Neubrand, Michael; Klieme, Eckhard; Lüdtke, Oliver; Neubrand, Johanna

Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung

Unterrichtswissenschaft 30 (2002) 2, S. 100-119

Unterrichtswissenschaft

Zeitschrift für Lernforschung
30. Jahrgang / 2002 / Heft 1

Thema: PISA – Konzept und Ergebnisse

Verantwortliche Herausgeber:
Jürgen Baumert, Manfred Prenzel

Allgemeiner Teil

Michael Neubrand, Eckhard Klieme, Oliver Lüdtke,
Johanna Neubrand

Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung

Degrees of Competence and Models of Difficulty for the PISA-Test of Mathematical Literacy

Die nationalen Ergänzungen des Untersuchungsprogramms für PISA haben den mathematischen Teil der Studie wesentlich erweitert. Dies betrifft die theoretische Konzeption zur mathematischen Grundbildung - dargelegt in einem nationalen Rahmenkonzept, das das internationale Framework ergänzt und erweitert - ebenso wie die Aufgabensammlung und die Stichprobe; zudem wurden in Deutschland eine Reihe von Zusatzuntersuchungen zur curricularen und Konstrukt-Validität der Aufgaben durchgeführt. Im vorliegenden Beitrag wird auf der Basis dieser Ergänzungen das Konzept der mathematischen Grundbildung und seine Operationalisierung differenziert dargestellt. Getrennt nach der Art des mathematischen Arbeitens - Ausführung technischer Prozeduren, rechnerisches Modellieren und begriffliches Modellieren - wird untersucht, welche Anforderungsmerkmale die Schwierigkeit der Aufgaben bestimmen. Die Stufen mathematischer Kompetenz lassen sich damit aus didaktischer und psychologischer Sicht charakterisieren.

The German national supplements to the PISA study include a considerable extension to the mathematical domain of the assessment, in terms of the theoretical conception of mathematical literacy - presented in a national framework that complements and extends the international framework - as well as the items administered and the student sample drawn. A number of additional studies on the curricular and construct validity of the items were also conducted in Germany. In the present article, the national supplements are used as a basis for a differentiated presentation of the concept of mathematical literacy and its operationalization. The items are grouped according to the type of mathematical reasoning required - execution of technical procedures, procedural modelling, or conceptual modelling -and the features that determine the difficulty of the individual items are examined. This allows levels of mathematical literacy to be characterised from the didactic and psychological perspective.

Im Bericht zu PISA-2000 (Baumert & al. 2001; darin das Kapitel über den Mathematik-Teil des Tests: Klieme, Neubrand & Lüdtke 2001) wurde eine Anzahl von Beispielen aus dem nationalen und dem internationalen Teil des PISA-Mathematik-Tests vorgestellt. Die einzelnen Beispielaufgaben wurden mittels eines eindimensionalen Rasch-Modells auf einer Schwierigkeits- bzw. Fähigkeitsskala angeordnet. Die Einordnung der Items blieb dort - von einigen interpretierenden Bemerkungen abgesehen - auf der deskriptiven

Ebene. Der Frage, aufgrund welcher Anforderungsmerkmale eine Aufgabe schwieriger wird, soll hier nun systematisch nachgegangen werden. Ziel der Analysen ist es, die inhaltliche Bedeutung der Skala „mathematische Kompetenz“ aufzuklären. Es geht also, in den Begriffen der pädagogisch-psychologischen Diagnostik, um die Konstruktvalidität des Tests.

Voraussetzung für eine solche Analyse ist eine hinreichende Anzahl von Aufgaben, die - in formaler Hinsicht - über den gesamten Schwierigkeitsbereich hinweg streuen und - bei inhaltlicher Betrachtung - die mathematischen Anforderungen vielseitig abbilden. Eine entsprechende Aufgabenvielfalt konnte durch den zusätzlichen nationalen Test zu PISA-2000 bereit gestellt werden, wie Klieme, Neubrand und Lüdtke (2001; insbesondere Kap. 3., Abschnitt 5.) belegen. Der nationale Ergänzungstest bewirkte zweierlei: Auf der einen Seite rundete er den internationalen Test ab, indem nun auch sogenannte „technische Items“ (zur Bestimmung s.u.) einbezogen wurden, so dass eine stärkere inhaltliche Ausgewogenheit erreicht wurde (vgl. auch Tab. 1). Außerdem sicherte der Zusatztest, dass mehr Schülerinnen und Schülern Scores für den Bereich Mathematik zugewiesen werden konnten (Baumert & al. 2001, Abschnitt 1.4.).

Dass mit dieser breiteren Aufgabenbasis nun analytisch weitergearbeitet werden kann, sichern die Dimensionsanalysen, die in Klieme, Neubrand & Lüdtke (2001, Abschnitt 3) dargestellt wurden. Diese erlauben es, die internationalen und die nationalen Aufgaben zusammen als einen breit angelegten Test zu begreifen, der dennoch eine gemeinsame latente Fähigkeitsdimension abbildet. Es wird also im Folgenden die in Tab. 1 dargestellte Aufgabenmenge als Grundgesamtheit für die Analysen genommen:

Tabelle 1:
Verteilung der PISA-Items

	Typen mathematischen Arbeitens			
	Technische Items	*rechnerische Modellierungsaufgaben*	*begriffliche Modellierungsaufgaben*	
Internationale PISA-Aufgaben	1	14	16	31
Nationale PISA-Aufgaben	23	33	30	86
	24	47	46	117

1. Zur inhaltlichen Breite des PISA-Gesamttests

Zentrale Idee von PISA ist es, in allen Domänen Grundbildung im Sinne von „literacy“ zu erfassen. Unter „mathematical literacy“ wird in den internationalen PISA-Dokumenten (OECD 1999) verstanden, Mathematik in unterschiedlichen Zusammenhängen einsichtsvoll und funktional verwenden zu können. Somit besteht mathematische Kompetenz für PISA nicht nur aus der

Kenntnis mathematischer Sätze und Regeln und der Beherrschung mathematischer Verfahren. Mathematische Kompetenz zeigt sich vielmehr im verständnisvollen Umgang mit Mathematik und in der Fähigkeit, mathematische Begriffe und Verfahren als „Werkzeuge" in einer Vielfalt von innermathematischen und außermathematischen Kontexten einzusetzen. Dahinter stehen - im internationalen Framework (OECD 1999) auch explizit angesprochen - die Ideen von Hans Freudenthal (1977; 1983). Ziel der Verankerung mathematischer Begriffe in Kontexten ist die Ausbildung tragfähiger „mentaler Modelle" (Freudenthal 1983), und dies grenzt mathematical literacy deutlich gegen ein rein instrumentelles Verständnis von Mathematik ab (vgl. auch de Lange 1996, Neubrand 2001).

Eine Besonderheit dieses Ansatzes ist die Betonung des „Modellierens". Demnach besteht das Bearbeiten mathematischer Aufgaben aus einem in der Grundstruktur zyklischen Prozess von „Erkennen einer Problemstellung" → „Übertragen in einen mathematischen Ansatz" → „Verarbeiten dieses Ansatzes" → „Interpretieren der Ergebnisse und ggf. Überprüfung der Adäquatheit des gewählten Ansatzes" (vgl. Abb. 3.1. in Klieme, Neubrand & Lüdtke 2001). Diese Auffassung wird in der Mathematikdidaktik inzwischen weitgehend geteilt, jedenfalls soweit es um die Verwendung von Mathematik in Anwendungszusammenhängen geht („mathematical modelling": Blum 1996; vgl. auch die nähere Beschreibung bei Klieme, Neubrand & Lüdtke 2001).

Man kann die Idee des Modellierens von einer allgemeineren Warte aus betrachten, ohne Bindung an Anwendungskontexte, indem man die kognitiven Prozesse beim Lösen der Aufgaben fokussiert. Aus dieser Sicht kann man das Aufgabenlösen auch dann als Modellierungsprozess ansprechen, wenn der Kontext innermathematischer Natur ist. Die spezifischen Wissensinhalte, Heuristiken und Operationen, die beim Mathematisieren, d.h. beim Übertragen des vorgelegten innermathematischen Problems in einen Lösungsansatz eingesetzt werden, können durchaus andere sein als bei außermathematischen Kontexten. Dennoch bleiben die kognitiven Prozesse im Kern gleich: Auch bei innermathematischen Aufgaben, die nicht nur algorithmisches Abarbeiten eines schon vorgegebenen Ansatzes verlangen, muss im Kern eine „Übersetzungsleistung" erbracht werden. In beiden Fällen muss aus einer Problemstellung heraus ein Weg der Lösung („Ansatz") gefunden werden. In diesem Sinn verwendet PISA das Bild des Modellierungsprozesses als allgemeines „Modell" für den Prozess des Aufgabenlösens.

Die Auffassung, mathematisches Aufgabenlösen als Modellierungsprozess zu betrachten, hat weitere kognitionspsychologische Implikationen. Diese betreffen u.a. Überlegungen zu den Prozessen der Übertragung von Texten in ein mathematisches Modell (Reusser 1992) sowie die Differenzierung nach der „Art des Wissens" (Hiebert 1986; Neubrand & al. 2001; J. Neubrand, im Druck). Das zur Lösung der Aufgabe notwendige Wissen kann dahin unterschieden werden, ob es „prozedural", d.h. auf die Durchführung mathematischer Verfahren bezogen, oder „begrifflich", d.h. auf die Verwendung mathematischer Zusammenhänge hin ausgerichtet ist („procedural and conceptual knowledge“ im Sinne von Hiebert 1986). Modellierungsaufgaben im oben geschilderten verallgemeinerten Sinne können danach unterschieden werden, ob in ihnen (a) eher prozedurales oder (b) eher begriffliches Wissen benötigt wird. Aufgaben des ersten Typs sollen hier als „rechnerische Modellierungsaufgaben“ bezeichnet werden. Sie verlangen im Anschluss an die Mathematisierung eine algorithmisch-prozedurale Durchführung des gewonnenen Ansatzes. Typische Beispiele sind die „klassischen“ Textaufgaben, von eingekleideten Aufgaben bis hin zu komplexeren Anwendungsproblemen, bei denen im allgemeinen ein Ansatz für eine geeignete „Rechnung“ zu finden ist. Aufgaben des zweiten Typs sollen hier als „begriffliche Modellierungsaufgaben“ bezeichnet werden. Hier kann die Modellierung mittels begrifflicher Mittel - ohne algorithmische Schritte - zu Ende gebracht werden. Zu diesen Mitteln gehören Verweise auf begrifflich geprägte Zusammenhänge, was bis zur strukturellen Verallgemeinerung einer Situation[1] führen kann, oder Argumentationen, Aufstellen einer Systematik, Entwerfen einer umfassenden Strategie, etc., jedenfalls nicht nur das Abarbeiten von festen Verfahren.

Es gibt dann außer den „Modellierungsaufgaben“ lediglich noch solche Aufgaben, bei denen es ausschließlich darauf ankommt, einen schon vorgegebenen Ansatz aufgrund bekannter Algorithmen oder Prozeduren abzuarbeiten. Diesen Typ von Aufgabe kann man als „technische Aufgabe“ bezeichnen. Typische Beispiele sind Aufgaben wie „Löse die Gleichung ... !“.

Von einer Differenzierung nach diesen drei Aufgabengruppen

- technische Aufgaben,
- rechnerische Modellierungsaufgaben,
- begriffliche Modellierungsaufgaben

wird in der folgenden Analyse ausgegangen. Sie verlangen drei qualitativ unterschiedliche *Typen mathematischen Arbeitens*[2]. Zu untersuchen ist nun, welche Aufgabenmerkmale im Einzelnen die Schwierigkeit von Aufgaben aus den unterschiedlichen Typen bedingen. Dazu werden Anforderungsmerkmale bestimmt, die sich aus der Auffassung, Aufgabenlösen sei ein Modellierungsprozess, heraus bestimmen lassen.

2. Zur Systematik der Anforderungsmerkmale

Die im Folgenden beschriebene Systematik fußt im weitesten auf dem internationalen und dem nationalen Rahmenkonzept von PISA (OECD 1999;

---

1 Typische Frage: „Ist das immer so?“

2 Im nationalen PISA-Framework (Neubrand & al. 2001) sind sogar fünf Klassen von Items unterschieden. Die jetzt vorgenommene Dreier-Gliederung stellt einen Zusammenschnitt der dortigen fünf „Kompetenzklassen“ dar. Im nationalen Framework findet man auch Beispielaufgaben zu den unterschiedlichen Typen mathematischen Arbeitens.

Neubrand et al. 2001). Leitender Gedanke ist dabei das Konzept der mathematischen Modellierung. Ferner gehen Erfahrungen aus der fachdidaktischen Analyse von TIMSS-Aufgaben (Neubrand, Neubrand & Sibberns 1998, Blum & Wiegand 1998, Baumert, Klieme & Watermann 1999, Klieme 2000), aus der Analyse mathematischer Aufgaben innerhalb der TIMS-Videostudie (J. Neubrand, in Druck) sowie aus anderen mathematikdidaktischen und psychologischen Arbeiten zur Aufgabenanalyse ein (Christiansen & Walther 1986, Klieme 1989, Bromme, Seeger & Steinbring 1990, Stein, Grover & Henningsen 1996, Williams & Clarke 1997).

Jede der 117 Aufgaben - nationale wie internationale - wurde im Hinblick auf die ausgewählten Merkmale eingeschätzt. Einen wesentlichen Teil dieser Arbeit hat die PISA-Expertengruppe Mathematik geleistet. Weitere Einschätzungen wurden von Lehrerinnen und Lehrern aus dem Raum Berlin-Brandenburg vorgenommen, die in der Entwicklung von Testaufgaben erfahren waren. Sie wurden in einer ausführlichen Schulung[3] mit dem Konzept des Modellierungsprozesses und den darin gekennzeichneten Merkmalen einer Aufgabe vertraut gemacht und schätzten dann unabhängig voneinander alle 117 Aufgaben anhand eines von den Autoren entwickelten Kategorienschemas ein[4].

Unter den vielen Anforderungsmerkmalen, die Aufgaben tragen (J. Neubrand, im Druck), wurden solche ausgewählt[5], die das Potential einer Aufgabe im Hinblick auf grundlegende Ausrichtungen des Mathematikunterrichts charakterisieren. Einerseits soll der curriculare Aufbau des Mathematikunterrichts berücksichtigt werden (hierzu das Merkmal *curriculare Wissensstufe).* Andererseits werden aus psychologischer Sicht verschiedene Merkmale berücksichtigt, von denen abhängt, wie weit beim Lösen einer Aufgabe der Suchraum zu ziehen ist. Hierzu zählen die Komplexität der erforderlichen Modellierungs- bzw. Problemlöseaktivitäten sowie der Kontext und ausgewählte mathematische Tätigkeiten. Die im Folgenden näher beschriebenen Merkmale sind also zentral, weil sie sowohl auf Anwendungs- bzw. Strukturorientierung (Winter 1995) reagieren *(Kontext, Argumentieren),* als auch auf

---

[3] Die Schulung wurde von J. Neubrand, M. Neubrand und E. Klieme konzipiert und durchgeführt, vor allem auf der Basis der von J. Neubrand entwickelten Kategorien für Aufgabenanalysen zur TIMS-Videostudie. Wir danken den Teilnehmerinnen und Teilnehmern und insbesondere Herrn Götz Bieber (Pädagogisches Landesinstitut Brandenburg) für seine Unterstützung dieser Studie.

[4] Die Teilnehmerinnen und Teilnehmer stimmten nach der Schulung in hohem Maße überein: 80 Prozent der Urteile werden von mindestens 6 der 8 Rater geteilt.

[5] Spätere Analysen, z.B. im zu PISA-2000 geplanten „Thematischen Bericht Mathematik", werden mathematikdidaktische Fragen vertieft untersuchen und dabei noch feiner nach Typen des mathematischen Arbeitens unterscheiden.

die Spannung zwischen Fertigkeits- und Fähigkeitsorientierung *(Komplexität der Modellierung, Umfang der Verarbeitung)* und auf heuristische Potentiale in der Aufgabe *(Offenheit)*.

2.1 Übergreifende Merkmale des Modellierungsprozesses

Oben wurde dargestellt, dass das Lösen von Aufgaben essentiell als ein Modellierungsprozess interpretiert werden kann. Dies liefert eine „Sprache", mit der Aufgaben und ihre Merkmale beschrieben werden können. Im Folgenden nennen wir zunächst jene Anforderungsmerkmale, die den Modellierungsprozess als Ganzes betreffen. Der nachfolgende Abschnitt stellt dann Merkmale vor, die jeweils wichtige Aspekte einzelner Teilprozesse betreffen.

Komplexität der Modellierung
Eine umfassende Charakterisierung des mathematikdidaktischen Anspruchs einer Aufgabe stellt die Einordnung in Kompetenzklassen (Competency Classes) dar, wie sie im internationalen Rahmenkonzept (Framework) von PISA formuliert sind (OECD 1999). Dort werden drei Ausprägungen (Kompetenzklassen) unterschieden:

Reproduktion
Aufgaben aus dieser Klasse verlangen im Wesentlichen Faktenkenntnisse und das Ausführen von Routinen, die als Standardverfahren aus dem Unterricht bekannt sind. Beispiele: „Rennwagen"[6], Fragen 2 und 3.

Verknüpfung
Zu dieser Klasse gehören Problemstellungen, zu deren Lösung man Querverbindungen zwischen verschiedenen Stoffgebieten und Teilbereichen der Mathematik herstellen und Informationen verknüpfen muss. Die Schülerinnen und Schüler müssen Entscheidungen über passende Modelle und Verarbeitungsstrategien fällen. Beispiele: „Äpfel", Fragen 1 und 2; „Rennwagen", Fragen 1 und 5.

Verallgemeinerung
Bei Aufgaben dieser Klasse ist einsichtsvolles mathematisches Denken gefordert. Die Bearbeiter müssen eigene Modelle entwickeln, diese verallgemeinern und über den Modellierungsprozess reflektieren. Beispiel: „Äpfel", Frage 3.

Dies ist eine eher integrative Sicht auf die Items. Die Anforderungen an den Modellierungsprozess insgesamt steigen entlang der drei Klassen: Die Mo-

---

6 Die Aufgaben „Rennwagen" und „Äpfel" können unter „Internationale Beispielaufgaben" bei www.mpib-berlin.mpg.de/pisa eingesehen werden.

dellierung wird zunehmend eigenständiger, offener, selbständiger und erfordert in wachsendem Maße komplexe Konstruktionsleistungen. Wir betrachten daher diese Klasseneinteilung als ein grundlegendes und allgemeines Maß des Anspruchs an den Modellierungsprozess und prüfen - auch wenn die Autoren des internationalen Rahmenkonzepts dies nicht explizit tun -, ob „Reproduktion", „Verknüpfung" und „Verallgemeinerung" eine Hierarchie mit zunehmendem Schwierigkeitsgrad bilden, und wenn ja auf welcher Itemgruppe diese wirkt. Es ist von vornherein klar, dass innerhalb der Gruppe der technischen Items diese Variable keinen Einfluss auf die Schwierigkeit haben kann, denn alle technischen Items verlangen nur die Reproduktion gelernter Verfahren, so komplex diese in einer Einzelaufgabe auch sein mögen.

Curriculare Wissensstufe

Mit diesem Merkmal wird beschrieben, wie anspruchsvoll in einem technischen Sinn das mathematische Wissen ist, das man zur Lösung einer Aufgabe benötigt. Es kommt also darauf an, in welchen curricularen Zusammenhang der in der Aufgabe explizit angesprochene Stoff gehört, d.h. auf welcher Stufe des Curriculums man diese Anforderungen gewöhnlich gelernt hat. Es werden drei Stufen unterschieden:

Grundkenntnisse

Hier werden lediglich die Grundrechenarten und einfachste geometrische Grundkenntnisse vorausgesetzt, die bereits in der Grundschule vermittelt werden oder aus dem Alltag bekannt sind.

Einfaches Wissen der Sekundarstufe I

Bei Aufgaben dieser Ausprägung werden Begriffe der Sekundarstufe I verwendet, wie etwa der Bruchzahlbegriff oder der Prozentbegriff, die als Grundbestandteile einer „Mathematik für alle" (Heymann 1996) gelten müssen und die in jeder Schulform zum Basiscurriculum gehören.

Anspruchsvolles Wissen der Sekundarstufe I

Hierzu gehören fortgeschrittene Verfahren und Begriffe, die in PISA bis zu quadratischen Gleichungen in der Algebra und zu Anfängen der Ähnlichkeitsgeometrie gehen. Solche Aufgaben erfordern einen besonders hohen Grad an fachsystematischem Wissen.

Erwartet wird, dass die curriculare Herkunft des Wissens bei allen drei Aufgabengruppen einen Einfluss auf die Schwierigkeit hat, sozusagen per se sollte dies bei den technischen Aufgaben besonders sichtbar werden.

Kontexte

Modellierungsprozesse beginnen mit „Situationen", also innerhalb bestimmter Kontexte. Im deutschen PISA-Ergänzungstest sollten, wie oben erläutert, neben außermathematischen Kontexten auch Modellierungen innerhalb der

Mathematik in Betracht gezogen werden. Hinzu kommt als dritte Kategorie, die beispielsweise bei den technischen Items die Regel darstellt, die Angabe „ohne Kontext“.

Zu betonen ist, dass in dieser Analyse nicht über Kontexte an sich argumentiert werden kann, sondern nur über die Kontexte in einem für den PISA-Test konstruierten Aufgabenset.

2.2 *Merkmale der Teilprozesse innerhalb des Modellierungsprozesses*

Offenheit der Mathematisierung, vielfältige Lösungen

Das Aufstellen eines mathematischen Modells ist eine konstruktive kognitive Leistung (Reusser 1992). Sie ist um so anspruchsvoller, je offener die Aufgabe gestellt ist, je mehr Spielräume sich also für die Mathematisierung eröffnen. Es werden daher alle PISA-Aufgaben danach eingeschätzt, ob sie mehrere unterschiedliche Ansätze der Mathematisierung zulassen, also ob es unterschiedliche Modelle und somit vielfältige Lösungsmöglichkeiten (Neubrand & Neubrand 1999) gibt.

Umfang der Verarbeitung

Die Komplexität der Verarbeitungsprozesse innerhalb eines Modells hängt nicht zuletzt davon ab, ob und in welchem Umfang neue Größen und Zwischenergebnisse, die noch nicht in der Aufgabenstellung selbst vorgegeben sind, im Verlauf der Lösung der Aufgabe eingeführt werden müssen. Wir nennen solche Größen die „impliziten Größen“, die bei der Verarbeitung - in einem möglichst ökonomischen Lösungsgang - anfallen, also sozusagen der Aufgabe implizit innewohnen. Zur tabellarischen Übersicht wird aus dieser Variablen ein dichotomes Merkmal „Umfang der Verarbeitung: niedrig / hoch“ gebildet, wobei „hoch“ bei mehr als zwei impliziten Größen zutrifft.

Argumentieren

Mit diesem Merkmal (dichotom) wird ein spezieller Verarbeitungsprozess festgehalten. Es wird beurteilt, ob die Verarbeitung des Modells explizit mathematisches Argumentieren erfordert. Dabei kommt es nicht darauf an, in welcher Form das Argument zu präsentieren ist. Argumentationen können von formalen Beweisen bis zu informellen Begründungen reichen.

Die folgende Tabelle gibt wieder, in welchen Anzahlen die angeführten Merkmale im Grundbestand der 117 PISA-Items vorkommen. Die Tabelle 2 macht abermals deutlich, dass der nationale Ergänzungstest notwendig war, um Spielraum für Schwierigkeitsanalysen zu gewinnen. Allein mit den internationalen Aufgaben wäre das Spektrum der Anforderungen zu eng, um Einflussmerkmale finden zu können.

Tabelle 2:
Anforderungsmerkmale bei den PISA-Aufgaben

Merkmal	*Anzahl unter den internationalen Aufgaben*	*nationalen Aufgaben*
Komplexität der Modellierung (Internationale Competency Classes)		
• Reproduktion	9	41
• Verknüpfung	20	41
• Verallgemeinerung	2	4
Curriculare Wissensstufe		
• Grundkenntnisse	6	21
• einfache Begriffe Sek.I	15	25
• fortgeschrittene Begriffe Sek I	10	40
Kontexte (Mehrfachzuweisungen möglich[7])		
• außermathematisch	23	43
• innermathematisch	10	23
• „ohne"	1	16
Offenheit: Multiple Lösbarkeit vorhanden:	10	40
Erhöhter Umfang der Verarbeitung:	4	21
Argumentieren erforderlich:	3	6
Gesamt	31	86

3 Stufen der mathematischen Kompetenz in PISA

Im Folgenden soll nun untersucht werden, welcher Zusammenhang zwischen den bisher beschriebenen Anforderungsmerkmalen und der empirisch ermittelten Schwierigkeit einer Aufgabe besteht. Damit beschreiben wir zugleich, worin sich Schülerinnen und Schüler, die hohe Testwerte erreichen, von Personen mit niedrigeren Testwerten unterscheiden. Personen mit ihren Fähigkeiten und Aufgaben mit ihren Schwierigkeitskennwerten werden nämlich bei der Konstruktion des PISA-Tests auf ein und derselben Skala angeordnet (Baumert & al. 2001; Abschnitt 1.4.). Will man inhaltlich verstehen, wodurch sich Personen mit hoher (resp. niedriger) mathematischer Kompetenz im Sinne von PISA auszeichnen, muß man demnach die Anforderungsmerkmale der schwierigen (resp. leichten) Aufgaben betrachten. Ein Erklärungsmodell für die Schwierigkeit von Testaufgaben, wie wir es im

[7] Es ist ein charakteristisches Kennzeichen gerade der internationalen Items, die aus dem Umfeld von Freudenthals Ideen kommen, dass oft zusammen mit einer außermathematischen Anknüpfung eine innermathematische „Struktur" präsentiert wird. Beispielsweise folgt beim Aufgabenset „Äpfel" sogleich der Situationsbeschreibung über die Anordnung der Apfel- und Nadelbäume eine schematische Skizze, die das Problem vom außermathematischen Bezug in den innermathematischen Kontext der geometrischen Muster stellt.

Folgenden entwickeln, kann daher genutzt werden, um unterschiedliche Ausprägungen der mathematischen Kompetenz von Schülerinnen und Schülern inhaltlich zu beschreiben.

Zur Verdeutlichung der qualitativen Abstufungen in der Schwierigkeit von Aufgaben bzw. in der mathematischen Kompetenz ihrer Bearbeiter unterscheiden wir fünf Niveaus, die im Folgenden unter Rückgriff auf typische Anforderungsmerkmale charakterisiert werden sollen. Die Stufen selbst wurden - ebenso wie die Stufen der Lesekompetenz - vom internationalen PISA-Konsortium (vgl. OECD 2001) abgegrenzt[8]. Der internationale Bericht verzichtet allerdings auf eine inhaltliche Beschreibung aller fünf Stufen, weil dies allein aufgrund der 31 international verwendeten Testitems nicht möglich ist. Indem wir unsere 86 Ergänzungsaufgaben ebenfalls auf der PISA-Skala verankern - gerechtfertigt durch den Befund, dass beide Testteile eine gemeinsame Dimension bilden (dargestellt in Klieme, Neubrand & Lüdtke 2001) -, erhalten wir die Möglichkeit, diese Stufen näher zu charakterisieren. Konkrete Aufgaben werden einen Eindruck davon vermitteln, wovon die Schwierigkeit einer Aufgabe abhängen könnte. Allerdings muss betont werden, dass es sich hierbei nicht um eine Modellierung kognitiver Prozesse handelt, d.h. über psychologische Kausalität nichts gesagt werden kann.

3.1 Einfluss von Aufgabenmerkmalen auf die Schwierigkeit - generell

Welcher Zusammenhang besteht nun im Einzelnen zwischen Anforderungsmerkmalen und Aufgabenschwierigkeit? Indem wir die eingeschätzten Merkmale mit der empirisch ermittelten Schwierigkeit der Aufgaben in Beziehung setzen, können wir rekonstruieren, was eine PISA-Mathematikaufgabe mehr oder weniger schwierig macht.[9]

Generell gilt: Alle in Tabelle 2 genannten Merkmale, und auch weitere, sind schwierigkeitsrelevant. Und es ist durchaus möglich, etwa durch eine Regres-

---

8 Die Logik der Definition von Stufen war dieselbe wie bei der Lesekompetenz: Das gesamte Fähigkeitsspektrum, das in den Testaufgaben abgebildet ist, sollte in fünf gleich große Abschnitte unterteilt werden. Wie breit die Abschnitte sind, ergibt sich aus der Forderung, dass alle Schüler, die zu einer bestimmten Kompetenzstufe gehören, mindestens etwa 50 % der Aufgaben dieser Stufe lösen können. Da die Fähigkeitsskala in der Mathematik stärker gestreckt ist als beim Lesen, sind die Kompetenzstufen hier breiter. Sie umfassen jeweils etwa 90 Punkte, im Lesen hingegen etwa 70 Punkte.

9 Statistisch stützen sich die folgenden Angaben auf multiple Regressionsanalysen mit n=117 Aufgaben, in die Aufgabenmerkmale als Prädiktoren und der gemäß Rasch-Modell bestimmte Schwierigkeitskennwert als abhängige Größe eingingen. Die Schwierigkeitskennwerte wurden aufgrund der Antworten von etwa 8000 deutschen Schülerinnen und Schülern ermittelt.

sionsanalyse, das Ausmaß der Schwierigkeitsrelevanz einzelner Merkmale zu bestimmen. Dann ergibt sich als generelles Bild, dass die beiden Merkmale „Komplexität des Modellierungsprozesses" und „curriculare Wissensstufe" den größten Einfluss haben. Diese beiden Merkmale erklären jeweils einzeln 16 bzw. 22 Prozent und gemeinsam 34 Prozent der Schwierigkeitsvarianz zwischen allen Aufgaben. Berücksichtigt man außerdem die gewählten speziellen Aspekte des Modellierungsprozesses (Offenheit und Umfang der Verarbeitung) und die Art des Kontextes (inner-vs. außermathematisch), so lassen sich insgesamt ca. 45 % der Schwierigkeitsvarianz über alle Items hinweg aufklären.

3.2 Einfluss von Aufgabenmerkmalen auf die Schwierigkeit: Unterschiedliche Einflüsse je nach Typ des mathematischen Arbeitens

Aus mathematikdidaktischer Sicht ist ein solches allgemeines Modell zur Rekonstruktion der Aufgabenschwierigkeit allerdings unbefriedigend. Denn offenbar ist es so, dass einige Merkmale nur sinnvoll auf diejenigen Aufgaben bezogen werden können, die bestimmte Typen mathematischen Arbeitens ansprechen. So kann auf den rein technischen Berechnungsaufgaben das Merkmal der Offenheit der Modellierung nicht wirken, und für begriffliches Modellieren wird die Zahl der durchzuführenden Schritte nur schwer verlässlich zu bestimmen sein, weil multiple Lösungswege gerade hier sehr unterschiedliche Verarbeitungsumfänge bedeuten. Kurz: Wenn in einer Aufgabe andere kognitive Aktivitäten verlangt werden, werden auch andere Merkmale diesen Prozess determinieren.

Dies überprüfen empirisch die nun folgenden Analysen. Die jeweils gleichen Regressionsansätze mit den oben beschriebenen zentralen sechs Variablen werden verwendet, um die Aufgabenschwierigkeit innerhalb der drei Typen des mathematischen Arbeitens zu rekonstruieren. Tatsächlich ergeben sich differenzierte Erklärungsansätze, wenn man technische Items, rechnerische und begriffliche Modellierungsaufgaben getrennt betrachtet. Statistisch gesprochen: Die Regressionsgewichte (standardisierte Beta-Koeffizienten), die ausdrücken, welchen Beitrag die einzelnen Anforderungsmerkmale zur Aufklärung der Aufgabenschwierigkeit leisten, sind unterschiedlich groß. Jeweils unterschiedliche Gruppen von Variablen bilden das beste, d.h. sparsamste und erklärungsstärkste Modell[10]. Diese unterschiedlichen Modelle sollen im Folgenden vorgestellt werden. Dabei stellen wir jeweils zunächst das Schwierigkeitsmodell vor, beschreiben dann die Aufgabenmerkmale auf den einzelnen Kompetenzstufen und nennen schließlich eine Reihe von Beispielaufgaben.

[10] Das sparsamste Modell wurde - getrennt für die drei Typen - mit der Strategie de schrittweisen Regression im Programmpaket SPSS bestimmt.

Schwierigkeitsgenerierung bei den technischen Items

Die Schwierigkeit *technischer Aufgaben* hängt praktisch allein davon ab, auf welchem curricularen Wissensniveau die jeweils geforderten Routinen angesiedelt sind. Die schrittweise Regressionsanalyse ergibt ein Modell, das nur diese eine Variable enthält. Der standardisierte Beta-Koeffizient ist mit .591 sehr hoch. Selbst das Merkmal „Umfang der Verarbeitung" leistet keinen signifikanten Beitrag zur Aufgabenschwierigkeit und bleibt daher in dem Endmodell unberücksichtigt.

Schwierigkeitsgenerierung auf den technischen Aufgaben

Merkmal	*Standardisiertes Beta*	*Signifikanzniveau*	*Erklärte Varianz der Aufgabenschwierigkeit (R^2)*
Curriculare Wissensstufe	.591	.002	.319

Aus heuristischen Gründen wird hier und im Folgenden jeweils eine Tabelle angefügt, die die Verteilung wichtiger Merkmale auf den einzelnen Kompetenzstufen aufzeigt:

Tabelle3:
Schwierigkeitsstufung bei den Technischen Items. Angegeben sind in der rechten Spalte die Anzahlen der Items pro Kompetenzstufe, die die entsprechende Merkmalsausprägung besitzen

Kompetenz-Stufe	Kurzbeschreibung	Curriculare Wissensstufe: *Grundk. / einf. / fortgeschr.*
V (1 Items)	Höhere algebraische Techniken	0 / 0 / 1
IV (3 Items)	Mathematische Grundtechniken (wie Termumformungen)	0 / 0 / 3
III (11 Items)	Umgehen mit den mathematischen Grundbegriffen der Sekundarstufe I.	2 / 3 / 6
II (8 Items)	Wissen über einfache Fakten und Konventionen.	4 / 2 / 2
I (1 Item)		1 / 0 / 0

Ein typisches Beispiel für die Stufe I ist eine Aufgabe im multiple-choice-Format, bei der die Fläche eines angezeichneten Rechtecks mit den Seitenlängen 3 cm bzw. 4 cm zu berechnen ist. Das Item

Berechne und kreuze die richtige Lösung an! $4 + 3 \cdot (2+1) =$ *MC-Format mit den Abstufungen: 11/ 13 / 14 / 15 / 21*

benutzt einfache Konventionen (Stufe II), während Umgehen mit dem zentralen Basisbegriff „Funktion" (Stufe III) im Item

Die Funktion mit der Gleichung y = 2x-1 soll untersucht werden. Berechne für x = 100 den y-Wert *offen: „kurze Antwort*

zum Ausdruck kommt. Die Steigerung des curricularen Anspruchs in dieser Itemgruppe ging im nationalen PISA-Test bis zum Lösen einer quadratischen Gleichung (Stufe V).

Schwierigkeitsgenerierung bei den rechnerischen Modellierungsaufgaben

Beim *rechnerischen Modellieren* kommen zusätzlich zur curricularen Wissensstufe die Komplexität der Modellierung (mit den Abstufungen Reproduktion - Verknüpfung - Verallgemeinerung) sowie der Umfang der Verarbeitung ins Spiel. Im Einzelnen ergibt sich dieses Modell:

Schwierigkeitsgenerierung auf den rechnerischen Modellierungsaufgaben

Merkmal	*Standardisiertes Beta*	*Signifikanzniveau*	*Erklärte Varianz der Aufgabenschwierigkeit (R^2) im Gesamtmodell*
Komplexität der Modellierung	.248	.033	
Curriculare Wissensstufe	.455	.000	
Außermathematischer Kontext vorhanden	- .208	.070	
Umfang der Verarbeitung	.283	.016	
			48,9 %

Das sparsamste und erklärungsstärkste Regressionsmodell ist nun vielschichtiger als im Fall der technischen Items. Der Umfang der Verarbeitung hat mit Beta = .283 nun erkennbar Einfluss auf die Aufgabenschwierigkeit, ebenso die Komplexität der Modellbildung mit Beta = .248. Die rechnerischen Modellierungsaufgaben mit außermathematischem Kontext sind tendenziell bei den leichteren Aufgaben zu finden.

Die folgende Tabelle weist wieder zur Demonstration die absoluten Anzahlen der einzelnen relevanten Merkmalsausprägungen aus:

Tabelle 4:
Schwierigkeitsstufung bei den Rechnerischen Modellierungsaufgaben

Kompetenz-Stufe	Kurzbeschreibung	Komplexität der Modellierung *Reproduktion / Verknüpfung / Verallgemeinerung*	Curriculare Wissensstufe: *Grundk. / einf. / fortgeschr.*	*Außermathematischer Kontext* *Nicht vorh. / vorhanden*	Umfang der Verarbeitung *niedrig / hoch*
V (7 Items)	Komplexe Modellierung auf hohem curricularem Niveau	2 / 4 / 1	0 / 2 / 5	4 / 4	5 / 2
IV (10 Items)	Komplexe Modellierungen mit einfachen schulischen Kenntnissen	3 / 7 / 0	1 / 6 / 3	3 / 7	6 / 4
III (14 Items)	Modellierungen mit einfachen schulischen Kenntnissen	8 / 6 / 0	1 / 11 / 2	3 / 11	13 / 1
II (12 Items)	Standard-mathematisierungen	6 / 6 / 0	6 5 1	1 / 11	11 / 1
I (4 Items)		4 / 0 / 0	3 0 1	0 / 4	4 / 0

An Beispielen sind im PISA-Bericht (Klieme, Neubrand & Lüdtke 2001) bereits die folgenden Aufgaben genannt und teilweise auch diskutiert:

Für Stufe I kann die Aufgabe „Brötchen" als typisch gelten:

> 7 Brötchen kosten 3,15 DM. Was kosten 11 Brötchen?
> *MC: 5,05 DM / 4,95 DM / 4,85 DM / 4,75 DM / 4,65 DM*

Für die Stufe II können die ersten beiden Varianten von „Glasfabrik", für Stufe III die dritte Variante als Beispiele dienen. Diese Aufgabenserie prüft bei konstant gehaltener Situation die Grundaufgaben der Prozentrechnung ab.

> *Variante 1:*
> Eine Glasfabrik stellt am Tag 8000 Flaschen her. 2 % der Flaschen haben Fehler. Wie viele sind das?
> MC: 16 / 40 / 80 / 160 / 400 Flaschen
>
> *Variante 2:*
> Eine Glasfabrik stellt Flaschen her. 2 % der Flaschen sind fehlerhaft. Wie viele Flaschen wurden insgesamt hergestellt?
> MC: 320 / 800 / 3200 / 8000 / 12500 Flaschen
>
> *Variante 3:*
> Eine Glasfabrik stellt am Tag 8000 Flaschen her. Erfahrungsgemäß sind ca. 160 Flaschen fehlerhaft. Wie viel Prozent sind das?
> MC: 0,02 % / 0,5 % / 1,28 % / 2 % / 5 %

Überhaupt sammeln sich in Stufe III viele Aufgaben, die man als einfache Aufgaben, die auf ein lineares Modell („Proportionalität") führen, bezeichen kann. Man kann diese Aufgaben sicher als „Standardmodellierungen mit einfachen Begriffen" beschreiben. Die Variante 3, die Berechnung des Prozentsatzes, setzt in etwas höherem Maße als die beiden übrigen Varianten Kenntnisse der Prozentrechnung voraus. Dies schlägt sich offenbar in der höheren Schwierigkeit nieder. Interessanterweise sind die Fehlantworten weniger auf die Vertauschung zweier zentraler Konzeptualisierungen des Prozentbegriffs (Operator 0,02 und Prozentzahl 2 %) zurückzuführen, sondern auf falsche rechnerische Ansätze, die in den Distraktoren 5 %, 0,5 % bzw. 1,28 % (Division bzw. Multiplikation der Zahlenangaben und „plausibles" Setzen des Kommas) zum Ausdruck kommen.

Das Item „Sparen" markiert die Grenze des Übergangs zur (höchsten) Kompetenzstufe V:

> Karina hat 1000 DM in ihrem Ferienjob verdient. Ihre Mutter empfiehlt ihr, das Geld zunächst bei einer Bank für 2 Jahre festzulegen (Zinseszins!). Dafür hat sie zwei Angebote:
>
> a) „Plus-Sparen": Im ersten Jahr 3 % Zinsen, im zweiten Jahr 5 % Zinsen.
> b) „Extra-Sparen": Im ersten und zweiten Jahr jeweils 4 % Zinsen.
>
> Karina meint: „Beide Angebote sind gleich gut!" Was meinst Du?
> Begründe Deine Antwort.
>
> (offen)

Noch gehen bei diesem Item die curricularen Erfordernisse nicht über die einfachen Begriffe der Sekundarstufe I hinaus, jedoch wird nun komplexes Modellieren durch den Vergleich beider Sparmodelle erforderlich. Die Durchführung des Modells erfordert sowohl die Hintereinanderausführungen von Berechnungen, als auch zusätzlich den Vergleich der beiden Resultate. Die „Mathematisierung" ist folglich recht anspruchsvoll, die einzelnen Schritte der Deduktion sind aber curricular noch standardmäßig, allerdings bei hohem Umfang der Verarbeitung. Auch ist dies ein Item mit vielfachen Lösungsvarianten. Das Item kann somit zugleich als Beispiel dafür dienen, dass innerhalb des abgeleiteten Regressionsmodells immer noch Spielraum für Einflüsse weiterer Art auf die Schwierigkeit besteht.

Schwierigkeitsgenerierung bei den Begrifflichen Modellierungsaufgaben

Für die Items des Bereichs *„begriffliches Modellieren"* ergibt sich wiederum ein anderes Erklärungsmodell. Zunächst fällt auf, dass die wenigen Items, bei denen von den Schülerinnen und Schülern explizites Argumentieren, d.h. Begründungen und Beweise, verlangt wird, hohe Schwierigkeiten aufweisen. Untersucht man diese Variable einzeln, so erklärt sie 23 % der Varianz der Schwierigkeit bei einem standardisierten Beta = .482. Diesen erheblichen Einfluss zeigen auch die absoluten Zahlen in Tabelle 5 an. Da es sich aber um einen speziellen Einzelprozess des Verarbeitens innerhalb des Modellierungsprozesses handelt, sind auch die allgemeinen Merkmale zu untersuchen. Und hier zeigt sich, dass das Argumentieren als solches im allgemeinen Regressionsmodell nicht mehr signifikant wird, wenn man andere Indikatoren ins Spiel bringt.

Vielmehr zeigt sich, dass der sonst dominante Einfluss der curricularen Wissensstufe nun keinen und der Kontext einen sehr starken Einfluss hat. Relevant ist für diese Gruppe vorwiegend die Komplexität des Modellierens insgesamt, gemessen durch die internationalen Kompetenzklassen, und auch die Offenheit der Mathematisierung spielt eine Rolle.

Schwierigkeitsgenerierung auf den begrifflichen Modellierungsaufgaben

Merkmal	*Standardisiertes Beta*	*Signifikanzniveau*	*Erklärte Varianz der Aufgabenschwierigkeit (R^2) im Gesamtmodell*
Komplexität der Modellierung	.451	.002	
Rein innermathematische Deduktion	.434	.002	
Offenheit / Multiple Lösbarkeit	.281	.039	
			30,6 %

Offenbar hängt die Schwierigkeit von Aufgaben des begrifflichen Typs zentral davon ab, wie hoch der Anspruch an den Modellierungsprozess insgesamt ist, während das stoffliche Wissen im engeren Sinne keine Rolle spielt. Aufgaben, die einen rein inner-mathematischen, sozusagen kontextfreien Verarbeitungsprozess erfordern, gibt es nur wenige, aber sie erweisen sich als schwierig.

Wieder zeigt die Tabelle mit den absoluten Zahlen, wie sich die einzelnen relevanten Merkmale auf die Aufgaben dieses Typs verteilen:

Tabelle 5:
Schwierigkeitsstufung bei den Begrifflichen Modellierungsaufgaben

Kompetenz-Stufe	Kurzbeschreibung	Komplexität der Modellierung *Reproduktion / Verknüpfung / Verallgemeinerung*	Offenheit / multiple Lösbarkeit *nicht vorh. / vorhanden*	Rein innermathematischer Verarbeitungsprozess *nein / ja*	Argumentieren *Nicht vorh. / vorhanden*
V (13 Items)	Selbständiges Strukturieren eines Problems	2 / 7 / 4	3 / 10	10 / 3	8 / 5
IV (13 Items)	Mathematische Modelle beurteilen, insbesondere funktionale Zusammenhänge	1 / 11 / 1	3 / 10	12 / 1	12 / 1
III (16 Items)	Vorgegebene Zusammenhänge verstehen	0 / 16 / 0	7 / 9	16 / 0	16 / 0
II (3 Items)	Hinweise unterstützen den begrifflichen Schritt	2 / 1 / 0	2 / 1	3 / 10	3 / 0
I (1 Item)		0 / 1 / 0	1 / 0	1 / 0	1 / 0

Auch für diese Gruppe sind bereits im PISA-Bericht (Klieme, Neubrand & Lüdtke 2001) Beispiele diskutiert worden. Die internationale Aufgabe „Äpfel 1“ (Fortsetzen einer bereits begonnenen Tabelle mit den Anzahlen der Nadel-bzw. Apfelbäume) kann für Stufe III stehen, weil - wie oben gezeigt - in der Aufgabenformulierung bereits Hilfen (hier: graphischer Art, indem eine Tabellenanordnung gewählt ist) für die notwendige Modellbildung gegeben sind. Das Beurteilen funktionaler Zusammenhänge tritt typischerweise auf Stufe IV auf. Ein Beispiel ist das aus TIMSS-III (“mathematische Grundbildung“; vgl. Klieme 2000) in den nationalen PISA-Testteil übernommene Item „xy=1“. Es ist dies zugleich ein Beispiel für eine Aufgabe, in der der Verarbeitungsprozess rein innermathematisch ist.

Für zwei Zahlen x ($x > 0$) und y gilt die Gleichung $x \cdot y = 1$.
Welche der folgenden Aussagen ist wahr?

Wenn der Wert von x größer als 1 ist, so ist der Wert von y negativ.
Wenn der Wert von x größer als 1 ist, so ist der Wert von y größer als 1.
Wenn der Wert von x kleiner als 1 ist, so ist der Wert von y kleiner als 1.
Wenn der Wert von x zunimmt, so nimmt auch der Wert von y zu.
Wenn der Wert von x zunimmt, so nimmt der Wert von y ab.

Kompetenzstufe IV zeigt hier an, dass begriffliches Umgehen mit funktionalen Beziehungen nicht zum geläufigen Repertoire von Schülern der 9. Jahrgangsstufe gehört, obwohl es vom Stoff her so sein könnte oder sogar müsste. Dies ist konsistent zu zahlreichen anderen Untersuchungen, die auf Defizite deutscher Schülerinnen und Schüler im begrifflichen Arbeiten hinwiesen (Neubrand, Neubrand & Sibberns 1998).

Stufe V erfordert selbständiges systematisches Durcharbeiten. Ein Beispiel, das zugleich zeigt, dass dies keineswegs mit technisch-anspruchsvollem Stoff-Wissen einhergehen muss, ist die Aufgabe „31 Pfennig“:

Wie kannst du einen Geldbetrag von genau 31 Pfennig hinlegen, wenn du nur 10-Pfennig-, 5-Pfennig- und 2-Pfennig-Münzen zur Verfügung hast? Gib *alle* Möglichkeiten an!

Dieses Item gehört zur begrifflichen Gruppe, weil das Item die konsequente Durchführung einer selbst zu wählenden Strategie verlangt, wobei die notwendigen (einfachsten) Berechnungen offenbar ganz hinter die Aufgabe, eine vollständige Systematik zu entwerfen, zurücktreten. Es handelt sich also tatsächlich in kognitiver Hinsicht um das Aufstellen eines „mathematischen Modells“, nämlich darum, eine Art systematischen Überblicks über alle vorkommenden Fälle zu erhalten.

4 Fazit: Zur Bedeutung differenzierter Schwierigkeitsmodelle

Das Ziel des vorliegenden Beitrags war es, die inhaltliche Bedeutung der PISA-Skala zur mathematischen Grundbildung zu klären. Zu diesem Zweck wurden zunächst auf didaktischer und kognitionspsychologischer Basis Aufgabenmerkmale bestimmt, von denen angenommen werden kann, dass sie den Schwierigkeitsgrad beeinflussen. Im Einklang mit dem Rahmenkonzept von PISA wurde das *Modellieren* - sei es aus innermathematischen oder außermathematischen Situationen heraus - als Kern des mathematischen Denkens betrachtet. Dementsprechend geht es bei den wichtigsten Aufgabenmerkmalen vor allem um die Qualität des Modellierungsprozesssses, d.h. seine allgemeine Komplexität, seine Offenheit, den Umfang der zugehörigen Verarbeitungsprozesse und die Notwendigkeit, mathematische Argumente zu formulieren. Hinzu kommen die curriculare Wissensstufe und der Kontext (inner- vs. außermathematisch) als weitere schwierigkeitsrelevante Aufgabenmerkmale. Diese sechs zentralen Merkmale wurden in die empirischen Analysen einbezogen: Experten schätzten die 117 internationalen und nationalen PISA-Mathematikaufgaben im Hinblick auf das Vorhandensein dieser Merkmale ein. Die Einschätzungen wurden sodann dazu genutzt, die empirische Schwierigkeit der Aufgaben vorherzusagen.

Ein erstes Resultat der Analysen lautet, dass es tatsächlich möglich ist, einen bedeutenden Teil der Varianz zwischen den Aufgaben in einem solchen „Schwierigkeitsmodell" aufzuklären. Über alle 117 Aufgaben hinweg konnten etwa 45 % der Varianz zwischen den Items erklärt werden. Allerdings muß deutlich darauf hingewiesen werden, dass ein solches Schwierigkeitsmodell keine Lösungs*prozesse* analysiert, sondern die Gesamt-Schwierigkeit rekonstruiert. Ferner ist zu betonen, dass die Ergebnisse nur für den jeweils untersuchten Aufgabenkontext gelten, hier also für den PISA-Test zur mathematischen Grundbildung. Es geht ja darum, für das konkrete Testinstrument zu bestimmen, welche Art von Anforderungen die leistungsstarken bzw. leistungsschwachen Schüler bewältigen können. In diesem Sinne wurden fünf Stufen der mathematischen Kompetenz unterschieden und jeweils unter Nutzung der Aufgabenmerkmale charakterisiert.

Der vorliegende Aufsatz differenziert dieses allgemeine Schwierigkeitsmodell aus. Aufgrund fachdidaktischer Überlegungen unterscheiden wir drei Typen mathematischen Arbeitens - technische Aufgaben, rechnerisches und begriffliches Modellieren -, für die getrennte Schwierigkeitsmodelle erstellt werden. Die Auswertungen belegen, dass in verschiedenen Bereichen des mathematischen Arbeitens bei PISA auch unterschiedliche Faktoren auf die Schwierkeitswerte einwirken:

Technische Aufgaben sind fast ausschließlich durch ihre curricularen Wissensanforderungen bestimmt, Aufgaben des begrifflichen Typs hingegen durch die Komplexität des Modellierens und das Vorhandensein mehrere Lösungswege. Auch die Schwierigkeitsanalyse bei den „klassischen", in der Schule üblichen, hier als „rechnerisch" bezeichneten Modellierungsaufgaben zeigt, dass dem Prozess des Modellierens selbst eine eigenständige Bedeutung zukommt. Der Anspruch an den Modellierungsprozess, d.h. die Komplexität des Prozesses und der Umfang der Modellierung, erwies sich als ein entscheidender Indikator für Schwierigkeit.

Die Ergebnisse des vorliegenden Aufsatzes verdeutlichen somit die Chancen, aber auch die Grenzen der Charakterisierung von Kompetenzstufen im Rahmen eines probabilistischen Testmodells. Es ist möglich, mittels weniger zentraler Anforderungsmerkmale darzustellen, was schwierige Aufgaben von leichten Aufgaben und somit kompetentere Personen von weniger kompetenten unterscheidet. Damit werden Testergebnisse - über den sozialen Vergleich mit Ergebnissen anderer Testteilnehmer hinaus - inhaltlich beschreibbar. Der Zusammenhang zwischen Aufgabenmerkmalen und Schwierigkeit ist jedoch keineswegs eindeutig, wie die Verteilung der Merkmale auf Kompetenzstufen zeigt. Und er wird moderiert vom Typ der zu lösenden Aufgabe.

Die fachdidaktische Unterscheidung zwischen technischen, rechnerischen und und begrifflichen Aufgaben sollte also bei der zukünftigen Untersuchung mathematischer Kompetenzen berücksichtigt werden. Aus psychologischer Sicht stellt sich die Aufgabe, auf experimentellem Weg und auf dem Weg der Konstruktvalidierung zu untersuchen, welche Lösungsprozesse tatsächlich beim Bearbeiten bestimmter Aufgaben ablaufen.

Literatur

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., Weiß, M. (Hrsg.) (2001): *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich.* Opladen: Leske & Budrich.

Baumert, J, Klieme, E. & Watermann, R. (1999): Jenseits von Gesamttest- und Untertestwerten: Analyse differentieller Itemfunktionen am Beispiel des mathematischen Grundbildungstests von TIMSS. In: H.-J. Herber & F. Hofmann (Hrsg.), *Schulpädagogik und Lehrerbildung. Festschrift zum 60. Geburtstag von Josef Thonhauser* (S. 302 - 324). Innsbruck: Studien Verlag.

Blum, W. (1996). Anwendungsbezüge im Mathematikunterricht - Trends und Perspektiven. In G. Kadunz et al (Hrsg.), *Trends und Perspektiven - Beiträge zum 7. Internationalen Symposium zur Didaktik der Mathematik in Klagenfurt Sept. 1994.* (S. 15 -38). Wien: Hölder-Pichler-Tempsky

Blum, W. & Wiegand, B. (1998). Wie kommen die deutschen TIMSS-Ergebnisse zustande? In W. Blum, & M. Neubrand (Hrsg.) , *TIMSS und der Mathematikunterricht: Informationen, Analysen, Konsequenzen* (S. 28 - 34). Hannover: Schroedel.

Bromme, R., Seeger, F & Steinbring, H. (1990). *Aufgaben als Anforderungen an Lehrer und Schüler* (= IDM-Untersuchungen zum Mathematikunterricht, Bd. 14). Köln: Aulis.

Christiansen, B. & Walther, G. (1986). Task and activity. In B.Christiansen, A.G.Howson & M.Otte (Eds.), *Perspectives on mathematics education* (S. 243-307). Dordrecht / Boston / Lancaster / Tokyo: D. Reidel Publishing Company.

De Lange, J. (1996). Real problems with real world mathematics. In C. Alsina & al. (Eds.), *Proceedings of the 8th International Congress on Mathematical Education, Sevilla July 1996* (pp. 83 - 110). Sevilla: Thales

Freudenthal, H. (1977). *Mathematik als pädagogische Aufgabe.* (Bd.1, Bd.2) Stuttgart: Ernst Klett Verlag.

Freudenthal, H. (1983). Didactical phenomenology of mathematical structures. Dordrecht: Reidel.

Heymann, H.W. (1996). *Allgemeinbildung und Mathematik.* Weinheim und Basel: Beltz.

Hiebert, J. (Ed.) (1986). *Conceptual and procedural knowledge: The case of mathematics.* Hillsdale, NJ: Lawrence Erlbaum

Klieme, E. (1989). *Mathematisches Problemlösen als Testleistung.* Frankfurt: Lang.

Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In: J. Baumert & al., *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 141 - 191). Opladen: Leske & Budrich.

Klieme, E. (2000). Fachleistungen im voruniversitäten Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte. In J. Baumert & al (Hrsg.), *TIMSS/III - Dritte Internationale Mathematik-und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2* (S. 57-128). Opladen: Leske+Budrich.

Neubrand, J. & Neubrand, M. (1999). Effekte multipler Lösungsmöglichkeiten: Beispiele aus einer japanischen Mathematikstunde. In C. Selter & G. Walther (Hrsg.): *Mathematikdidaktik als design science – Festschrift für Erich Christian Wittmann* (S. 148 - 158). Leipzig, Stuttgart, Düsseldorf: Ernst Klett Grundschulverlag 1999.

Neubrand, J. (2001). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen - Schülerarbeitsphasen und Selbsttätigkeit in den Stunden der TIMSS-Video-Studie* (Dissertation: Freie Universität Berlin). Hildesheim: Franzbecker (in Druck).

Neubrand, J., Neubrand, M. & Sibberns, H. (1998). Die TIMSS-Aufgaben aus mathematik-didaktischer Sicht: Stärken und Defizite deutscher Schülerinnen und Schüler. In W.Blum & M.Neubrand, *TIMSS und der Mathematikunterricht: Informationen, Analysen, Konsequenzen.* (S. 17 - 27).Hannover: Schroedel.

Neubrand, M. (2001). PISA: „Mathematische Grundbildung“ / „mathematical literacy“ als Kern einer internationalen und nationalen Leistungsstudie. In: G. Kaiser, N. Knoche & al. (Hrsg.): *Leistungsvergleiche im Mathematikunterricht - Ein Überblick über aktuelle nationale Studien.* Hildesheim: Franzbecker, S. 177 - 194.

Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., Lind, D., Löding, W., Möller, G., Wynands, A. (Deutsche PISA-Expertengruppe Mathematik) (2001). Grundlagen der Ergänzung des internationalen PISA-Mathematik-Tests in der deutschen Zusatzerhebung. *Zentralblatt für Didaktik der Mathematik 33(2),* 33 - 45.

OECD (Ed.) (1999): *Measuring student knowledge and skills. A new framework for assessment.* Paris: OECD Publ.Service.

OECD (Hrsg.) (2001): *Lernen für das Leben: Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000.* Paris: OECD Publications.

Reusser, K. (1992). Kognitive Modellierung von Text-, Situations- und mathematischem Verständnis beim Lösen von Textaufgaben. In K. Reiss & M. Reiss (Eds.), *Maschinelles Lernen - Modellierung von Lernen mit Maschinen* (pp 225 - 249). Berlin: Springer.

Stein, M., Grover, B.W. & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal 32(2),* 455 - 488.

Williams, G. & Clarke, D.J. (1997). Mathematical task complexity and task selection. In D.M. Clarke et al. (Eds.), *Mathematics: Imagine the possibilities.* (pp 406 - 415). Brunswick, Victoria: Mathematics Association oif Victoria.

Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik, Nr. 61,* 37 - 46.

Anschriften der Autoren:

Michael Neubrand, Universität Flensburg
Mürwiker Straße 77, D-24943 Flensburg. neubrand@uni-flensburg.de

Eckhard Klieme, Deutsches Institut für Internationale Pädagogische Forschung
Schloßstraße 19, D-60486 Frankfurt a.M., klieme@dipf.de

Oliver Lüdtke, Max-Planck-Institut für Bildungsforschung, Berlin
Lentzeallee 94, D-14195 Berlin, luedtke@mpib-berlin.mpg.de

Johanna Neubrand, Universität Lüneburg
Scharnhorststraße 1, D-21332 Lüneburg, neubrand@uni-lueneburg.de