

# Arbeitskreis ‚Vergleichsuntersuchungen zum Mathematikunterricht‘

Bielefeld, 24.–25. 11. 2007

Gabriele Kaiser und Norbert Knoche

Am 24. und 25. November 2007 fand die diesjährige Herbsttagung des Arbeitskreises „Vergleichsuntersuchungen zum Mathematikunterricht“ statt. Tagungsort war die Universität Bielefeld.

Auf dem Programm standen am ersten Tag ein Vortrag von Olaf Köller mit dem Thema „Definition von Kompetenzstufen im Fach Mathematik in der Grundschule“, ein Vortrag von Anja Felbrich zum Thema „Ergebnisse der Studie MT21“ und ein Vortrag von Stefan Kraus zum Thema „Die Untersuchung des professionellen Wissens deutscher Mathematiklehrerinnen und -lehrer im Rahmen der COACTIV-Studie“.

Am zweiten Tag standen die folgenden Vorträge auf dem Program, ein Vortrag von Hans-Dieter Sill und Christine Sikora mit dem Thema „Leistungserhebungen im Mathematikunterricht – Merkmale, Funktionen und Analysemethoden“, ein Vortrag von Rudolf vom Hofe, Alexander Jordan und Pascal Stölting mit dem Thema „Analyse der Leistungsentwicklung in Mathematik – Das Projekt PALMA“ und ein Vortrag von Andreas Büchter und Andreas Pallack mit dem Thema „Zentrale Prüfungen in NRW: Schülerleistungen am Ende der Klasse 10“.

Nachfolgend werden die zentralen Aspekte der Vorträge kurz referiert.

*Olaf Köller: „Definition von Kompetenzstufen im Fach Mathematik in der Grundschule“*

Im Vortrag wurde zunächst noch einmal kurz für die Grundschule der Fahrplan von den Bildungsstandards bis hin zum ersten Ländervergleich vorgestellt. Für die im Herbst 2004 verabschiedeten Bildungsstandards im Fach Mathematik in der Grundschule wurden im Laufe des Jahres 2005 rund 1.000 Aufgaben zu den verschiedenen inhaltlichen und prozessbezogenen Kompetenzen entwickelt. Die Aufgaben wurden im Frühjahr 2006 an einer Stichprobe von rund 12.000 Schülerinnen und Schülern der 3. und 4. Jahrgangsstu-

fe erprobt. Ein großer Teil der Aufgaben erwies sich fachdidaktisch und psychometrisch als geeignet und wurde im Frühjahr 2007 in einer großen Normierungsstudie eingesetzt. Im Rahmen der Normierungsstudie bearbeiteten rund 7.500 Schülerinnen und Schüler aus 3. und 4. Jahrgangsstufen die Aufgaben. Die Normierungsdaten sind in diesen Tagen im IQB eingetroffen und werden aktuell ausgewertet. Der entsprechende Bericht zur Normierung wird im Sommer 2008 erscheinen. Die normierten Aufgaben werden dann Grundlage des Ländervergleichs sein, der parallel zu PIRLS/IGLU und TIMSS 2011 stattfinden wird. Auf der Basis der Pilotierungsdaten wurden dann im Vortrag zwei empirische Schwerpunkte gelegt. Zunächst wurden Analysen zur Problematik unterschiedlicher Dimensionalitätsannahmen präsentiert. Unterschieden wurde zwischen within-item-dimensionality (WID; ein Item kann auf mehreren Faktoren laden) und between-item-dimensionality (BID; jedes Item lädt nur auf einem Faktor). In der Vergangenheit (z. B. in PISA) wurde in den Dimensionsanalysen – egal ob für inhaltliche oder prozessbezogene Kompetenzen – primär die Annahme BID verfolgt und auch geprüft, wohl wissend, dass viele Items mehr als eine Dimension testen. Die Kosten dieser u. E. nicht haltbaren Annahmen wurden dann immer in extrem hohen Korrelationen (über .80) zwischen unterschiedlichen mathematischen Dimensionen sichtbar. Im Rahmen unserer Arbeiten haben wir Modelle unter der plausibleren WID-Annahme getestet. Hierbei zeigen sich für die Grundschuldaten im Fach Mathematik deutliche geringere Korrelationen zwischen mathematischen Dimensionen. Prüft man beispielsweise ein Modell mit fünf inhaltlichen Kompetenzdimensionen, so liegen die Korrelationen zwischen .50 und .70, also deutlich unter 1. Es lässt sich so belegen, dass man unterschiedliche mathematische Dimensionen gut analytisch trennen kann.

Abschließend wurde auf die Definition von Kompetenzstufen auf der Basis der Pilotierungsdaten eingegangen. Dabei wurden zunächst folgende Argumente vorgebracht:

- ▷ Die in Schulleistungstudien untersuchten Merkmale stellen latent traits dar.
- ▷ Jede Stufe des latenten Kontinuums ist letztendlich willkürlich und dient als Interpretationshilfe für Praktiker (in Schule und Politik): "Dividing (...) these continua into levels, though useful for communication about students' development, is essentially arbitrary." (Adams/Wu 2002, S. 197).
- ▷ Entwicklungspsychologische Interpretationen der Stufen stellen einen Rückfall in überholte theoretische Traditionen dar
- ▷ Ziel der Niveaufestsetzung müssen plausibel interpretierbare und kommunizierbare Stufen sein
- ▷ Hierzu sollten Experten unterschiedlicher Gruppen eingebunden werden

Anschließend wurde das Vorgehen bei der Niveaufestsetzung im IQB vorgestellt. Dies stellt eine Mischung aus Empirie und fachdidaktischer Setzung von Grenzen dar. Im ersten Schritt werden die Aufgaben nach ihren empirischen Schwierigkeiten sortiert. Anschließend definieren fachdidaktische Expertinnen/Experten, bei welchen Items auf dem Schwierigkeitskontinuum deutliche Sprünge in den kognitiven Anforderungen zu verzeichnen sind. An diesen Übergängen werden dann die vorläufigen Grenzen gesetzt, die dann mit der Politik und Verbandsvertretern abgestimmt werden. Aktuell stellt sich der Arbeitsstand so dar, dass für die vierte Jahrgangsstufe fünf fachdidaktisch plausible Kompetenzstufen definiert wurden, die sich natürlich an Vorarbeiten aus IGLU 2001 und PISA orientieren. In weiteren Schritten wird es nötig sein, auf der Basis dieses Kompetenzstufenmodells die genauen Grenzen zu definieren, bei denen davon ausgegangen werden kann, dass Schülerinnen und Schüler die in den Standards festgelegten Leistungserwartungen erreicht haben.

*Anja Felbrich: „Epistemologische Überzeugungen zur Mathematik am Ende der Lehrerbildung – Ergebnisse der Studie MT21“*

Es gibt bereits seit langem eine starke Kritik an der Lehrerbildung, ohne dass bisher ihre Wirksamkeit umfassend empirisch geprüft wurde. Es gibt auch nur wenig empirisch gesicherte Erkenntnisse, wie sich die international äußerst unterschiedliche Gestaltung der Ausbildungssysteme auf das Wissen und die Kompetenzentwicklung

von zukünftigen Lehrerinnen und Lehrern auswirkt. Die International Association for the Evaluation of Educational Achievement (IEA) hat daher am Beispiel der Ausbildung von Mathematiklehrerinnen und -lehrern für die Primarstufe und die Sekundarstufe I im Jahr 2006 eine internationale Vergleichsstudie (Teacher Education and Development Study: Learning to Teach Mathematics – TEDS-M) begonnen, die erstmals weltweit die Wirksamkeit unterschiedlicher Lehrerbildungssysteme erheben soll. Ergebnisse der Studie, an der bisher ca. 20 Länder teilnehmen, sind für Dezember 2009 angekündigt. Auch Deutschland wird an dieser Studie durch die DFG gefördert teilnehmen (Projektleitung Sigrid Blömeke, Gabriele Kaiser, Rainer Lehmann).

Aufgrund der Schwierigkeiten der Konzeptualisierung und theoretischen Fundierung wurde im Auftrag der IEA eine Vorbereitungsstudie in sechs Ländern durchgeführt, die Studie Mathematics Teaching in the 21st Century (MT21) ursprünglich P-TEDS genannt, ebenfalls mit deutscher Beteiligung (Leitung Sigrid Blömeke). MT21 hat an vier Regionen in Deutschland Studierende und Referendar(innen) des Fachs Mathematik für die Sekundarstufe I an Universitäten und Studienseminaren getestet; insgesamt wurden 849 Studierende und Referendarinnen und Referendare aus 4 Universitäten und 22 Studienseminaren getestet. Dabei wurde auf Seiten der zukünftigen Lehrpersonen Wissen in den drei Bereichen Mathematik, Mathematikdidaktik und Pädagogik-Psychologie erfasst. Darüber hinaus wurden die angehenden Mathematiklehrerinnen und -lehrer zu ihren Überzeugungen und Persönlichkeitsmerkmalen befragt. Des Weiteren wurden Institutionen- und Curriculumanalysen durchgeführt, die auf Seiten der ausbildenden Institutionen das intendierte und das implementierte Curriculum beschreiben.

Im Vortrag wurden Ergebnisse zu den epistemologischen Überzeugungen der zukünftigen Lehrpersonen vorgestellt. Die grundlegende Annahme von MT21 ist, dass Überzeugungen eine orientierende und handlungsleitende Funktion besitzen und somit eine Brücke zwischen Wissen und Handeln bilden. Allerdings handelt es sich bei Überzeugungen um ein relativ schlecht definiertes Konstrukt mit wenig präzisen Abgrenzungen von ähnlichen Konstrukten. In Anlehnung an gängige Ausdifferenzierungen des Überzeugungssystems von Mathematik-Lehrkräften wird in MT21 zwischen vier Gruppen an Überzeugungen unterschieden, wobei Ergebnisse zu der ersten Gruppe dargestellt wurden:

- ▷ Epistemologische Überzeugungen zur Mathematik
- ▷ Unterrichtsbezogene Überzeugungen zum Lehren und Lernen von Mathematik
- ▷ Professionsbezogene Überzeugungen zur Rolle von Schule und Lehrerberuf in der Gesellschaft
- ▷ Selbstbezogene Überzeugungen der Mathematiklehrkräfte

In Bezug auf die Struktur der Mathematik als akademischer Disziplin ist es gelungen, die vier Skalen aus den Studien mit praktizierenden Lehrkräften von Grigutsch, Raatz und Törner (1998) auch bei angehenden Mathematiklehrerinnen und -lehrern zu reproduzieren. Ebenso deutet sich wie bei Grigutsch, Raatz und Törner (1998) eine Ordnung der vier Skalen nach statischen und dynamischen Aspekten der Mathematik an. Anders als in den Studien mit praktizierenden Lehrkräften aber, in denen das Zusammenhangsmuster auf antagonistische Leitvorstellungen hindeutet, vertreten angehende Mathematik-Lehrkräfte eher das von Experten bevorzugte Muster einer Janusköpfigkeit der Mathematik. Es kann angenommen werden, dass sich hier die in den letzten Jahrzehnten intensiviertere und professionalisiertere mathematikdidaktische Ausbildung niederschlägt, in der genau solche Meta-Reflexionen über die Natur der Mathematik ihren Platz haben.

Im Grad der Zustimmung zu den vier Subdimensionen lassen sich Unterschiede finden: Insgesamt dominiert der Prozess- und Anwendungsbezug die Sicht der Referendarinnen und Referendare auf Mathematik, während schemaorientierte Überzeugungen eine untergeordnete Rolle spielen. Damit bestätigen sich die deskriptiven Ergebnisse von Grigutsch, Raatz und Törner (1998).

Referendarinnen und Referendare unterschiedlicher Ausbildungsgänge unterscheiden sich hinsichtlich ihrer mathematischen Weltbilder nicht substantiell voneinander. Referendare für das GHR-Lehramt favorisieren lediglich den Anwendungsaspekt der Mathematik stärker als GyGS-Referendare. Dies lässt sich möglicherweise auf unterschiedliche fachbezogene Ausbildungserfahrungen zurückführen: Während die entsprechende universitäre Ausbildung der angehenden Gymnasial- und Gesamtschullehrer weitgehend in der Fachwissenschaft Mathematik stattfindet, erwerben die angehenden GHR-Lehrkräfte ihre mathematische Ausbildung zu einem erheblichen Anteil in der Mathematikdidaktik und in der Grundschulpädagogik (Lernbereich Mathematik). Hier steht der Anwendungsaspekt vermutlich deutlich stärker im Vordergrund.

Unter anthropologischen Gesichtspunkten wurden vier Skalen eingesetzt: zwei mit begabungstheoretischen Bezügen und zwei mit erkenntnistheoretischen Bezügen. In einer begabungstheoretischen Perspektive vertreten die angehenden Mathematik-Lehrkräfte angemessene Überzeugungen, indem sie deterministische Vorstellungen deutlich ablehnen und stattdessen das Vorverständnis von Kindern in Betracht ziehen. Zwischen GHR- und GyGS-Lehrerinnen und -Lehrern zeigen sich hier keine bedeutsamen Unterschiede. Dazu passend erhält eine kognitiv-konstruktivistische Sichtweise auf die Genese mathematischer Kompetenz sehr hohe Zustimmung, während die Transmissions-Orientierung abgelehnt wird. Auch dieses Muster zeigt sich gleichmäßig über die in MT21 untersuchten Teilpopulationen der GHR- und GyGS-Lehrer. Beide Ergebnisse spiegeln die aktuelle didaktische Diskussion, sodass ein Zusammenhang zu Ausbildungserfahrungen angenommen werden kann.

Weitere Analysen bestätigen, dass begabungs- und erkenntnistheoretische Überzeugungen zur Genese mathematischer Kompetenz systematische Zusammenhänge untereinander aufweisen. Dabei korrelieren die als antipodisch angesehenen Skalen zur Mathematik als anthropologischer Konstante und zum *Conceptual Change* sowie die Transmissions- und Konstruktionsorientierung jeweils wie erwartet latent negativ. Was den Zusammenhang zwischen Überzeugungen zur Struktur der Mathematik und anthropologischen Überzeugungen angeht, zeigt sich die erwartete hohe Bedeutung einer prozess- und etwas schwächer auch noch einer anwendungsorientierten Perspektive auf die Mathematik. Sie korrelieren jeweils sowohl hoch positiv mit angemessenen *begabungstheoretischen* Überzeugungen als auch mit einer kognitiv-konstruktivistisch orientierten Erkenntnistheorie. In Bezug auf Letzteres zeigt sich auch noch eine Bedeutung der statischen Perspektiven auf Mathematik, indem diese stark mit einer Transmissionsorientierung zusammenhängen. In Bezug auf begabungstheoretische Überzeugungen zeigen sich aber keine Zusammenhänge.

Die MT21-Ergebnisse unterstützen die Befunde anderer Studien, dass eine prozessorientierte Sicht von Mathematik mit konstruktivistischen Ansichten zum Lernen von Mathematik einhergeht. Offensichtlich muss man insgesamt von einem umfassenden *Überzeugungs-Syndrom* sprechen, was epistemologische Überzeugungen angeht, da die von den angehenden Mathematik-Lehrerinnen und -Lehrern vertretenen Überzeugungen in sich hoch konsistent sind, worauf auch die Ergebnisse

der COACTIV-Studie hinweisen (Baumert et al., 2004).

Die Vernetztheit der epistemologischen Überzeugungen wirft Fragen zur Veränderbarkeit im Zuge der Lehrerbildung auf. Auch wenn angesichts der Passung dieser empirischen Ergebnisse zu aktuellen allgemein- und mathematikdidaktischen Diskussionen ein Ausbildungseinfluss angenommen werden kann, sind Veränderungen doch deutlich schwieriger zu erreichen, wenn ein System an Überzeugungen stabil zueinander passt, als wenn es sich um nur lose miteinander verknüpfte Dimensionen handelt. Eine ausführliche Darstellung der deutschen Gesamtergebnisse der Studie MT21 findet sich in Blömeke, Kaiser, Lehmann (2008).

#### Literatur

Baumert, J., Kunter, M., Brunner, M., Krauss, St., Blum, W. & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In: Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (Hrsg.): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. Münster: Waxmann, S. 314–354.

Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.) (2008). Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematik-Studierender und -Referendare – erste Ergebnisse zur Wirksamkeit der Lehrerbildung. Münster: Waxmann Verlag.

Grigutsch, S., Raatz, U., & Törner, G. (1998). Einstellungen gegenüber Mathematik bei Mathematiklehrern. *Journal für Mathematik-Didaktik*, 19, 3–45.

*Stefan Krauss: „Die Untersuchung des professionellen Wissens deutscher Mathematiklehrerinnen und -lehrer im Rahmen der COACTIV-Studie“*

In Deutschland fehlen bislang empirische Untersuchungen, die professionelles Wissen und Kompetenzen von Lehrkräften systematisch erfassen und in Zusammenhang zu Unterrichtsaspekten und zur Leistung der Schüler stellen. Die COACTIV-Studie 03/04 (Projektleiter: J. Baumert, W. Blum, M. Neubrand), in der die Mathematiklehrkräfte der deutschen PISA-Klassen 03/04 ausführlich befragt und getestet wurden, bot die einmalige Gelegenheit, im Verbund mit PISA ein breites Spektrum sowohl von Schüler- als auch von Lehrerdaten zu erheben und diese gemeinsam zu analysieren. Im Rahmen von COACTIV wurden

zahlreiche Instrumente für Lehrkräfte adaptiert bzw. neu entwickelt (z. B. zur Messung von fachlichen und fachdidaktischen Kompetenzen, aber auch zu motivationalen Orientierungen, Überzeugungen, Aspekten des Berufserlebens usw).

Ein Instrument ist dabei von besonderer fachdidaktischer Relevanz, nämlich die neu entwickelten Tests zum fachdidaktischen Wissen und zum Fachwissen von Mathematiklehrkräften. Fachdidaktisches Wissen (gemessen mit insgesamt 22 Items) ist in COACTIV in drei Bereiche aufgeteilt, nämlich in Wissen über Erklären und Repräsentieren mathematischer Sachverhalte (11 Items), in Wissen über typische Schülerfehler und Schwierigkeiten (7 Items) und in Wissen über die multiple Lösbarkeit von Aufgaben (4 Items). Fachwissen (gemessen mit 13 Items) ist konzeptualisiert als vertieftes Hintergrundwissen über den Schulstoff. Im Folgenden sollen stichpunktartig einige Hauptergebnisse der Durchführung dieser Tests im Jahre 2004 mit 198 deutschen PISA-Mathematiklehrkräften berichtet werden (die sich immer auf die in COACTIV vorgenommene Konzeptualisierung und Operationalisierung beziehen).

Lehrkräfte am Gymnasium verfügen – wie zu erwarten – über deutlich mehr Fachwissen als ihre Kollegen von anderen Schulformen, aber auch über signifikant mehr fachdidaktisches Wissen (auch wenn der Unterschied hier weniger stark ausgeprägt ist). Partialisiert man das Fachwissen aus, schneiden die Lehrkräfte nicht-gymnasialer Schulformen beim Fachdidaktiktest jedoch etwas besser ab als die Gymnasiallehrkräfte. Erstaunlicherweise konnte keine positive Korrelation der beiden Wissensbereiche Fachwissen und fachdidaktisches Wissen mit der Berufserfahrung gefunden werden, die Wissensentwicklung scheint also nach der Ausbildung im Wesentlichen bereits abgeschlossen zu sein. Spezifiziert man Strukturgleichungsmodelle zur Vorhersage des Leistungszuwachses der PISA-Schüler, so zeigt sich, dass das fachdidaktische Wissen – dies wiederum unterstützt durch das Fachwissen – einer Lehrkraft einen substantiellen Beitrag zur Erklärung der Lernzuwächse der Schüler liefert.

*Hans-Dieter Sill, Christine Sikora: „Leistungserhebungen im Mathematikunterricht – Merkmale, Funktionen und Analysemethoden“*

Im ersten Teil des Vortrages stellte H.-D. Sill ein neues Konzept zur Beschreibung von Merkmalen und Funktionen von Leistungserhebungen vor, wobei er unter Leistungserhebungen alle Formen der Leistungsmessung von Kurzkontrollen bis zu

internationalen Vergleichsstudien versteht. Das Konzept entstand im Ergebnis einer Analyse aller aktuellen Erhebungen in Deutschland bis 2006, die gleichzeitig auf dieser Grundlage einer kritischen Analyse unterzogen wurden. Die Funktion einer Leistungserhebung sollte in Bezug zu einem der folgenden Entwicklungsprozesse bestimmt werden:

- ▷ Entwicklung des mathematischen Wissens und Könnens von Lernenden,
- ▷ Entwicklung des beruflichen Wissens und Könnens von Lehrern ,
- ▷ Entwicklung der Rahmenbedingungen von Schule und Unterricht.

Jedes der 10 ermittelten Merkmale von Leistungserhebungen muss dazu jeweils eine spezifische Ausprägung besitzen. Es zeigt sich, dass insbesondere die aktuellen Vergleichsarbeiten eine Reihe notwendiger Bedingungen nicht erfüllen und somit die in sie gesetzten Erwartungen nicht erfüllen können.

Sill sieht die Hauptursache für die aktuellen Fehlentwicklungen in den Desiderata der empirischen didaktischen Forschung der letzten Jahrzehnte. Er forderte eine Umorientierung von der z. Z. dominierenden interpretativen und damit nur zustandsbeschreibenden psychologisch orientierten Lehr-Lernforschung zu einer konstruktiven schulorientierten didaktischen Forschung, die die Entwicklungsprozesse psychischer Dispositionen unter den historisch konkreten Bedingungen des Klassenunterrichts in Deutschland zum Gegenstand hat. Dazu sind eine Emanzipation der Fachdidaktik von der Bildungsforschung und eine Ablösung der gegenwärtigen Kultur einer zentral gesteuerten Zustandsevaluation durch eine Kultur der selbst bestimmten Prozessevaluation und Prozessveränderung an Schulen erforderlich.

H.-D. Sill stellte weiterhin die Grundzüge eines neuen Kompetenzmodells vor, bei dem die Grundlage der Stufung nicht primär die Anforderungen der Aufgaben sind, sondern die auszubildende Qualität der psychischen Dispositionen und dabei insbesondere der angestrebte Beherrschungsgrad durch die Schüler. Dieses Modell wird z. Z. mit Erfolg in Mecklenburg-Vorpommern zur Strukturierung der Ziele in der neuen gymnasialen Oberstufe verwendet.

Ch. Sikora erläuterte im zweiten Teil an Hand von Beispielen aus dem Projekt Vergleichsarbeiten in Mecklenburg-Vorpommern eine neue Methode der Auswertung von Schülerantworten, den Zusammenhang von Lerngelegenheiten und Schülerleistungen sowie Aktivitäten zur nachhaltigen

Rückwirkung auf den Unterricht. Sie zeigte an einem Beispiel, welche Potenzen für die didaktische Forschung in einer detaillierten und tiefgründigen Analyse von Schülerantworten aus Massenerhebungen liegen können.

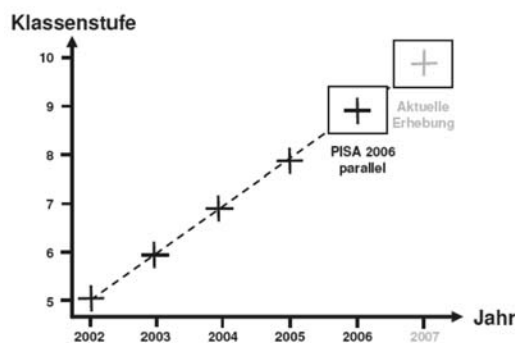
Als eine wirksame Form der Unterrichtsentwicklung auf der Basis von Ergebnissen der Vergleichsarbeiten haben sich unterrichtsbegleitende ganzjährige Lehrerfortbildungen erwiesen, die auf der Grundlage von theoretisch fundierten Materialien durchgeführt werden. Dass zur Entwicklung dieser Materialien eine Zusammenarbeit von Didaktikern und Schulpraktikern notwendig ist, hat die Erarbeitung von drei Broschüren zum sicheren Wissen und Können in Mecklenburg-Vorpommern gezeigt.

Die Ergebnisse der Untersuchungen von Sill und Sikora wurden 2007 in der Reihe „Texte zur mathematischen Forschung und Lehre“ bei Franzbecker veröffentlicht.

*Rudolf vom Hofe, Alexander Jordan, Pascal Stöltzing:*  
 „Analyse der Leistungsentwicklung in Mathematik- Das Projekt PALMA“

Im Vortrag wird über das DFG-Projekt PALMA (Kurzform von **P**rojekt zur **A**nalyse der **L**eistungsentwicklung in **M**athematik; Laufzeit: 2000-2008; Leitung: Reinhard Pekrun/LMU München, Rudolf vom Hofe, Universität Bielefeld, Werner Blum, Universität Kassel) berichtet. Dazu werden neben (1.) Motivation und Anlage der Studie sowie (2.) Ergebnissen zur Entwicklung mathematischer Kompetenzen von Lernenden der Sekundarstufe I auch (3.) Detailuntersuchungen in den Klassen 6 und 8 ausgeführt und (4.) Perspektiven für eine individuelle Diagnose und Förderung aufgewiesen.

(1) *Motivation und Anlage:* Ziel dieses Projekts ist es, in einer Längsschnittstudie Entwicklungsverläufe, Schülervoraussetzungen und Kontextbedingungen von Mathematikleistungen bei Schülern der 5.-10. Klassenstufe zu untersuchen.



Hierzu werden im jährlichen Rhythmus an einer für Bayern repräsentativen Schülerkohorte in Gymnasien, Realschulen und Hauptschulen Erhebungen durchgeführt, die inhaltlich und methodisch so angelegt sind, dass eine Verschränkung mit den Erhebungen von PISA 2006 möglich ist. Zur empirischen Analyse wurden der Regensburger Mathematikleistungstest sowie Münchener Skalen zu Mathematikemotionen, Schülervoraussetzungen und Kontexten entwickelt und eingesetzt.

(2) *Ergebnisse zur mathematischen Kompetenzentwicklung:* Die Erhebungen finden jeweils am Ende eines Schuljahres statt, so dass damit die Entwicklung vom Ende der Klasse 5 bis zum Ende der Klasse 10 beschrieben werden kann: (a) Dabei zeigt sich eine weitgehend parallele Entwicklung von Realschule und Gymnasium, wobei das Gymnasium erwartungsgemäß durchgehend über der Realschule liegt. (b) Die Hauptschule liegt mit deutlich größerem Abstand unterhalb der Realschule. Die Leistungsentwicklung der Hauptschüler nach der Klasse 7 verläuft deutlich moderater als in den anderen beiden Schularten, so dass man fast von Phasen der Stagnation sprechen kann. (c) In allen Schularten folgt nach einem deutlichen Anstieg der Mittelwerte in der Jahrgangsstufe 6 eine Phase der Stagnation in der Jahrgangsstufe 7. (d) Am Ende der Klasse 5 zeigt sich eine starke Überschneidung der Leistungsspektren für Hauptschule, Realschule und Gymnasium. In den folgenden Jahren zeigen sich trotz einer Auseinanderentwicklung der Mittelwerte immer noch starke Überlappungen. So liegen z. B. selbst in der Klasse 9 noch zahlreiche gute Hauptschüler in einem ähnlichen Leistungsbereich wie weniger gute Gymnasiasten. Diese Entwicklung bestärkt die Forderung nach einem Schulsystem, das eine hohe Durchlässigkeit zwischen Zweigen bzw. Schulformen ermöglicht, um zum einen individuelle Bildungsgerechtigkeit zu gewährleisten und zum anderen das vorhandene Bildungspotential angemessen auszuschöpfen.

(3) *Detailuntersuchungen:* Ergänzend zu diesen schriftlichen Befragungen finden qualitative Erhebungen in Form von halbstandardisierten Interviews statt. Beispielsweise wurden am Ende der Jahrgangsstufen 6 und 7 Interviewserien zur Bruchrechnung sowie am Ende der Jahrgangsstufen 8 und 9 zu Funktionen (jeweils N=36) durchgeführt. Dabei zeigt sich, dass die Defizite im Modellierungsbereich in erheblichem Teil auf nicht adäquat ausgebildete Grundvorstellungen bzw. systematische, mental bedingte Fehlstrategien zurückgeführt werden können. Eine typische

Fehlstrategie bei der Bruchrechnung ist die unbewusste Übertragung von Vorstellungen aus den natürlichen Zahlen auf den neuen Zahlbereich. Ein charakteristischer Fehler beim Umgang mit Funktionen ist das Deuten von Graphen als direktes Abbild der Realität. Die bei letztgenanntem Themengebiet diagnostizierten Fehler finden auch in einer parallelen Interviewstudie, die in Frankreich durchgeführt wurden, Bestätigung.

(4) *Perspektiven für eine individuelle Diagnose und Förderung:* Ausgehend von den bei PALMA gewonnenen Erkenntnissen wurden Konzepte zur systematischen Förderung der mathematischen Kompetenz des Modellierens bei Schülerinnen und Schülern generiert, die im Förderunterricht der sechsten Klassen an einem Bielefelder Gymnasium bei der Behandlung der Bruchrechnung erprobt wurden (sog. Projekt PALMA-I). Bei der Evaluation der Fördermaßnahme zeigen sich deutliche Leistungszuwächse der Förderschüler. Diese Zuwächse betreffen nicht nur den aktuellen Unterrichtsstoff, sondern auch zurückliegende Inhalte und finden zudem Bestätigung in schulischen Beurteilungssituationen.

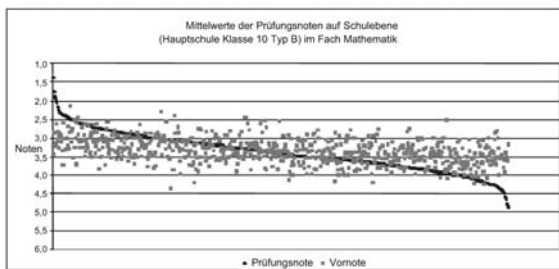
*Andreas Büchter, Andreas Pallack: „Zentrale Prüfungen in NRW: Schülerleistungen am Ende der Klasse 10“* 2007 wurden in Nordrhein-Westfalen erstmals zentrale Prüfungen zum Ende der Sekundarstufe I durchgeführt. Andreas Büchter und Andreas Pallack, beide Referenten am Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, berichteten über das Konzept und die Ergebnisse der Prüfungen. Informationen zum Verfahren sowie Beispielaufgaben stehen im Internet unter <http://www.standardsicherung.nrw.de/zp10/>.

Zur Evaluation der Zentralen Prüfungen 10 im Jahr 2007 wurden Daten in zwei Modulen (Modul I und Modul II) erhoben. Für Modul I gaben alle Schulen (für alle rund 190.000 Schülerinnen und Schüler) die Vornoten, Prüfungsnoten und Abschlussnoten sowie das Geschlecht der Schülerinnen und Schüler ein. Modul I dient primär der Rechenschaftslegung: Mit Hilfe dieser Daten können u. a. Mittelwerte berechnet und berichtet werden, die Indikatoren für den Ausfall der Prüfungen sind, um das Informationsinteresse einer breiten Öffentlichkeit zu bedienen.<sup>1</sup>

Die Referenten betonten, dass man sich bei der Deutung dieser Daten bewusst sein muss, dass Noten stark komprimierte Kennzahlen sind: Die konkrete Schülerarbeit wird mit Hilfe einer Auswertungsanleitung analysiert. Auf der Basis der Kriterien werden Punkte vergeben. Die Punkte

werden summiert und mit Hilfe eines Punkte-Notenschlüssels in Noten umgerechnet. Verändert man z. B. nur den Punkte-Notenschlüssel, so erhält man einen anderen Ausfall, obwohl die Anforderungen in den Aufgaben und die gezeigten Schülerleistungen selbst gleich geblieben sind. Die Festlegung von Kriterien, Punktverteilung und Notenschlüssel sind Ergebnis eines Diskurses, an dem viele Schulpraktiker beteiligt waren, da neben allgemeinen Aspekten der Leistungsmessung hier auch das Verständnis und die Anwendbarkeit durch die Lehrkräfte vor Ort eine besondere Rolle spielen.

Besondere Aufmerksamkeit fand die Analyse der Beziehung zwischen Vor- und Prüfungsnoten. So erscheint z. B. die Prüfung an Hauptschulen Typ B (mittlerer Abschluss) bei Betrachtung der Mittelwerte von Vornote (3,3) und Prüfungsnote (3,4) recht fair. Dahinter steckt die weit verbreitete Vorstellung, dass eine Prüfung (im Mittel) idealerweise die Vornote bestätigt. Heruntergebrochen auf einzelne Schulen ergibt sich jedoch ein differenzierteres Bild:



So findet man sowohl Schulen, an denen die Vorbenotung im Mittel bei „befriedigend“ lag ausreichende Prüfungen erzielten, als auch Kurse, die mit ausreichend vorbenotet wurden und in den Prüfungen einen befriedigenden Schnitt erreichten.

In Modul II wurde eine Stichprobe von 120 Schulen gezogen. Diese 120 Schulen gaben jeweils für eine Klasse bzw. Kurs die von den Schülerinnen und Schüler erzielten Punkte für jedes in den Auswertungsmanualen ausgewiesene Kriterium ein. Diese Daten bieten Möglichkeiten zur fachlichen Analyse. Die Referenten stellten ausgewählte Ergebnisse vor. Einen Schwerpunkt bildeten methodologische Überlegungen, die von der Deutung von Aufgaben bis hin zu Möglichkeiten und Grenzen der Interpretation von Raschmodellierungen auf der Basis von Prüfungsdaten reichten.

Die nächste Tagung des Arbeitskreises findet vom 25.–26. April 2008 zum Thema Kompetenzstufenmodelle in Soest statt.

Interessentinnen und Interessenten für die Tagung können sich bei der Leiterin des Arbeitskreises, Gabriele Kaiser, melden ([gabriele.kaiser@uni-hamburg.de](mailto:gabriele.kaiser@uni-hamburg.de)).

<sup>1</sup> Bei der Auswertung von Modul I wurde trotz der bekannten und berechtigten methodischen Kritik das arithmetische Mittel der entsprechenden Noten gebildet, da in Schulen und Öffentlichkeit genau dieser Wert „verlangt“ wird. Die Mittelwerte wurden jedoch mit Hilfe einer Modellrechnung auf der Basis der erreichten Gesamtpunktzahlen repliziert, wobei sich hohe Übereinstimmungen zeigten.