

Um Hypothesen zu Entwicklungsverläufen einer ersten empirischen Prüfung unterziehen zu können, wurden drei Kohorten von angehenden Mathematiklehrkräften der Sekundarstufe I für die Teilnahme ausgewählt: Studierende im Grundstudium, Studierende im Hauptstudium sowie Referendarinnen und Referendare. Die Stichprobe wurde in Bezug auf die beiden dominierenden Typen an Ausbildungen für die Sekundarstufe I stratifiziert: angehende Lehrkräfte für die Klassen 1 bis 10 (GHR) und angehende Lehrkräfte für die Klassen 5 bis 13 (GyGS).

Insgesamt nahmen 849 Personen an der Studie teil; davon befanden sich 368 in der ersten, 195 in der zweiten und 286 in der dritten Kohorte. Die Ausschöpfungsquote in Bezug auf die in den vier ausgewählten Regionen vorhandenen angehenden Lehrkräfte lag für die dritte – im vorliegenden Beitrag analysierte – Kohorte bei sehr guten 80 Prozent. Die Ergebnisse der Stichprobe wurden darüber hinaus anhand vorliegender Strukturdaten der Grundgesamtheit gewichtet, sodass ihnen Repräsentativität für die vier Ausbildungsregionen zugesprochen werden kann. Die Gewichtungsfaktoren liegen dabei wegen der geringen Stichprobenausfälle dicht um 1 (Spannweite: 0,792–1,188). Hier liegt also nur eine geringe Abweichung von Sollstruktur und Stichprobenstruktur vor, sodass ihre Qualität als sehr gut eingeschätzt werden kann (Gelman & Carlin, 2002).

2.2 Methodisches Vorgehen

Für die Erfassung der professionellen Kompetenz der angehenden Mathematiklehrkräfte wurde ein Test entwickelt, der fachwissenschaftliches, fachdidaktisches und fachübergreifendes Wissen erfasst. Darüber hinaus wurden Überzeugungen zur Natur der Mathematik, zum Lehren und Lernen von Mathematik sowie zur Lehrerrolle und zur Rolle der Schule in unserer Gesellschaft erhoben (ausführlich siehe Blömeke, Kaiser & Lehmann, 2008). Die Bearbeitung des Instruments, das mit Hilfe eines Rotationsdesigns in zwei Testheften präsentiert wurde, dauerte 90 Minuten. Die Mehrheit der Items weist ein Multiple-Choice-Format auf.

Für den vorliegenden Beitrag wurden die 72 Items des fachbezogenen *MT21*-Tests herangezogen und folgenden Analyseschritten unterzogen:

(1) In einem ersten Schritt wurden die Items von drei Expertinnen bzw. Experten detailliert auf die oben angeführten Kriterien hin untersucht, von denen ein Einfluss auf ihre Bewältigung durch angehende GHR- und GyGS-Lehrkräfte angenommen wurde. Bereits die Itementwicklung war theoriegeleitet erfolgt, indem u.a. die Zuordnung zur Mathematik bzw. Mathematikdidaktik sowie Inhaltsgebiete Berücksichtigung fanden. Darüber hinaus erfolgte für den vorliegenden Beitrag eine Klassifizierung jedes Items im Hinblick darauf, ob es aus deutscher Sicht elementares oder nicht-elementares mathematisches Wissen verlangte. Über eine solche Zuordnung konnte aus Gründen der Unvergleichbarkeit der Anforderungen im Studium sowie im Hinblick auf unterschiedliche didaktische Traditionen auf internationaler Ebene keine Einigung erzielt werden, sodass sie in *MT21* unterblieben war.

(2) Anschließend wurde eine Analyse auf differentielle Itemfunktionen hin vorgenommen, um drei Sets an Items bilden zu können: ein erstes Set, für das deutlich erkennbare Stärken auf Seiten der angehenden GHR-Lehrkräfte festzustellen sind (GHR+); ein zweites Set, für das deutlich erkennbare Stärken auf Seiten der angehenden GyGS-Lehrkräfte festzustellen sind (GyGS+); und ein drittes Set, in dem für keine der beiden Gruppen besondere Stärken zu erkennen sind (zum genauen Vorgehen siehe Abschnitt 2.2.1).

(3) In einem dritten Schritt wurde schließlich mittels Regressionsanalysen geprüft, inwieweit die ausgewählten drei Itemmerkmale elementares vs. nicht-elementares Mathematikwissen, rein mathematische vs. mathematikdidaktische Anforderungen sowie Inhaltsgebiet tatsächlich in der Lage sind, die Stärken der beiden Gruppen GHR und GyGS zu erklären (zum Vorgehen siehe Abschnitt 2.2.2).

2.2.1 Bestimmung der differentiellen Itemfunktionen

Die Analyse differentieller Itemfunktionen erfolgt mit Hilfe der *Item Response Theory* auf der Basis messfehlerfreier Itemparameter (Holland & Wainer, 1993; Millsap & Everson, 1993; Camilli & Shepard, 1994; siehe beispielsweise Klieme & Baumert, 2001). Der Analyse liegt eine eindimensionale Rasch-Skalierung der fachbezogenen Testitems zugrunde, die eine geringfügig schlechtere Anpassung an die Daten zeigt als eine zweidimensionale Modellierung, die zwischen mathematischen und mathematikdidaktischen Items unterscheidet, dafür aber eine höhere Reliablität aufweist (WLE-Reliabilität = .88; EAP-Reliabilität = .90).

Im ersten Schritt der Analysen werden der Mittelwert Θ des Testergebnisses für eine der beiden Gruppen auf 0 und die Varianz auf 1 fixiert, während beide Parameter für die zweite Gruppe geschätzt werden. Im nächsten Schritt erfolgt eine Testung der Hypothese, dass die Schwierigkeiten aller Items in beiden Gruppen dieselben sind (H_0: $b_{iGHR} = b_{iGYGS}$). Dabei werden bei der Prüfung eines Items jeweils alle anderen Items als Anker verwendet. Die Analyse wird mit *ConQuest 2.0* durchgeführt, das durch die Modellierung unterschiedlicher Facetten und deren Interaktion in der Lage ist, differentielle Itemfunktionen aufzudecken (Wu, Adams, Wilson & Haldane, 2007).

2.2.2 Überprüfung der Vorhersagekraft der Item-Merkmale

Basierend auf den Ergebnissen der DIF-Analyse werden Items mit signifikanten Abweichungen entweder dem GHR+- oder dem GYGS+-Set zugewiesen und qualitativ auf Gemeinsamkeiten und Unterschiede hinsichtlich ihres mathematischen Niveaus, der erforderlichen Verknüpfungsleistungen und ihrer inhaltlichen Ausrichtung untersucht. Anschließend erfolgt eine quantitative Überprüfung der Vorhersagekraft dieser Merkmale für die Zuordnung zu einem der Sets über einfaktorielle und multiple Hypothesenprüfungen. Die einfaktorielle Prüfung greift auf Kreuztabellierungen zurück und macht Aussagen dazu, inwieweit jedes der drei Merkmale – elementare vs. nicht-elementare Mathematik, rein mathematische vs. mathematikdidaktische Anforderungen und GHR-typische vs. GyGS-typische Ausbildungsinhalte – separat die Zuordnung der Items zu den Sets vorhersagen kann. Als Zusammenhangsmaß wird der Kontingenzkoeffizient verwendet, eine Signifikanzprüfung findet über Chi-Quadrat-Tests statt.

Die multiple Hypothesenprüfung basiert auf einer multinomialen logistischen Regressionsanalyse, mit der die Vorhersagekraft der drei Item-Merkmale geschätzt werden kann, wenn sie *gleichzeitig* berücksichtigt werden. Auf diese Weise kann die Bedeutsamkeit der drei Merkmale gegeneinander abgewogen werden. Die Gültigkeit des Gesamtmodells wird anhand des Pearson-Goodness-of-Fit-Index beurteilt. Bei Modellgültigkeitstests formuliert die Nullhypothese die Gültigkeit eines Modells, sodass ein Signifikanztest die Überschreitungswahrscheinlichkeit angibt. Hohe Werte deuten damit darauf hin, dass ein Modell akzeptiert werden kann. Anschließend wird die Relevanz des Modells anhand des Pseudo-R^2 nach Nagelkere betrachtet. Diese Kennziffer kann Werte zwischen 0 und 1 annehmen, wobei höhere Werte auf eine größere Varianzaufklärung hinweisen. Schließlich kann die Güte des Modells anhand seiner Klassifikationsleistung eingeschätzt werden, also anhand des Anteils korrekter Zuordnungen aufgrund der drei Item-Merkmale.

3 Ergebnisse

Im Folgenden werden zunächst die Ergebnisse der Analyse auf differentielle Itemfunktionen dokumentiert (Abschnitt 3.1), bevor eine qualitative Analyse der beiden Item-Sets GHR+ und GyGS+ erfolgt, die besondere Stärken der GHR- bzw. der GyGS-Lehrkräfte anzeigen (3.2). Die Copyright-Bestimmungen lassen es zu, exemplarisch Aufgaben aus dem *MT21*-Test abzubilden, die das Spektrum der untersuchten Merkmale abdecken. Darauf hinzuweisen ist allerdings, dass alle *MT21*-Items, die sich besonders gut bewährt haben, in das Eigentum der IEA als Organisatorin der Folgestudie TEDS-M übergegangen sind. Eine zwangsläufige Folge ist, dass die zur Illustration wiedergegebenen Beispiele Items sind, die zwar in *MT21* Anwendung fanden, aber aus verschiedenen Gründen nicht weiter verwendet wurden und im Folgenden ggf. leicht modifiziert wiedergegeben sind. Nicht freigegebene Items können nur rephrasierend wiedergegeben werden.

Die Tatsache, dass nicht alle Items der wissenschaftlichen Öffentlichkeit zur Verfügung stehen, ist kritikwürdig, da damit die Nachvollziehbarkeit der Ergebnisse eingeschränkt ist; im Hinblick auf die Möglichkeit von Folgestudien ist eine eingeschränkte Freigabe von Items allerdings unabdingbar. Um die Folgen der Copyright-Bestimmungen abzumildern, werden möglichst viele Items im Detail beschrieben, so dass ihre Kernideen deutlich werden.

3.1 Differentielle Itemfunktionen

Die Analyse zeigt in einem ersten Schritt, dass angehenden Gymnasial- und Gesamtschullehrkräften die *MT21*-Items im Mittel deutlich leichter fallen als angehenden Grund-, Haupt- und Realschullehrkräften. Die Itemschwierigkeiten liegen für die erste Gruppe um .972 Logits unter denen für die zweite Gruppe (S.E. = .038). Auch wenn dieser Unterschied hochsignifikant ist ($\chi^2_{[1]} = 134.35$, $p < .001$) und eine latente Regressionsanalyse mit dem Ausbildungsgang als Prädiktor das Ergebnis nachdrücklich bestätigt, zeigt dieses noch keine differentiellen Itemfunktionen an, sondern weist nur auf den generellen Leistungsvorsprung der GyGS-Lehrkräfte hin. Erst ein Test auf Wechselwirkungen zwischen Ausbildungsgang und Itemschwierigkeiten macht deutlich, dass Letztere für die beiden Gruppen nicht gleich sind: $\chi^2_{[70]} = 364.85$, $p < .001$. Insgesamt unterscheiden sich die Itemschwierigkeiten von 23 Items signifikant, davon weisen 12 Items auf besondere Stärken von GHR- und 11 Items auf besondere Stärken von GYGS-Lehrkräften hin. Deren Eigenschaften werden im Folgenden detailliert analysiert.

3.2 Qualitative Analyse der Item-Sets

3.2.1 Merkmale der Items GHR+

Von den 12 Items, die auf besondere Stärken angehender Grund-, Haupt- und Realschullehrkräfte hinweisen, geht mit einem Item eine Aufgabe[2] in das GHR+-Set ein, bei der

[2] Ein „Item“ stellt die kleinste Einheit unseres Tests dar. Mit ihr kann ein Punkt erreicht werden. Als „Aufgabe“ bezeichnen wir dagegen, wenn sich mehrere Items auf einen Aufgabenstamm

eingeschätzt werden muss, wie viel Prozent der deutschen Schülerinnen und Schüler der Jahrgangsstufe 8 im Rahmen der TIMS-Studie in der Lage waren, bestimmte Aufgaben zu lösen. Bei dem Item geht es um elementargeometrische Sachverhalte aus dem Bereich Symmetrie. Um die Aufgabe lösen zu können, müssen zum einen auf der mathematikdidaktischen Ebene Schülerleistungen eingeordnet werden, zum anderen ist ein Verständnis elementarer Mathematik im Inhaltsgebiet Geometrie notwendig.

Ein Item entstammt einer Aufgabe aus dem Inhaltsgebiet der Stochastik. Mathematisch muss hier das elementare Konzept des Durchschnitts auf eine reale Situation – Anzahl an Personen in Autos – angewendet werden, d.h. welche Möglichkeiten es bei bekannter Anzahl von Autos es gibt, diese zu besetzen. Vom mathematikdidaktischen Bezug her muss die Korrektheit mehrerer authentischer Schülerlösungen beurteilt werden.

---

beziehen. Die in Abbildung 1 gezeigte Aufgabe besteht beispielsweise aus drei Items, während die in den Abbildungen 4 und 5 gezeigten Beispiele aus jeweils einem Item bestehen.

I61. Alle Schüler(innen) des achten Jahrgangs einer Schule nahmen an einem Test teil, der auf einer Skala mit 100 Punkten bewertet wurde. Die Tabelle zeigt die Ergebnisse.

Punktzahl	***Anzahl von Schüler(innen)***
Weniger als 31	26
31–40	18
41–50	16
51–60	20
61–70	21
71–80	19
81–90	11
91–100	5
	Insgesamt:136

Der Direktor der Schule ordnet die Schüler(innen) vier Bereichen zu:

Bereich 3 – umfasst Diejenigen, die mindestens 81 Punkte erreicht haben
Bereich 2 – umfasst Diejenigen, die mindestens 51 Punkte, aber weniger als 81 Punkte erreicht haben
Bereich 1 – umfasst Diejenigen, die mindestens 31 Punkte. aber weniger als 51 Punkte erreicht haben
Bereich 0 – umfasst Diejenigen, die weniger als 31 Punkte erreicht haben

Sie haben drei Aufgaben:

Aufgabe 1: Zeigen Sie, dass *Bereich 2* mehr Schüler(innen) enthält als jeder der anderen Bereiche.
Aufgabe 2: Zeigen Sie, dass die Schule den Standard, dass "weniger als 20% aller Schüler(innen) zu *Bereich 0* gehören" erfüllt.
Aufgabe 3: Zeigen Sie, dass die besten 20% der Schüler(innen) aus mehr als einem Bereich kommen.

Welche der folgenden Aktivitäten ist die angemessenste für jede der obigen Aufgaben?

Kreuzen Sie ***nur ein*** *Kästchen für* ***jede Aufgabe*** *an.*

	Aufg. 1	Aufg. 2	Aufg. 3
1. Ich würde ein Balkendiagramm für alle 8 Zeilen in der Tabelle machen.	☐	☐	☐
2. Ich würde ein Balkendiagramm für die 4 vom Direktor definierten Ebenen machen.	☐	☐	☐
3. Ich würde aufzeigen, dass $\frac{26}{136}$ weniger als 20% ist.	☐	☐	☐
4. Ich würde zeigen, dass *Bereich 0* weniger Schüler(innen) enthält als jeder der anderen Bereiche.	☐	☐	☐
5. Ich würde den Prozentsatz von Schüler(innen) in jedem Bereich berechnen.	☐	☐	☐
6. Ich würde den Prozentsatz von Schüler(innen) in Bereich 3 berechnen.	☐	☐	☐

Abb. 1: Beispiel-Aufgabe aus dem MT21*-Test mit elementaren mathematischen Anforderungen*

Ein weiteres Item aus dem Gebiet der Stochastik verlangt von den angehenden Lehrkräften Vorschläge, wie numerische Ergebnisse graphisch dargestellt werden können (siehe Abb. 1). Die mathematischen Anforderungen sind nur elementar, darüber hinaus stellt genau diese schülernahe Variation von Repräsentationsformen eine Stärke der GHR-Ausbildung dar, auch wenn es sich in diesem Fall um ein rein mathematisches Item handelt, das keine mathematikdidaktischen Verknüpfungen erfordert.

Eine Aufgabe aus dem Bereich Funktionen verlangt, eine graphische Schülerlösung zu beurteilen. Hieraus stammt ein Item. Mathematisch muss die mittels eines Graphen gegebene Beziehung zwischen der Geschwindigkeit eines Radfahrers in Abhängigkeit von der Zeit in reale Bewegungsverläufe übersetzt werden. Vom didaktischen Bezug her müssen Lösungen beurteilt werden, die bekannte Schülerfehlvorstellungen zur Verwechslung des Graphen mit dem tatsächlichen Weg wiedergeben (Janvier, 1981).

Fünf Items stammen aus dem Inhaltsbereich der Geometrie. In einer Aufgabe, deren mathematischen Anforderungen nicht mehr eindeutig als elementar bezeichnet werden können, muss die Korrektheit von Eigenschaften bestimmt werden, die sich aus Mittelsenkrechten in Dreiecken ableiten lassen. Abb. 2 visualisiert die Itemfunktion zugunsten der GHR-Lehrkräfte. Die Zuordnung dieses Items zu dem Item-Set GHR+ ergibt sich vermutlich aus der starken Stellung von Geometrie in deren universitären Curriculum.

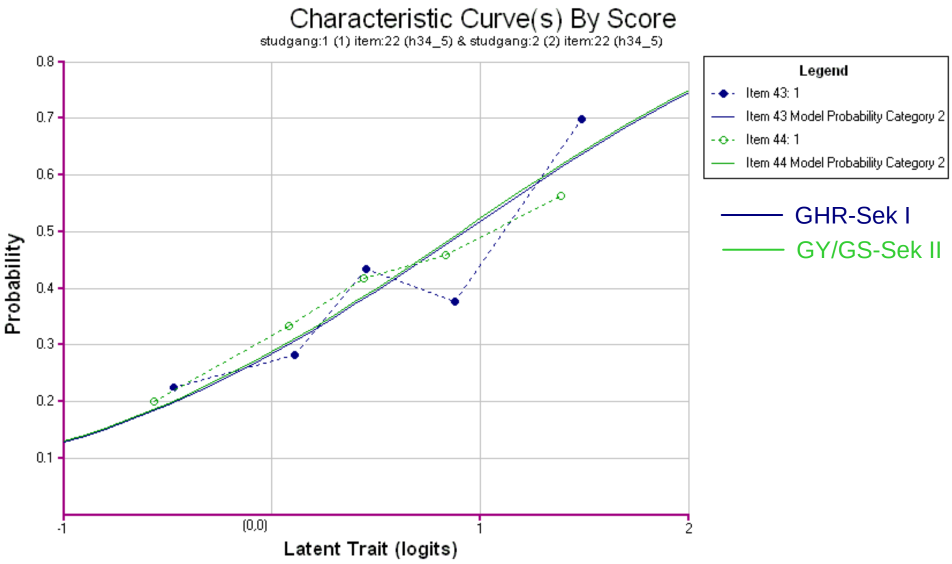


Abb. 2: Beispiel-Item mit differentieller Itemfunktion zugunsten der GHR-Lehrkräfte

Ebenfalls dem Inhaltsgebiet der Geometrie entstammt ein weiteres Item, in dem Skizzen zur Visualisierung eines elementargeometrischen Zusammenhangs gegeben sind. Anschließend muss die Frage beantwortet werden, ob ein gegebener Satz, der elementare geometrische Zusammenhänge beschreibt, zutreffend ist. Einen didaktischen Bezug enthält dieses Item nicht.

Zusammenfassend kann festgehalten werden, dass sich im Item-Set GHR+ überwiegend Items finden, bei denen die didaktische Analyse unterrichtlicher Lernprozesse im Vordergrund steht und die dabei elementare mathematische Inhalte zur Grundlage haben, überwiegend aus den Inhaltsgebieten der Geometrie und der Stochastik. Bei der didaktischen Analyse handelt es sich fast immer um die Einschätzung von Schülerleistungen oder um auf diese bezogene Rückmeldungen bzw. Bewertungen. Die Schülerleistungen beziehen sich auf die Bearbeitung von Aufgaben, die elementares mathematisches Wissen erfordern, weswegen die Analyse der Schülerlösungen von den angehenden Lehrkräften auch nur elementarmathematische Kenntnisse erfordert.

3.2.2 Merkmale der Items GyGS+

Vier Items in diesem Set stammen aus einer Aufgabe, die klassisch universitätsmathematisches Wissen repräsentiert. Sie ist dem Inhaltsgebiet Funktionen zuzuordnen und weist keinen didaktischen Bezug auf. Es geht um die Definition des Konzepts der Stetigkeit. Hier sind sowohl korrekte als auch fehlerhafte Definitionen gegeben, die auf bekannten Fehlvorstellungen zur Verwechslung von ε und δ beruhen (Tietze, Klika & Wolpers, 1997). Abb. 3 visualisert die differentielle Itemfunktion zugunsten der GyGS-Lehrkräfte.

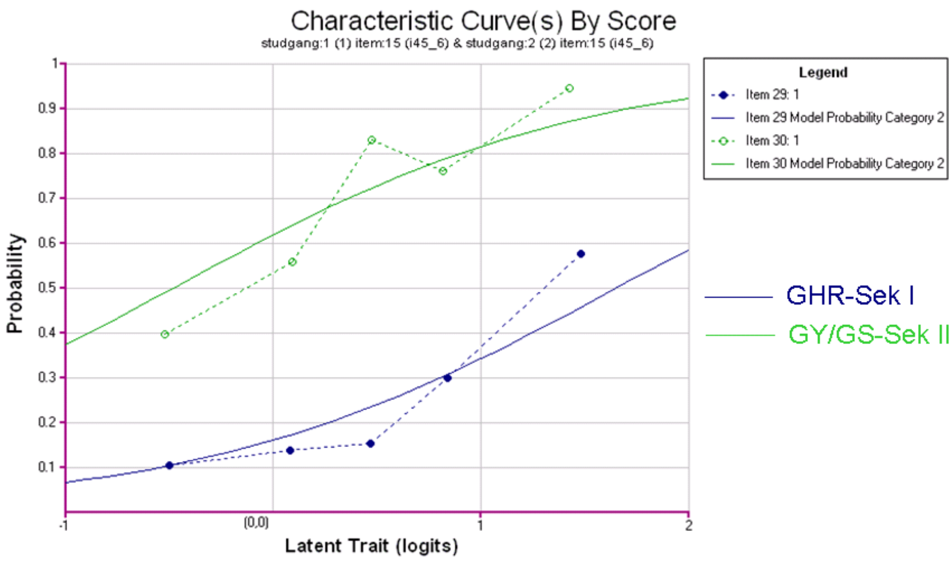


Abb. 3: Beispiel-Item mit differentieller Itemfunktion zugunsten der GYGS-Lehrkräfte

Ein Item entstammt dem Inhaltsbereich der Algebra. Hier muss von der fachmathematischen Perspektive her die Anzahl der Lösungen einer quadratischen Gleichung bestimmt werden (siehe Abb. 4). Dabei handelt es sich zwar um schulbezogene Mathematik, die aber nicht der elementaren Mathematik zugeordnet werden kann, da in der Aufgabe mit Parametern gearbeitet werden muss. Die Diskriminante muss zudem abstrakt, also mit Parameterbezug, analysiert werden. Ein didaktischer Bezug liegt nicht vor.

Ebenfalls dem Inhaltsbereich der Algebra entstammt ein Item des GyGS+-Sets, in dem zu einer gegebenen logarithmischen Gleichung die äquivalente Exponentialgleichung gefunden werden soll (siehe Abb. 5). Der mathematische Bezug ist hier ebenfalls nicht elementar, wenn auch erneut deutlich schulbezogen. Bei dieser Aufgabe liegt ebenfalls kein didaktischer Bezug vor.

B47. Wenn $a > 0$ ist, wie viele verschiedene reelle Lösungen hat dann die Gleichung $x^2 + x - a = 0$?

Kreuzen Sie **ein** *Kästchen an.*

1. 0 .... ☐
2. 1 .... ☐
3. 2 .... ☐
4. Unendlich viele. .... ☐
5. Die Anzahl reeller Lösungen hängt vom Wert von a ab. .... ☐

Abb. 4: Mathematisches Beispielitem aus dem MT21-*Test*

Zwei weitere Items können ebenfalls dem Inhaltsbereich der Algebra zugeordnet werden, dem damit – entsprechend seiner Stellung in der Ausbildung – eine starke Bedeutung für die Identifikation GYGS-Stärken zukommt. Die Aufgabe behandelt die Lösungen verschiedener Gleichungen, die jeweils den passenden kleinsten Zahlenbereichen zugeordnet werden müssen. Auch dies lässt sich der nicht-elementaren Mathematik zuordnen. Darüber hinaus ist die Fragestellung typisch für universitäre Mathematik, weil nach dem kleinsten Zahlenbereich gefragt wird, aus dem jeweils alle Lösungen der betreffenden Gleichung stammen. Diese Art von Fragen wird eher im fachmathematischen Studium von angehenden GyGS-Lehrerinnen und -Lehrern diskutiert als im fachmathematischen Studium der angehenden GHR-Lehrerinnen und -Lehrer. Wiederum ist in der Aufgabe zudem keinerlei mathematikdidaktischer Bezug gegeben. Zudem ist festzuhalten, dass sich die Items dieser Aufgabe deutlich hinsichtlich ihres mathematischen Schwierigkeitsgrades unterscheiden. Die beiden Items, die dem GYGS-Set zugeordent werden können, sind deutlich anspruchsvoller als die übrigen Items, da nur zur Lösung dieser Items Kenntnisse über reelle bzw. komplexe Zahlen nötig sind. Diese werden jedoch, wenn überhaupt nur in der Oberstufe vermittelt, viele Studierenden lernen komplexe Zahlen auch erst im Rahmen ihres Mathematikstudiums kennen.

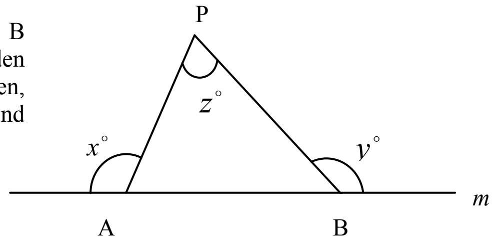
V01. Welcher der folgenden Ausdrücke ist äquivalent zu $\log_a b = c$?

Kreuzen Sie **ein** *Kästchen an.*

1. $a^b = c$	☐
2. $a^c = b$	☐
3. $b^a = c$	☐
4. $c^a = b$	☐

Abb. 5: Mathematisches Beispielitem aus dem MT21-*Test*

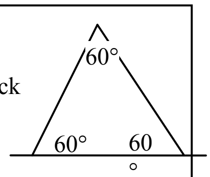
K4. In der Zeichnung sind A und B zwei feste Punkte auf der Geraden m. Punkt P kann bewegt werden, aber bleibt oberhalb von m und bleibt verbunden mit A und B.



Anna, Bruno, Charlotte und Daniel diskutieren, ob die folgende Aussage wahr ist:
$x° + y°$ ist gleichgroß wie $180° + z°$.

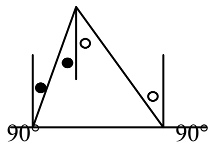
Annas Antwort
Ich habe die Winkel in der Zeichnung gemessen und herausgefunden, dass der Winkel x 110° beträgt. Winkel y ist 125° und Winkel z ist 55°. $110° + 125° = 235°$ und $180° + 55° = 235°$. *Also sagt Anna, dass es wahr ist.*

Brunos Antwort
Ich kann Punkt P so verschieben, dass das Dreieck gleichseitig ist und seine Winkel 60° sind.



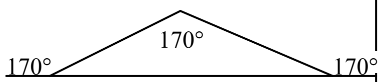
Damit ist x 120° und y ist 120°. $120° + 120°$ ist dasselbe wie $180° + 60°$. *Also sagt Bruno, dass es wahr ist.*

Charlottes Antwort
Ich habe drei parallele Linien gezeichnet, die zur Basis rechtwinklig sind.



Die beiden mit einem ● markierten Winkel sind gleichgroß und die beiden mit einem ○ markierten sind gleichgroß.
Winkel x ist 90° + ● und Winkel y ist 90° + ○. Also $x + y$ ist 180° + ● + ○, was $180° + z°$ ist.
Also sagt Charlotte, dass es wahr ist.

Daniels Antwort
Ich habe an eine Zeichnung gedacht, wo die Winkel x, y und z alle 170° sind.



In meiner Zeichnung sind $x° + y°$ nicht gleich zu $180° + z°$. *Also sagt Daniel, dass es nicht wahr ist.*

Kreuzen Sie die am **meisten angemessene** *Antwort für* **jede(n) Schüler(in)** *an.*

	Anna	**Bruno**	**Charlotte**	**Daniel**
1. Deine Argumentation enthält einen Fehler. Denke nochmals darüber nach.	☐	☐	☐	☐
2. Es reicht nicht aus, die Aussage an einem Beispiel zu überprüfen. Denke nochmals darüber nach.	☐	☐	☐	☐
3. Ausgezeichnet! Dies ist ein überzeugender Beweis.	☐	☐	☐	☐

Abb. 6: Mathematikdidaktisches Beispielitem aus dem MT21-*Test*

Ein Item des GyGS+-Sets entstammt schließlich dem Inhaltsbereich der Geometrie (siehe Abb. 6). Hier geht es darum, unterschiedliche Beweise zu Eigenschaften geometrischer Figuren auf ihre Angemessenheit hin einzuschätzen. Mathematisch gehen die Anforderungen über Methoden der elementaren Mathematik hinaus. Erneut ist der äußerliche didaktische Bezug der Aufgabe fachmathematischer Natur, da das Durchführen, Analysieren und Nachvollziehen von formalen Beweisen ein typischer Bestandteil des universitärmathematischen Studiums ist.

Zusammenfassend lässt sich zu dem Set, in dem besonders die angehenden GyGS-Lehrkräfte Stärken zeigen, festhalten, dass die Aufgaben überwiegend nicht-elementare Mathematik behandeln und keinen bzw. nur einen sehr fachnahen didaktischen Bezug aufweisen. Sofern ein didaktischer Bezug gegeben ist, kann dieser fast vollständig auf der Basis fachmathematischer Kenntnisse bewältigt werden und es steht nicht die Beurteilung von Schüleraktivitäten im Vordergrund, sondern ein Lehrbezug. Es sei nochmals darauf hingewiesen, dass von uns als "nicht-elementare Mathematik" bezeichnete Wissensstände auch, aber nicht ausschließlich universitäre Mathematik umfassen.

3.2.3 Merkmale einer Aufgabe, aus der Items zu beiden Sets gehören

Eine interessante Perspektive auf die jeweiligen Stärken der beiden untersuchten Teilpopulationen von angehenden Mathematiklehrerinnen und -lehrern der Sekundarstufe I liefert eine Aufgabe, aus der einzelne Items dem GHR+- und andere dem GyGS+-Set zugeordnet werden können. Sie behandelt die Frage, welche mathematischen Begriffe und Methoden man noch in der Sekundarstufe I vermitteln kann, wenn der Begriff der Quadratwurzel erst in der Oberstufe eingeführt werden würde. Die gegebenen Begriffe und Methoden sprechen unterschiedliche Inhaltsgebiete und unterschiedliche mathematische Niveaus an, die offensichtlich spezifische Stärken angehender GHR- und GyGS-Lehrkräfte gut zur Geltung kommen lassen. Die GHR-Studierenden zeigen Stärken im elementargeometrischen Bereich, einem mathematischen Teilgebiet also, das häufig in der GHR-Ausbildung verpflichtend, in der GyGS-Ausbildung jedoch nur ein mögliches Wahlgebiet unter mehreren ist. Weiterhin wird typischerweise in der GHR-Ausbildung im Hinblick auf die spätere berufliche Arbeit ein starker mathematischer Fokus auf eher basale arithmetische Zusammenhänge gelegt, die dann vertieft betrachtet werden, während in der GyGS-Ausbildung abstraktere Strukturen im Mittepunkt stehen. Dies wird in der Zuordnung der in der Aufgabe gegebenen Begriffe und Strukturen zu den verschiedenen Sets ebenfalls deutlich. Die Stärken der angehenden GyGS-Lehrerinnen und -Lehrer liegen in denjenigen Bereichen der Arithmetik, die zwar nur unwesentlich weniger elementar sind, als die dem GHR-Set zugeordneten Bereich, aber oft als Hilfsmittel zur Analyse abstrakterer Strukturen herangezogen werden. Schon eine leichte Variation des mathematischen Kontextes kann also die Zuordnung eines Items zu einem anderen Item-Set zur Folge haben.

3.3 Überprüfung der formulierten Hypothesen

3.3.1 Einfaktorielle Hypothesenprüfung

Die qualitativen Analysen der Items in den Sets GHR+ und GyGS+ stützen die eingangs aus den *MT21*-Ergebnissen abgeleiteten Hypothesen, dass drei Merkmale zentral für die Bestimmung der Itemschwierigkeiten sind und dass es sich hierbei um die Merkmale elementares vs. nicht-elementares mathematisches Wissen, fachmathematische vs. mathematikdidaktische Anforderungen und GHR- vs. GyGS-spezifische Inhaltsgebiete handelt. Dabei scheinen das erste und das letzte Merkmal von größerer Bedeutung zu sein als die Verknüpfung von Mathematik und Mathematikdidaktik. Im Folgenden werden die Hypothesen statistisch überprüft. Im ersten Schritt erfolgt eine Prüfung, inwieweit jedes der drei Merkmale die Zuordnung der Items zu einem der beiden Item-Sets bzw. zu dem Set mit Items ohne auffällige Abweichungen vorhersagen kann.

30 der 71 *MT21*-Items – ein Item ist eine Konstante, d.h. es weist eine Lösungshäufigkeit von 100% auf und wurde daher aus den Analysen entfernt – erfordern elementarmathematisches Wissen, während 41 Items Wissen benötigen, das darüber hinausgeht. Dieses Merkmal trägt signifikant dazu bei, Stärken angehender GHR- bzw. angehender GyGS-Lehrkräfte vorherzusagen ($\chi^2_{[2]} = 11.23$, $p < .01$). Deutlich mehr Items aus dem Bereich der nicht-elementaren Mathematik als bei einer Gleichverteilung erwartet finden sich im GYGS+-Item-Set (siehe Tab. 1). Der Zusammenhang zwischen mathematischem Niveau und Gruppenzugehörigkeit der Items ist substanziell ($r_C = 0.37$).

		Elementare Mathematik	**Nicht-elementare Mathematik**
GHR+	Beobachtet	4	8
	Erwartet	5	7
Mittelgruppe	Beobachtet	26	22
	Erwartet	20	28
GyGS+	Beobachtet	0	11
	Erwartet	5	6

Tab. 1: Zugehörigkeit zu den Items-Sets nach mathematischem Wissensniveau

Mathematikdidaktische Anforderungen werden in 32 der 71 Items gestellt, während 39 Items auf rein mathematische Anforderungen beschränkt sind. Auch mit diesem Merkmal lässt sich die Zuordnung eines Items zu einer der drei Gruppen – GHR+-Items, die Stärken angehender GHR-Lehrkräfte signalisieren, GyGS+-Items, die Stärken angehender GyGS-Lehrkräfte anzeigen, sowie eine Mittelgruppe – signifikant vorhersagen, allerdings nicht so deutlich wie mit dem vorhergehenden Merkmal ($\chi^2_{[2]} = 8.84$, $p < .05$; siehe Tab. 2). Der Zusammenhang zwischen der zu erbringenden Verknüpfungsleistung und der Gruppenzugehörigkeit der jeweiligen Items weist etwa dieselbe Stärke auf wie beim ersten Merkmal ($r_C = 0.33$). In der Item-Gruppe, die Stärken der angehenden Gymnasial- und Gesamtschullehrkräfte andeutet, sind substanziell mehr Items zu finden, die rein mathematisch ausgerichtet sind und keine Verknüpfungsleistungen erfordern.

		Mathematik	Mathematikdidaktik
GHR+	Beobachtet	8	4
	Erwartet	7	5
Mittelgruppe	Beobachtet	21	27
	Erwartet	26	22
GyGS+	Beobachtet	10	1
	Erwartet	6	5

Tab. 2: Zugehörigkeit zu den Items-Sets nach Verknüpfungsleistungen

Konzepte aus den mathematischen Gebieten Arithmetik, Geometrie und Stochastik sind als GHR-typische Ausbildungsinhalte definiert worden. 36 der 71 Items entstammen diesen drei Inhaltsgebieten. 35 Items gehören zu den beiden als GyGS-typische Ausbildungsinhalte definierten Inhaltsgebieten Algebra und Funktionen. Das Merkmal mathematisches Themengebiet wird ebenfalls signifikant, wenn es um die Vorhersage von Stärken der angehenden GHR- bzw. GyGS-Lehrkräfte geht ($\chi^2_{[2]} = 10.69$, $p < .01$; siehe Tab. 3). Der Zusammenhang von Inhaltsgebieten und Itemzugehörigkeit liegt erneut im Bereich mittlerer Stärke ($r_C = 0.36$). Deutlich mehr Items als erwartet gehören – wie vermutet – zu den Inhaltsgebieten Arithmetik, Geometrie und Stochastik in der Gruppe der GHR+-Items. Noch auffälliger ist die Zugehörigkeit zu Algebra und Funktionen im GYGS+-Set.

		Algebra und Funktionen	Arithmetik, Geometrie und Stochastik
GHR+	Beobachtet	3	9
	Erwartet	6	6
Mittelgruppe	Beobachtet	22	26
	Erwartet	24	24
GyGS+	Beobachtet	10	1
	Erwartet	5	6

Tab. 3: Zugehörigkeit zu den Items-Sets nach Inhaltsgebiet

3.3.2 Multiple Hypothesenprüfung

Im letzten Schritt der Itemanalysen geht es darum, die drei bisher getrennt betrachteten Merkmale mathematisches Niveau, Verknüpfungsleitsungen, und Inhaltsgbiet gleichzeitig einzuführen, um zu überprüfen, ob allen auch dann noch bedeutsames Gewicht in der Vorhersage der Gruppenzugehörigkeit eines Items zukommt oder ob ein Merkmal ggf. überflüssig ist. Die qualitativen Analysen in Abschnitt 3.2 haben darauf aufmerksam gemacht, dass das Auftreten von Items, die elementarmathematisches Wissen erfordern, im *MT21*-Test häufig an eine didaktische Anforderung geknüpft ist, z.B. in Form von didaktisch motivierten Rückmeldungen zu Schülerlösungen. Insofern ist denkbar, dass durch diese Form der Testkonstruktion eines dieser beiden Itemmerkmale zur Vorhersage der Klassifizierung in GHR+, Mittelgruppe bzw. GyGS+ hinreichend ist. Zur Prüfung wird eine multinomiale logistische Regressionsanalyse durchgeführt, mit der der Zusammenhang zwischen mehreren unabhängigen (mathematisches Niveau, Verknüpfungsleistung sowie Inhaltsgebiet, jeweils in dichotomer Ausprägung) und einer abhängigen Variablen mit drei Kategorien auf Nominalskalen-Niveau (GHR+, Mittelgruppe, GyGS+) geschätzt werden kann. Die Zahl der Items ist knapp hinreichend, um dieses Verfahren anwenden zu können (Backhaus et al., 2006; Urban, 1993).

Bevor auf die Bedeutung der drei Itemmerkmale eingegangen wird, soll zunächst die Gültigkeit des vorgeschlagenen Modells, seine Relevanz und seine Güte in Form der Klassifikationsleistung beurteilt werden. Das Modell weist eine gute Anpassung an die Daten auf ($\chi^2_{[8]} = 10.4$, $p = .24$). Es kann daher als *gültig* akzeptiert werden. Durch die drei Regressoren werden 42 Prozent der Varianz in der Itemzuordnung erklärt, was als *relevanter* Anteil angesehen werden kann. Die *prädikative Effizienz* des Modells liegt bei 69.0%, das heißt, dass mehr als zwei Drittel der Itemklassifikationen mit Hilfe der drei Merkmale korrekt vorhergesagt werden können. Dabei liegt die Vorhersagequalität für die Mittelgruppe und das GyGS+-Item-Set deutlich höher als die für das GHR+-Item-Set.

Wie nach den qualitativen Analysen erwartet, aber nur partiell der Ausgangshypothese entsprechend, beeinflussen lediglich zwei der drei Itemmerkmale die Zuordnung eines Items zu einem der drei Sets signifikant. In H1 war formuliert worden, dass alle drei Merkmale relevant hierfür sind.

Die Wahrscheinlichkeit, dass ein Item dem mittleren Set bzw. gar dem GHR+-Item-Set zugeordnet werden kann, sinkt nahezu auf Null, falls die anzuwendende Mathematik elementares Niveau überschreitet ($\chi 2_{[2]} = 9.09$, $p < .05$). Das mathematische Niveau stellt damit ein mächtiges Item-Merkmal dar. Ein weiteres bedeutsames Merkmal ist das Inhaltsgebiet: Stammt ein Item aus den Gebieten Algebra oder Funktionen, liegt die Wahrscheinlichkeit, dass es dem GHR+- oder dem mittleren Item-Set anstelle des GyGS+-Item-Sets zugeordnet werden kann, ebenfalls nahe 0 ($\chi 2_{[2]} = 12.65$, $p < .01$). Nicht signifikant wird dagegen die gefordertete Verknüpfungsleistung. Die zusätzliche Erklärungskraft über die beiden anderen Merkmale hinaus ist entgegen unserer Ausgangsannahme gering.

Die fast gleichlautenden Ergebnisse für die Zuordnung eines Items zum GHR+-bzw. zum mittleren Item-Set machen darauf aufmerksam, dass die zentrale Trennlinie offensichtlich zwischen Letzterem und dem GyGS+-Item-Set verläuft. Tatsächlich weist eine binär logistische Regressionsanalyse mit nur noch dem GyGS+-Item-Set auf der einen Seite und allen übrigen Items auf der anderen Seite eine noch höhere Relevanz auf als das zunächst getestete (Nagelke-$R^2 = .52$) und es klassifiziert 85 Prozent der Items korrekt.

4 Zusammenfassung und Diskussion

Im vorliegenden Beitrag wurden die fachbezogenen Ergebnisse angehender Grund-, Haupt- und Realschullehrkräfte (GHR) sowie angehender Gymnasial- und Gesamtschullehrkräfte (GyGS) im *MT21*-Test auf ihre besonderen Stärken hin untersucht. Zu diesem Zweck wurden die differenziellen Itemfunktionen betrachtet, das heißt, es wurden zwei Sets an Items gebildet (GHR+ und GyGS+), die sich durch deutlich größere relative Lösungshäufigkeiten im Vergleich zu den mittleren Testergebnissen der beiden Gruppen von angehenden Lehrkräften auszeichnen. Solche subgruppenspezifischen Abweichungen in der Itemschwierigkeit heben die mittlere Differenz in der Kompetenzeinschätzung nicht auf, sie erlauben aber ein besseres Verständnis spezifischer Stärken der untersuchten Ausbildungsgänge.

Unsere Annahme war, dass drei Merkmale zur Bestimmung von Itemschwierigkeiten wichtig sind: die Anforderung, elementares vs. nicht-elementares mathematisches Wissen anzuwenden, rein mathematische Anforderungen vs. die Anforderung, mathematisches mit mathematikdidaktischem Wissen zu verknüpfen sowie das Inhaltsgebiet eines Items. Sowohl unsere qualitativen Analysen als auch die einfaktoriellen statistischen Prüfungen stützen diese Hypothese. Items, die elementares mathematisches Wissen, mathematikdidaktische Verknüpfungsleistungen oder die Anwendung von Wissen aus den Inhaltsgebieten Arithmetik, Geometrie und Stochastik fordern, gehören zu den Stärken angehender GHR-Lehrkräfte. Items, die über elementares mathematisches Wissen hinausgehen, die fachmathematische Anforderungen stellen oder aus den Inhaltsgebieten Algebra und Funktionen stammen, gehören dagegen zu den Stärken angehender GyGS-Lehrkräfte. Eine multiple Hypothesenprüfung bestätigt dieses Ergebnis statistisch signifikant für das mathematische Niveau und das Inhaltsgebiet.

Die häufig integriert angelegte fachliche Sekundarstufen-I-Ausbildung leistet offensichtlich in vielfacher Hinsicht, was von ihr erwartet wird. Das Wissen der angehenden GHR-Lehrkräfte repräsentiert typische berufliche Anforderungen, was auf eine schwerpunktmäßige Vorbereitung in diesen Gebieten schließen lässt. Gleichzeitig lässt sich dasselbe für die Gymnasiallehrerausbildung feststellen, deren Absolvierenden mit gänzlich anderen beruflichen Anforderungen konfrontiert sind. Eine Kompetenzeinschätzung sollte daher immer berücksichtigen, welches Aufgabenspektrum zu handhaben ist.

Implizit werden durch eine Analyse von Stärken selbstverständlich auch Schwächen deutlich. Auf zwei Punkte soll hingewiesen werden: Sobald das verlangte mathematische Niveau elementare Kenntnisse überschreitet, haben angehende GHR-Lehrkräfte Schwierigkeiten bei der Lösung der Items. Mit Blick auf die höheren Jahrgänge der Sekundarstufe I, vor allem die beiden Abschlussklassen 9 und 10, die erst zu einem Mittlerem Bildungsabschluss führen, wird hier ein Defizit deutlich. Umgekehrt ist festzuhalten, dass das lange fachwissenschaftliche Studium der Gymnasiallehrkräfte in unserem Test nicht dazu geführt hat, dass elementare Mathematikaufgaben besser gelöst werden. Hier ist mit Blick auf die Eingangsklassen der Sekundarstufe I zu fragen, inwieweit die Ausbildung Defizite aufweist.

Auf der Basis des *MT21*-Tests kann nicht entschieden werden, wo die Ursache für die festgestellten Defizite liegt. Dass die Verknüpfung elementarmathematischer Aufgaben mit didaktischen Anforderungen nicht mehr signifikant zu einer besseren Itemklassifizierung beiträgt, ist u.a. ein Ergebnis der Testanlage: Elementarmathematische Aufgaben sind im *MT21*-Test so gut wie immer in einen didaktischen Kontext eingebunden, Aufgaben auf universitärmathematischem Niveau sind dagegen so gut wie immer frei von mathematikdidaktischen Anforderungen. Offensichtlich erschien es bei der Testentwicklung nicht nahe liegend, unabhängig von mathematikdidaktischen Anforderungen elementares mathematisches Wissen zu testen bzw. universitäre Mathematik didaktisch zu rahmen. Für zukünftige Testentwicklungen ist daraus die Konsequenz zu ziehen, entsprechende Aufgaben zu entwickeln. Denn erst so wird das gesamte Fähigkeitsspektrum der Probanden erfasst und erst so lässt sich schlüssig entscheiden, ob das Wissensniveau oder die didaktische Einbettung für die Schwierigkeiten der angehenden GHR- bzw. GyGS-Lehrkräfte entscheidend war.

Unter methodischen Gesichtspunkten ist zu betonen, dass unsere Analysen eine Obergrenze differentieller Itemfunktionen darstellen, da wir alle Items als Anker verwendet haben. Dies kann zu einer Inflation der Entdeckung differentieller Itemfunktionen führen (Wang, 2004). Allerdings sollte sich das Problem in unserem Falle deutlich in Grenzen halten, da nur gut ein Viertel der Items Abweichungen aufweisen und inflatorischen Wirkungen erst bei größeren Abweichungen auftreten (Stark et al., 2006). Unter dem Gesichtspunkt der Testentwicklung ist abschließend zu thematisieren, ob die differentiellen Itemfunktionen so groß sind, dass von einem unfairen Testbias gesprochen werden muss (Camilli & Shepard, 1994). Davon kann im vorliegenden Fall jedoch nicht die Rede sein. Die untersuchten Fähigkeiten sind Bestandteil dessen, was als professionelle Kompetenz angehender Mathematiklehrkräfte definiert wurde. Die Unterschiede, die herausgearbeitet wurden, spiegeln Unterschiede in den Ausbildungsgängen, die mindestens zum Teil notwendig sind, wenn man die unterschiedlichen Aufgaben von GHR- und GyGS-Lehrkräften im Berufsalltag betrachtet.

5 Literatur

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2006). *Multivariate Analysemethoden* (11. Aufl.). Berlin: Springer.

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.

Biehler, R. & Hartung, R. (2006). Die Leitidee Daten und Zufall. In W. Blum et al. (Hg.), *Bildungsstandards Mathematik: konkret.* Berlin: Cornelsen Verlag Scriptor, 51–80.

Blömeke, S. (2002). *Universität und Lehrerausbildung.* Bad Heilbrunn: Klinkhardt.

Blömeke, S., Kaiser, G. & Lehmann, R. (Hg.) (2008), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung.* Münster: Waxmann.

Blömeke, S., Kaiser, G., Schwarz, B., Seeber, S., Lehmann, R., Felbrich, A. et al. (2008a). Fachbezogenes Wissen am Ende der Ausbildung. In S. Blömeke, G. Kaiser & R. Lehmann (Hg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematik-Studierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung.* Münster: Waxmann, 89–104.

Blömeke, S., Lehmann, R., Seeber, S., Schwarz, B., Kaiser, G., Felbrich, A. et al. (2008b). Niveau- und institutionenbezogene Modellierungen des fachbezogenen Wissens. In S. Blömeke, G. Kaiser & R. Lehmann (Hg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematik-Studierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung.* Münster: Waxmann, 105–134.

Blömeke, S., Seeber, S., Lehmann, R., Kaiser, G., Schwarz, B., Felbrich, A. et al. (2008c). Messung des fachbezogenen Wissens angehender Mathematiklehrkräfte. In S. Blömeke, G.

Kaiser & R. Lehmann (Hg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematik-Studierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung.* Münster: Waxmann, 49–88.

Bromme, R. (1992). *Der Lehrer als Experte. Zur Psychologie des professionellen Lehrerwissens.* Göttingen: Hans Huber.

Brunner, M., Kunter, M., Krauss, S., Klussmann, U., Baumert, J., Blum, W. et al. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht; eine Zwischenbilanz des COACTIV-Projekts. In M. Prenzel & L. Allolio-Näcke (Hg.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms.* Münster: Waxmann, 54–82.

Budgell, G.R., Namburty, S.R. & Douglas, A.Q. (1995). Analysis of differential item functioning in translated assessment instruments. *Applied Psychological Measurement,* 19, 309–321.

Camilli, G. & Shepard, L.A. (1994). *Methods for identifying biased test items.* Bd. 4. Thousand Oaks: Sage.

Franke, M. (2000). *Didaktik der Geometrie.* Heidelberg: Spektrum.

Gelman, A. & Carlin**,** J.B. (2002). Poststratification and Weighting Adjustments. In Groves, R.M., Dillman D.A., Eltinge, J.L. & Little, R.J.A. (Hg.), *Survey Nonresponse.* New York: Wiley, 288–302.

Graeber, A. & Tirosh, D. (2008). Pedagogical Content Knowledge: Useful Concept or Elusive Notion. In P. Sullivan & T. Woods (Eds.), *Knowledge and Beliefs in Mathematics Teaching and Teaching Development. The International Handbook of Mathematics Teacher Education, Vol. 1.* Rotterdam: Sense Publisher, 117–132.

Hill, H.C., Loewenberg Ball, D. & Schilling, S.G. (2008). Unpacking Pedagogical Content Knowledge: Conceptualising and Measuring Teachers' Topic-specific Knowledge of Students. *Journal for Research in Mathematics Education,* 39(4), 372–400.

Holland, P.W. & Wainer, H. (1993). *Differential Item Functioning.* Mahwah, NJ: Lawrence Erlbaum.

Janvier, C. (1981). Use of situations in mathematics education. *Educational Studies in Mathematics,* 12(1), 113–122.

Kirsch, A. (1987). *Mathematik wirklich verstehen.* Köln, Aulis Verlag Deubner.

Klein, F. (1933). *Elementarmathematik vom höheren Standpunkte aus* (4. Aufl.). Erster Band. Berlin: Springer.

Klieme, E. & Baumert, J. (2001). Identifying national cultures of mathematics education: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology of Education*, 16(2001), 385–402.

[KMK] Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hg.) (2003), *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss.* Bonn: KMK.

Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M. et al. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv-aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung.* Münster: Waxmann, 31–53.

Krauss, S., Baumert, J. & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: validation of the COACTIV constructs. *ZDM – The International Journal on Mathematics Education*, 40(5), 873–892.

Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M. et al. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29(3/4), 223–258.

Krauthausen, G. & Scherer, P. (2007). *Einführung in die Mathematikdidaktik* (3. Aufl.). Heidelberg: Elsevier.
Krutezki, W.A. (1966). *Zur Struktur der mathematischen Fähigkeiten.* Berlin, Volk und Wissen.
Kütting, H. (1994). *Didaktik der Stochastik.* Mannheim, BI Wissenschaftsverlag.
Liljedahl, P., Durand-Guerrier, V., Winslow, C., Bloch, I., Huckstep, P., Rowland, T. et al. (2009). Components of mathematics teacher training. In R. Even & D. Loewenberg Ball (Hg.), *The professional education and development of teachers of mathematics. The 15th ICMI Study.* New York: Springer, 25–34.
Loewenberg Ball, D. & Bass, H. (2003). Interweaving Content and Pedagogy in Teaching and Learning to Teach: Knowing and Using Mathematics. In J. Boaler (Ed.), *Multiple Perspectives on Mathematics Teaching and Learning.* Westport, Ablex Publishing., 83–104.
Malle, G. (1993). *Didaktische Probleme der elementaren Algebra.* Braunschweig, Wiesbaden: Vieweg.
Millsap, R. E. & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17*, 297–334.
Pepin, B. (1999). Existing models of knowledge in teaching: developing an understanding of the Anglo/American, the French and the German scene. In B. Hudson, F. Buchberger, P. Kansanen & H. Seel (Eds.), *Didaktik/Fachdidaktik as Science(.s) of the the Teaching Profession?* Umea, TNTEE Publication, 49–66.
Shulman, L. S. (1986). Those who understand. Knowledge growth in teaching. *Educational Researcher*, 15, 4–14.
Stark, S., Chernyshenko, O. S. & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91,* 1291–1306.
Swaminathan, N. & Rogers, H.J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement,* 27, 361–370.
Tietze, U.-P., Klika, M. & Wolpers, H. (1997). *Mathematikunterricht in der Sekundarstufe II.* Bd. 1. Braunschweig: Vieweg.
Urban, D. (1993). *Logit-Analyse. Statistische Verfahren zur Analyse von Modellen mit qualitativen Respone-Variablen.* Stuttgart: Fischer.
Vollrath, H.-J. (2001). *Grundlagen des Mathematikunterrichts in der Sekundarstufe.* Heidelberg: Spektrum.
Wang, W. (2004). Effects of anchor item methods on detection of differential item functioning within the family of Rasch models. *Journal of Experimental Education, 72,* 221–261.
Weinert, F.E. (1999). *Konzepte der Kompetenz. Gutachten zum OECD-Projekt "Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo)".* Neuchatel: Bundesamt für Statistik.

Sigrid Blömeke, Rainer Lehmann
Humboldt Universität zu Berlin/ Philosophische Fakultät IV
Geschwister-Scholl-Straße 7, 10099 Berlin

Björn Schwarz, Gabriele Kaiser
Universität Hamburg/ Fachbereich Erziehungswissenschaft
Von-Melle-Park 8, 20146 Hamburg

Susan Seeber
Deutsches Institut für Internationale Pädagogische Forschung
Warschauer Str. 34, 10243 Berlin