

Norbert SOMMER, Osnabrück

Merkmale und Schwierigkeit von Mathematikaufgaben

Aufgaben sind ein zentrales Element des Mathematikunterrichts. Anforderungsniveau und -struktur a priori richtig zu bewerten, ist aus Sicht der „Passung“ von Informationsangebot und sachstrukturellem Entwicklungsstand (Heckhausen) Voraussetzung für Leistungsmotivation und Lernerfolg. Mathematiklehrer unterschätzen die Schwierigkeit von Testaufgaben für ihre Schüler häufig (vgl. u.a. Eder, 2003). Eine Lehrergruppe erwartet bei den PISA-2000-Items eine im Durchschnitt um 16,2 % höhere Lösungswahrscheinlichkeit als empirisch ermittelt, bei der bekannten 31-Pfennig Aufgabe sogar ein um 50 % besseres Abschneiden. Die Ursachen dieser Bewertungsunsicherheit sind bisher nur unspezifisch aufgeklärt.

Ergebnisse von Parallelarbeiten, die als Teil eines Evaluationsprojekts zur Qualitätsentwicklung des Mathematikunterrichts einer Schule seit mehreren Jahren analysiert und dokumentiert werden (vgl. Sommer, 2004), bestätigen die Einschätzungsprobleme.

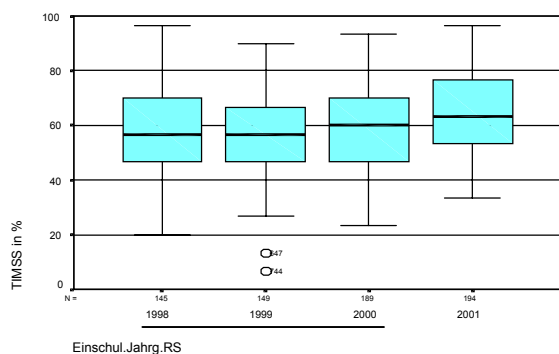


Abb. 1 TIMSS-Aufgaben Anfang 9

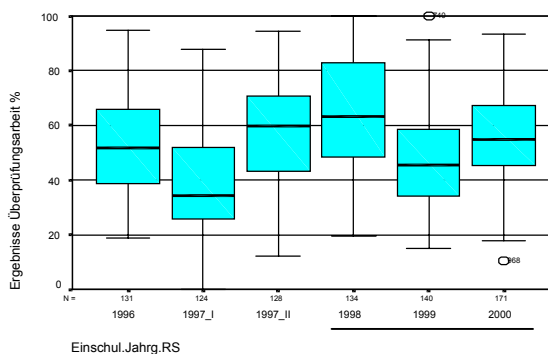


Abb. 2 Überprüfungsarbeit 2. Hälfte 10

Die Jahrgänge 1998-2000 unterscheiden sich in einem Test mit TIMSS-Aufgaben Anfang Klasse 9 nicht. Die Schwankungen in den Ende 10 geschriebenen Überprüfungsarbeiten sind daher auf die Testgestaltung und nicht auf Fähigkeitsunterschiede der Schülerjahrgänge zurückzuführen.

Die Identifizierung von Aufgabenmerkmalen, die das Anforderungsniveau qualitativ und quantitativ zuverlässig und valide beschreiben, trägt

- zur Entwicklung von Kompetenzmodellen im Rahmen der Arbeit an Bildungsstandards,
- zur Konstruktion von Tests in Vergleichsstudien und von Klassenarbeiten um das Spektrum der erwarteten Teilkompetenzen abzudecken, und
- zur Verbesserung der diagnostischen Fähigkeiten von Lehrern bei.

Die Diskussion in Fachkonferenzen über den Ausfall von Parallelarbeiten gewinnt an Substanz, wenn Erkenntnisse über die Beziehung zwischen Eigenschaften der Aufgaben und dem Lösungserfolg der Schüler vorliegen.

Mehrere Arbeitsgruppen haben die 117 Aufgaben der PISA-2000-Erhebung auf der Grundlage verschiedener Hintergrundtheorien analysiert (vgl. Neubrand u.a., (2002); Blum u.a., (2004); Cohors-Fresenborg u.a., (2004)) und dabei unterschiedliche Aufgabenmerkmale in Regressionsanalysen mit der Aufgabenschwierigkeit als abhängiger Variable eingebracht. Die Ergebnisse können hier nicht dargestellt werden. Die folgende Tabelle zeigt das Ergebnis einer Analyse mit den Merkmal-Ratings der drei Gruppen¹.

Aufgabentyp	Merkmal	r ²
rechnerische und begriffliche Modellierungsaufgaben	- Kognitive Komplexität (Cohors-Fresenborg u.a.)	71,0 %
	- Formelhandhabung (Cohors-Fresenborg u.a.)	
	- Grundvorstellungsintensität (Blum u.a.)	
	- Formalisierung von Wissen (Cohors-Fresenborg u.a.)	

Die „Grundvorstellungsintensität“ erklärt einen größeren Anteil der Schwierigkeitsvarianz als die „Sprachlogische Komplexität“ im ansonsten mit Cohors-Fresenborg u.a. (2004, S. 130) übereinstimmenden Ergebnis der schrittweisen Regression. Eine vertiefte Analyse der Beziehung dieser beiden Merkmale, die der Phase der Interpretation der Aufgabenanforderung zuzuordnen sind, könnte einen Beitrag zum besseren Verständnis kognitiver Prozesse während der „Ansatzfindung“ leisten.

Alle genannten Untersuchungen beziehen sich auf den PISA-2000-Itemsatz. Es ist zu vermuten, dass Zusammenhänge zwischen Aufgabenschwierigkeit und Merkmalsausprägung beobachtet werden, die auf Besonderheiten der Aufgabenauswahl zurückzuführen sind. Das Verfahren der Regressionsanalyse ist daher auf die Aufgaben aus den Überprüfungsarbeiten (Abb. 2) angewandt worden.

Die Datenqualität der Überprüfungsarbeiten (Repräsentativität von Kompetenz- und Schülerstichprobe, Objektivität, Reliabilität) ist geringer als die der PISA-Erhebung, aber besser als in normalen Klassenarbeiten (Zusammenarbeit von Lehrkräften aus sechs Schulen, Tests vor dem Einsatzes den Fachlehrern nicht bekannt, Kokorrektur durch eine zweite Lehrkraft). Die schultypische Struktur von Klassenarbeiten zeigt sich im Vergleich der Verteilungen der Aufgaben auf die Kompetenzklassen. Die Überprüfungsarbeiten sind nicht einseitig auf technische Fertigkeiten ausgerichtet.

	Technische Items	Rechnerische Modellierung	Begriffliche Modellierung
Internationale PISA-Aufgaben	3,2 %	45,2 %	51,6 %
Nationale PISA-Aufgaben	24,4 %	40,7 %	34,9 %
Überprüfungsarbeiten	29,4 %	50,0 %	20,6 %

¹ Die Ratings wurden zur Verfügung gestellt. Dafür möchte ich mich herzlich bedanken.

Deren Anteil ist aber höher als im curriculumnahen nationalen PISA-Aufgabensatz. Der Anteil begrifflicher Modellierungsaufgaben ist deutlich geringer; einfache Textaufgaben (Rechnerische Modellierung), in denen bekannte Verfahren anzuwenden sind, überwiegen. Die Schwierigkeit der Tests (s. Mediane in Abb. 2) hängt eng mit dem Anteil technischer Aufgaben in den Arbeiten zusammen ($r = .79$; $N = 6$; $p = .059$ n.s.).

U.a. für die innerschulische Rückmeldung sind zusätzlich zu den Merkmalen der oben genannten Untersuchungen weitere Aspekte berücksichtigt worden, beispielsweise das Darbietungsformat, die Themenzugehörigkeit, die Position im Test und die von den Lehrern vorgesehene Bepunktung, in der sich Annahmen über das Anspruchsniveau ausdrücken.

Merkmal	Standardisiertes Beta	Signifikanzniveau	r^2 (korrigiert)
algebraischer Anspruch	-.419	.000	
Bearbeitungsrichtung	-.328	.000	
Sprachlogische Kompl.	-.281	.010	
Trigonometrie I	.159	.034	38,5 %

Das Ergebnis auf Basis aller Aufgaben und aller Schüler belegt die Bedeutung der algebraischen Kompetenz am Ende der Sek I, die durch notwendige Formelumstellungen auch in „Bearbeitungsrichtung“ einfließt, und der Sprachlogischen Komplexität. Dass von den Themen nur die jeweils kurz vor den Tests bearbeitete Trigonometrie einen Einfluss hat (wenn Trigonometrie, dann leichter), spricht eher gegen die Bedeutung des Unterrichtsinhalts als schwierighkeitsrelevantes Aufgabenmerkmal.

Abweichungen zwischen den empirischen Lösungswahrscheinlichkeiten und den Vorhersagen durch das Regressionsmodell weisen auf Defizite in der Aufklärung der Aufgabenanforderungen hin. Ein Beispiel sind Aufgaben zur Begründung trigonometrischer Funktionswerte wie $\sin 30^\circ = 0,5$ oder $\tan 60^\circ = \sqrt{3}$, die in fünf der sechs Tests vorkommen. Schwierigkeitschwankungen zwischen $p = .29$ und $p = .76$ (leider nicht mit der Tendenz einer kontinuierlichen Verbesserung im Laufe der Jahre) spiegeln sich in den Merkmalszuweisungen nicht präzise genug wider.

Separate Regressionsberechnungen für die Einzeltests und für Schülergruppen (Mädchen/Jungen; größer/kleiner Median) lassen keine systematischen differentiellen Effekte erkennen. Andere Auswertungsmethoden belegen dagegen interessante Unterschiede. Die folgende Tabelle enthält Korrelationen der Differenzen der Aufgabenschwierigkeiten von Jungen und Mädchen mit Aufgabenmerkmalen.

Jungen schneiden in allen Überprüfungsarbeiten besser ab als Mädchen.

Aufgabenschwierigkeit			Aufgabenmerkmale				
r ²	p (m)	p (w)	Sachrechnen	Kontext	Formelhandhab.	Text / Term	Term
Diff. m-w	(.085)	- .316	.313	.383	- .321	- .374	- .457

Die Aufgabenschwierigkeit für Mädchen korreliert negativ mit der Schwierigkeitsdifferenz, d.h. bei schwereren Aufgaben ist der Leistungsvorsprung der Jungen größer. Wenn Aufgaben als „Sachrechnen“ klassifiziert sind, ist die Schwierigkeitsdifferenz mit Vorteil der Jungen größer als bei anderen Aufgaben, während negative Werte (z.B. bei „Formelhandhabung“) bedeuten, dass die Differenz zwischen Jungen und Mädchen bei Vorliegen dieses Aufgabenmerkmals kleiner ist oder evtl. sogar zugunsten der Mädchen ausfällt. Erkenntnisse über Stabilität und gegebenenfalls Ursachen dieser geschlechtspezifischen Leistungsprofile bei Sachrechenaufgaben bzw. algebraischen Anforderungen wären fachdidaktisch bedeutsam. Eine genauere Untersuchung ist hier aber nicht möglich.

Lösungsweg- und Fehleranalysen belegen Klasseneffekte. Bei nahezu gleichen Lösungswahrscheinlichkeiten in einer Prozentrechnungsaufgabe zum „vermehrten Grundwert“ treten große Unterschiede in der Häufigkeit des Auftretens verschiedener Lösungsansätze zutage (Zuordnungstabelle oder Verhältnisgleichung, Prozentwert und Addition oder in einem Schritt zum vermehrten Grundwert), die deutlich mit der Leistung im Gesamttest in Beziehung stehen und durch Regressionsanalysen nicht erkannt werden.

Literatur:

- Blum, W.; v. Hofe, R.; Jordan, A.; Kleine, M. (2004). Grundvorstellungen als aufgabenanalytisches und diagnostisches Instrument bei PISA. In: M. Neubrand (Hrsg.), S. 145 - 157
- Cohors-Fresenborg, E.; Sjuts, J.; Sommer, N. (2004). Komplexität von Denkvorgängen und Formalisierung von Wissen. In: M. Neubrand (Hrsg.), S. 109 – 144
- Eder, F. (2003). Tests und Lehrerurteil. Wie gut stimmen externe Leistungstests mit Lehrereinstufungen überein. In: E.J Brunner; P. Noack; G. Scholz; I. Scholl (Hrsg.). Diagnose und Intervention in schulischen Handlungsfeldern. Münster, S. 125 - 140
- Neubrand, M.; Klieme, E.; Lüdtke, O.; Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung. In: Unterrichtswissenschaft, 30. Jahrgang, Heft 1, S. 100 - 119
- Neubrand, M. (Hrsg.) (2004). Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland. Wiesbaden
- Sommer, N. (2004). Welchen Nutzen kann die Einzelschule aus den Ergebnissen und Instrumenten der „großen Vergleichsuntersuchungen“ ziehen? In: Journal für Mathematikdidaktik, Jg. 25, Heft 3/4, S. 269 – 293

² Signifikant sind Korrelationen ab einer Höhe von $r > .18$, betrachtet werden hier nur Korrelation ab einer Höhe von $r \geq .30$.