

Prenzel, Manfred; Häußler, Peter; Rost, Jürgen; Senkbeil, Martin

Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen?

Unterrichtswissenschaft 30 (2002) 2, S. 120-135

urn:nbn:de:0111-opus-76826



in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen / conditions of use

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Mitglied der Leibniz-Gemeinschaft
Informationszentrum (IZ) Bildung
Schloßstr. 29, D-60486 Frankfurt am Main
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Unterrichtswissenschaft

Zeitschrift für Lernforschung
30. Jahrgang / 2002 / Heft 1

Thema:

7a+b, abc

PISA – Konzept und Ergebnisse

Verantwortliche Herausgeber:
Jürgen Baumert, Manfred Prenzel

Jürgen Baumert, Manfred Prenzel:
Einführung

98

Michael Neubrand, Eckhard Klieme, Oliver Lüdtke, Johanna Neubrand:
Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test
zur mathematischen Grundbildung

100x

DLG ✓

Manfred Prenzel, Peter Häußler, Jürgen Rost, Martin Senkbeil:
Der PISA-Naturwissenschaftstest:
Lassen sich die Aufgabenschwierigkeiten vorhersagen?

120

Joachim Wirth, Eckhard Klieme:
Computer literacy im Vergleich zwischen Nationen,
Schulformen und Geschlechtern

136x

HP ✓

Allgemeiner Teil

Britta Kohler:
Zur Rezeption von TIMSS durch Lehrerinnen und Lehrer

158

Hinweise für die Autoren

190

97

Manfred Prenzel, Peter Häußler, Jürgen Rost, Martin Senkbeil

Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen?

The PISA Science Test: Can we Predict the Item Difficulties

Die Ergebnisse der ersten PISA-Runde im Jahre 2000 haben für die deutschen Schülerinnen und Schüler hinsichtlich ihrer naturwissenschaftlichen Kompetenzen im internationalen Vergleich nur unterdurchschnittliche Leistungen ausgewiesen. Durch die Entwicklung eines ergänzenden nationalen Naturwissenschaftstests war es möglich, über diese Befunde hinausgehend differenzierte Informationen über mentale Anforderungen und Aktivitäten zu erfassen, die typisch sind für naturwissenschaftliches Denken und Verstehen. Im Rahmen dieser Arbeit soll anhand einer Analyse der nationalen und internationalen Items geklärt werden, ob sich die Schwierigkeiten der naturwissenschaftlichen Items durch bestimmte Aufgabenmerkmale vorhersagen lassen. Als Analysekatégorien werden formale und wissensbezogene Aufgabenmerkmale sowie kognitive Anforderungen beim Lösen der Testitems unterschieden. Die statistischen Analysen belegen, dass ein erheblicher Anteil (45%) der Varianz der Itemschwierigkeiten durch die genannten Merkmale aufgeklärt werden kann. Darüber hinaus geben die Ergebnisse differenzierte Hinweise auf spezifische Stärken und Schwächen der deutschen Schülerinnen und Schüler. Im Anschluss daran werden Konsequenzen aus der ersten PISA-Runde für das Testdesign des nationalen Zusatztests der nächsten PISA-Erhebung im Jahre 2003 diskutiert.

The findings from the first lap of PISA 2000 have revealed that German students, in regard to their scientific competency, underperform when internationally compared. An additional national science test made it possible to compile more differentiated information relating to the students' mental demands and activities which are typical for scientific thinking and understanding and which go beyond the original findings. In the context of this project, it is to be clarified, if the difficulties of scientific items can be predicted by certain task features. This task will be carried out on the basis of an analysis of national and international items. Formal and knowledge-related features as well as cognitive demands will be applied as analysis categories and will then be differentiated when solving the test item. Statistical analysis prove that a considerable proportion (45%) of the variance of item difficulties can be explained by the mentioned featured. Moreover, the findings differentiate between specific strenghts and weakness of German students. In connection to this, consequences will be discussed for the test design of the national supplementary test of the next PISA survey in 2003.

1 Einführung

Im Programme for International Student Assessment (PISA) der OECD sind die Naturwissenschaften einer der drei Kompetenzbereiche, die in den vorge-

sehenen Erhebungszyklen regelmäßig untersucht werden. Die internationale Testkonzeption beruht auf einer Vorstellung von naturwissenschaftlicher Grundbildung im Sinne von „Scientific Literacy“ (OECD, 1999). Im Blickpunkt stehen naturwissenschaftliche Kompetenzen, die für eine verständige und verantwortungsvolle Teilnahme am gesellschaftlichen Leben erforderlich sind. Diese Kompetenzen umfassen vor allem das Wissen bzw. Verständnis naturwissenschaftlicher Begriffe und Prinzipien (Konzepte) und naturwissenschaftlicher Denk- und Arbeitsweisen (Prozesse). Die Konzeption von PISA deckt sich weitgehend mit neueren internationalen und nationalen Konzeptionen naturwissenschaftlicher Grundbildung (vgl. Duit, Häußler, Prenzel, 2001; Gräber & Bolte, 1997), wie an anderer Stelle ausführlich dargestellt und diskutiert wurde (Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001).

Auch die wichtigsten Ergebnisse des internationalen Vergleichs wurden im internationalen Report (OECD, 2001) und im Berichtsband des Deutschen PISA-Konsortiums (Baumert, Klieme, Neubrand, Prenzel, Schiefele, Schneider, Stanat, Tillmann & Weiß, 2001) präsentiert. Betrachtet man die Kennzahlen, dann ergibt sich im Bereich der Naturwissenschaften ein ähnliches Bild wie in den anderen Kompetenzbereichen: Die Leistungen der deutschen Schülerinnen und Schüler im Naturwissenschaftstest liegen deutlich unter dem internationalen Durchschnitt (Prenzel et al., 2001). Der Mittelwert der deutschen Stichprobe liegt für die Naturwissenschaftsleistungen bei 487 (SD = 102), während die durchschnittliche Leistung aller getesteten 15-jährigen Jugendlichen 500 (SD = 100) beträgt. Die im PISA-Bericht dargestellten Daten belegen, dass in Deutschland insbesondere im unteren Leistungsbereich (den unteren 25 Prozent der Leistungsverteilung) gravierende Schwächen bestehen. Hinweise auf ausgeprägte naturwissenschaftliche Spitzengruppen in Deutschland gibt es nicht.

Tabelle 1:

Prozentuale Verteilung von Schülerinnen und Schüler unterschiedlicher Länder auf Kompetenzstufen der naturwissenschaftlichen Grundbildung

Stufen naturwissenschaftlicher Kompetenz	Deutschland	Schweiz	Österreich	Vereinigtes Königreich	Alle Länder (ohne Deutschland)
Stufe V (>661)	3,4	4,6	4,5	9,0	4,1
Stufe IV (554-661)	23,9	24,5	33,0	33,7	25,7
Stufe III (498-553)	20,1	20,1	22,5	22,0	20,1
Stufe II (421-497)	26,3	27,4	25,3	21,8	25,8
Stufe I (<421)	26,3	23,4	14,7	13,5	24,3

Diese Einschätzung kann anhand der prozentualen Verteilung von Schülerinnen und Schülern aus den verschiedenen Ländern auf Kompetenzstufen der naturwissenschaftlichen Grundbildung belegt werden (Tabelle 1). Auf den

unteren Kompetenzstufen (einfaches Fakten- oder Alltagswissen) sind deutsche Schülerinnen und Schüler überrepräsentiert, auf den höheren Kompetenzstufen (konzeptuelles und prozedurales Verständnis) relativ selten vertreten.

Entsprechende Unterscheidungen und Aufteilungen nach Stufen naturwissenschaftlicher Kompetenz müssen beim PISA-Test 2000 vorsichtig interpretiert werden. Für die Naturwissenschaften als Nebenkomponekte bei PISA 2000 stand nur eine knappe Testzeit zur Verfügung. Aufgrund der kleinen Anzahl von Naturwissenschaftsitems im internationalen Test kann man die Unterscheidung und inhaltliche Beschreibung von Kompetenzstufen nur als erste Annäherung und keinesfalls empirisch abgesichert verstehen. Zur Vorbereitung der Testkonzeption und Aufgabenpools für die nachfolgenden Erhebungen – bei der PISA-Erhebung 2006 stehen die Naturwissenschaften im Zentrum – sind eingehende Analysen von Aufgabenmerkmalen und der zur Lösung benötigten kognitiven Operationen jedoch hilfreich und nützlich. Das deutsche PISA-Konsortium hat die Möglichkeit nationaler Ergänzungsstudien genutzt. Mit der Einführung eines zweiten Testtages bestand die Möglichkeit, für alle Kompetenzbereiche zusätzliche Erhebungen vorzunehmen. Für die Entwicklung eines nationalen Tests im Bereich naturwissenschaftlicher Kompetenz waren folgende Überlegungen ausschlaggebend (vgl. Prenzel et al., 2001):

- Die geringe Itemzahl im internationalen Test gestattet noch keine zuverlässigen Aussagen über Kompetenzen in den einzelnen naturwissenschaftlichen Fächern;
- Mit der Ausrichtung des internationalen Tests auf Verständnis und Wissensanwendung könnten Aspekte unberücksichtigt bleiben, die in der Tradition deutscher Lehrpläne und Naturwissenschaftsdidaktik eine wichtige Rolle spielen;
- Die für den internationalen Test vorgesehenen Aufgabentypen und -formate bieten nur unzureichende Möglichkeiten, spezifische Leistungsdefizite deutscher Schülerinnen und Schüler zu identifizieren und voneinander abzugrenzen.

Mit der Entwicklung und Anwendung eines ergänzenden nationalen Naturwissenschaftstest bestand also die Möglichkeit, den internationalen Test unter bestimmten Gesichtspunkten zu validieren und die Akzeptanz der internationalen Vergleichsbefunde in Deutschland zu sichern (an diesem nationalen Test nahm freilich nur die deutsche Stichprobe teil!).

Bei der Entwicklung des nationalen Tests wurden Gesichtspunkte berücksichtigt, die aus wissenschaftlicher Perspektive als interessant und weiterführend erschienen. Aus der Absicht, differenziertere Informationen über mentale Anforderungen und Aktivitäten zu erfassen, die typisch sind für naturwissenschaftliches Denken, Verstehen und Schlussfolgern, wurde ein systematischer Ansatz der Aufgabenkonstruktion entwickelt. Bei der Konstruktion der nationalen Aufgaben wurde versucht, über die (äußeren)

Merkmale der Testaufgaben zu steuern, welche kognitiven Prozesse bei der erfolgreichen Bearbeitung einer Aufgabe ablaufen. Auf diese Weise sollten fünf kognitive Aspekte naturwissenschaftlicher Kompetenz erfasst werden (vgl. Prenzel et al., 2001):

- Aus einer grafischen Repräsentation richtige Informationen ableiten (Grafikverständnis);
- Fakten- und Konzeptwissen aus dem Gedächtnis abrufen und anwenden (Faktenwissen anwenden);
- Aus gegebener Information die richtigen Schlüsse ziehen (Schlüsse ziehen);
- Ein (räumliches) mentales Modell heranziehen (Mentales Modell);
- Einen Sachverhalt verbalisieren (Verbalisieren).

Ordnet man die Aufgaben jeweils einer dieser fünf kognitiven Teilkompetenzen zu (nach Maßgabe des in der Aufgabe als „dominant“ erachteten Prozesses, d.h. es kommen keine Mehrfachzuordnungen vor), weisen bereits die mittleren Aufgabenschwierigkeiten auf spezifische Stärken und Schwächen der deutschen Schülerinnen und Schüler hin (vgl. dazu auch Prenzel et al., 2001, 225ff). Tabelle 2, in der die mittleren Schwierigkeitsparameter für die einzelnen Teilkompetenzen eingetragen sind, verdeutlicht, dass der Umgang mit mentalen Modellen und das Anwenden von Faktenwissen besonders schwierige Anforderungen in der Zielgruppe der 15-jährigen darstellen.

Weitergehende Analysen belegen, dass vor allem Mädchen in diesen beiden Teilkompetenzen Schwächen aufweisen. Hier zeigen sich die größten Geschlechterdifferenzen zu ungunsten der Schülerinnen (vgl. Prenzel et al., 2001, 241 ff.). Dagegen fällt es den Schülerinnen und Schülern eher leicht, aus Grafiken die richtige Information abzuleiten.

Tabelle 2:
Mittlere Schwierigkeiten der Items, die den kognitiven Teilkompetenzen zugeordnet sind

Kognitive Teilkompetenzen	Anzahl Items: Nationaler Test	Anzahl Items: Internationaler Test	Mittlerer Itemparameter
Aus einer grafischen Repräsentation die richtige Information ableiten	8	3	-0.35
Faktenwissen aus dem Gedächtnis abrufen und anwenden	7	0	1.59
Aus gegebener Information die richtigen Schlüsse ziehen	2	12	0.45
Ein (räumliches) mentales Modell heranziehen	4	2	1.16
Einen Sachverhalt verbalisieren	5	8	0.89

Die Zuordnung der Items zu den kognitiven Kompetenzen war ein Ergebnis einer umfassenden Analyse der Aufgabenmerkmale sowohl der internationalen als auch der nationalen Testitems. Die Aufgabenanalyse diente dazu, die Konstruktionsprinzipien der nationalen Zusatzaufgaben transparent und einer empirischen Kontrolle zugänglich zu machen, sowie die Unterschiede zwischen den nationalen und den internationalen Items zu beschreiben.

Die als Ergebnis der Aufgabenanalyse erhaltenen Merkmale lassen sich in zweifacher Hinsicht für eine empirische Kontrolle der Aufgabenvalidität nutzen. Zum einen sollten die Aufgabenmerkmale eine prädiktive Validität hinsichtlich der Aufgabenschwierigkeit haben, d.h. Aufgaben mit einem ‚schwierigen‘ Merkmal sollten generell seltener gelöst werden. Die diesbezüglichen Ergebnisse werden im folgenden berichtet. Zum anderen stellt sich aber auch die Frage, ob es merkmalsbezogene Teilkompetenzen gibt, d.h. ob Aufgaben mit denselben Merkmalen höher miteinander korrelieren, als Aufgaben mit unterschiedlichen Merkmalen. Auch hierzu bietet der PISA-Berichtsband bereits erste Ergebnisse (Prenzel et al. 2001, 225ff). Eine Kreuzvalidierung der psychometrischen Struktur der 5 kognitiven Teilkompetenzen anhand der PISA-E Daten (nationale Ergänzungsstichprobe) wird andernorts berichtet (Prenzel, Carstensen, Rost & Senkbeil, in Vorbereitung).

2 Fragestellung

Auf der Grundlage einer detaillierten Aufgabenanalyse, die neben den bereits beschriebenen kognitiven Anforderungen eine Vielzahl weiterer Aufgabenmerkmale beinhaltet, soll anhand einer multiplen Regressionsanalyse die Frage beantwortet werden, ob sich die Itemschwierigkeiten des naturwissenschaftlichen Tests durch bestimmte Aufgabenmerkmale erklären bzw. vorhersagen lassen. Für das (theoretische) Verständnis von Testaufgaben und für zukünftige Testentwicklungen ist diese Frage hoch bedeutsam. Dabei ist nicht zu erwarten, dass die Aufgabenschwierigkeiten bereits auf der Basis von Aufgabenmerkmalen perfekt vorhergesagt werden können. Bereits eine mittlere Prädiktionsleistung bietet jedoch nützliche Hinweise darauf, welche Merkmale überhaupt schwierigkeitsrelevant sind, und worin die kognitiven Anforderungen bestehen können, die eine Aufgabe mehr oder weniger schwierig sein läßt.

3 Vorgehen

Die Kategorien für die Aufgabenanalysen wurden von der nationalen Expertengruppe für den Bereich Naturwissenschaften entwickelt. Handlungsleitend waren die folgenden Gesichtspunkte:

- Bei der Formulierung der Kategorien sollte auf eine strenge Trennung von Prozess- und Konzeptkategorien geachtet werden, weil es nur so

möglich ist, eine nicht ausreichende Wissensbasis als Ursache für das Nichtlösen einer Aufgabe von der Unfähigkeit, einen bestimmten Prozess auszuführen, zu unterscheiden.

- Die Konzeptkategorien sollten so aufgeschlüsselt werden, dass unterschiedliche Stufen der Wissensbasis (z.B. Allgemeinwissen, naturwissenschaftliche Bezeichnungen, Sachverhalte, Beziehungen zwischen Größen) abgebildet werden können.
- Um ein tragfähiges Erklärungsmodell für die Itemschwierigkeit zu erhalten, sollten außer den Prozess- und Konzeptkategorien auch (äußere) Aufgabenmerkmale hinzugenommen werden, von denen mutmaßlich die Schwierigkeit eines Items beeinflusst wird.

Das Kategoriensystem wurde anhand der Prozessmerkmale entwickelt, die in der internationalen PISA-Konzeption beschrieben sind, und in einem iterativen Prozess dem vorliegenden Itemmaterial angepasst.

Anhand der ausgearbeiteten Kriterien kann man drei Arten von Analysekatgorien unterscheiden, nämlich

- formale Aufgabenmerkmale,
- kognitive Anforderungen beim Lösen der Aufgaben und
- Merkmale der für das Lösen der Aufgaben erforderlichen Wissensbasis.

Für alle nationalen (N = 46) wie internationalen Testaufgaben (N = 34) wurden die zutreffenden Aufgabenmerkmale der einzelnen Analysekatgorien signiert, d.h. für jedes Item wurde geprüft, ob es das entsprechende Aufgabenmerkmal aufweist oder nicht.

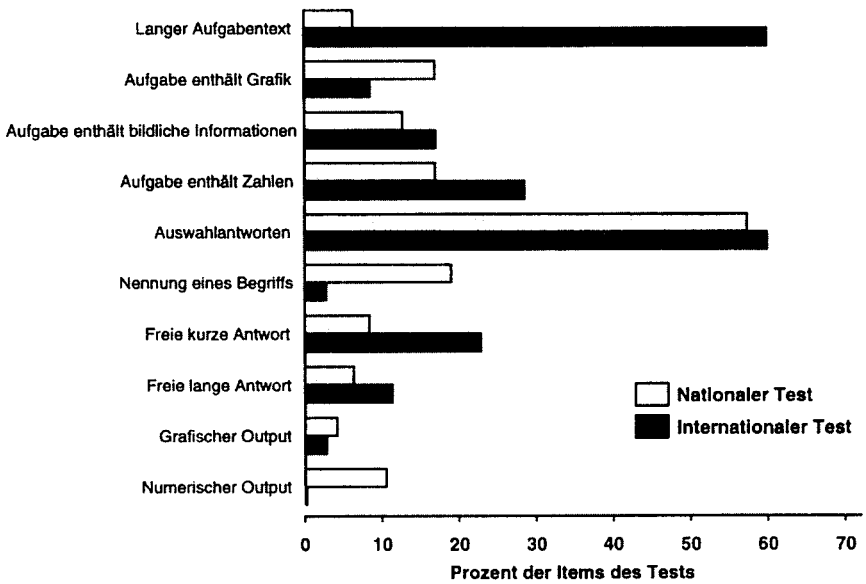
Im Folgenden wird zunächst jede Art von Aufgabenmerkmalen beschrieben sowie der Anteil der Items angegeben, der das jeweilige Aufgabenmerkmal aufweist. Anhand einer Grafik werden jeweils die Unterschiede zwischen dem nationalen und internationalen Naturwissenschaftstest verdeutlicht. (Abbildungen 1 - 3)

3.1 Formale Aufgabenmerkmale

Die formalen Aufgabenmerkmale beziehen sich im Wesentlichen auf „äußere“ Merkmale der Itemformulierung und des betreffenden Antwortformates. Sie bestimmen, wie der „Input“ der Aufgabenstellung und die Art des geforderten „Outputs“ aussieht.

Die nationalen Aufgaben unterscheiden sich von den internationalen Naturwissenschaftsitems deutlich dadurch, dass im internationalen Test sehr viel mehr gelesen werden muss. Dieser Unterschied leitet sich aus der unterschiedlichen Akzentuierung der internationalen bzw. nationalen Erhebungskonzeption ab: Bei den internationalen Aufgaben stand die Anwendung von

Abbildung 1:
Verteilung der formalen Aufgabenmerkmale im nationalen und internationalen Naturwissenschaftstest



Prozessen und Konzepten auf realistische und lebensnahe Fragestellungen (z.B. durch die Verwendung von Zeitungsartikeln) im Vordergrund, während sich die Aufgabenentwicklung beim nationalen Ergänzungstest stärker an den deutschen Lehrplänen orientierte. Die Einschränkung des Leseaufwands im nationalen Test ermöglicht eine Klärung der Frage, inwieweit die hohe Lesebelastung im internationalen Naturwissenschaftstest eine Korrelation mit dem Test zum Leseverständnis bedingen könnte.

Während sich Aufgaben mit Auswahlantworten bei beiden Tests in etwa die Waage halten, wurden im nationalen Test die Ein-Wort-Antworten gegenüber längeren verbalen Ausführungen bevorzugt. Das liegt unter anderem daran, dass der internationale Test ein präzises zielgruppenbezogenes Argumentieren als Bestandteil naturwissenschaftlicher Kompetenz betrachtet. Dieser Unterschied in den Konzeptionen sollte die Abhängigkeit der Leistung von den sprachlichen Fähigkeiten der Schüler im nationalen Test reduzieren. Anzumerken bleibt schließlich, dass in den internationalen Testaufgaben zwar mehr Zahlen als im nationalen Test vorgeben werden, aber in keinem einzigen Fall eine numerische Antwort verlangt wird.

3.2 Kognitive Anforderungen beim Lösen der Aufgaben

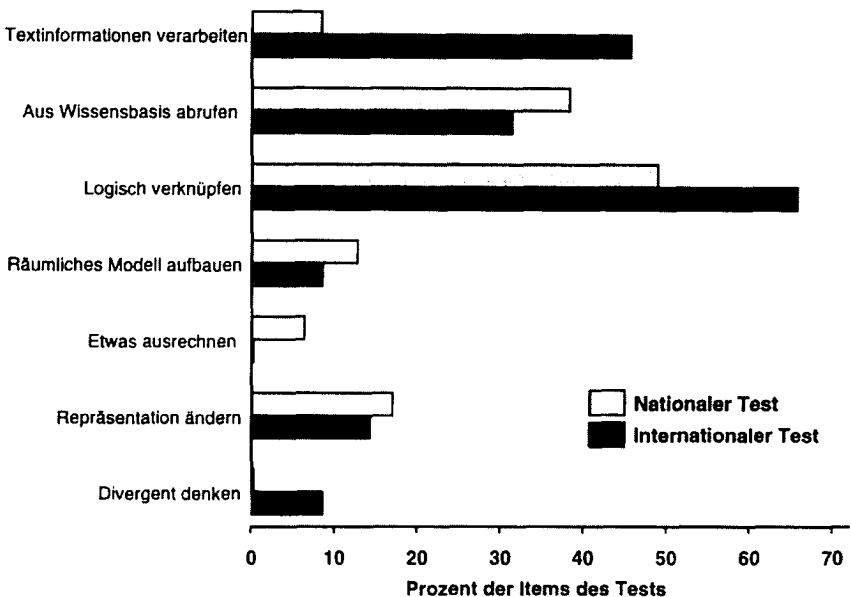
Unter dieser Kategorie werden Aufgabenmerkmale gefasst, die kognitive Prozesse bei der Aufgabenbearbeitung betreffen. Es geht hier also nicht um formale Merkmale des Testmaterials, sondern um Vorgänge, die bei der Auf-

gabenbearbeitung ablaufen und die für den Lösungsprozess entscheidend werden. Die Analyse kategorien beschreiben kognitive Prozesse, die bei der Verarbeitung und Anwendung naturwissenschaftlicher Informationen und Wissensinhalte relevant sind.

Nicht beabsichtigt, aber auffällig ist eine Korrespondenz dieser Prozesskomponenten zu den „klassischen“ Primärfaktoren der Intelligenz: verbal fluency (Wortflüssigkeit), memory (Kurzzeitgedächtnis), reasoning (konvergentes Denken), space (räumliches Vorstellungsvermögen), number (Rechenfähigkeit) und creativity (kreatives Denken) (Thurstone, 1938; es „fehlt“ die siebte Primärfähigkeit, die Wahrnehmungsgeschwindigkeit). Das bedeutet jedoch nicht, dass mit den zugeordneten Items allgemeine Intelligenzfaktoren gemessen werden. Vielmehr geht es hier um domänenspezifische Kompetenzen bei der Verarbeitung und Anwendung naturwissenschaftlicher Informationen und Wissensinhalte.

Abbildung 2:

Verteilung der kognitiven Anforderungen beim Lösen der Aufgaben im nationalen und internationalen Naturwissenschaftstest.



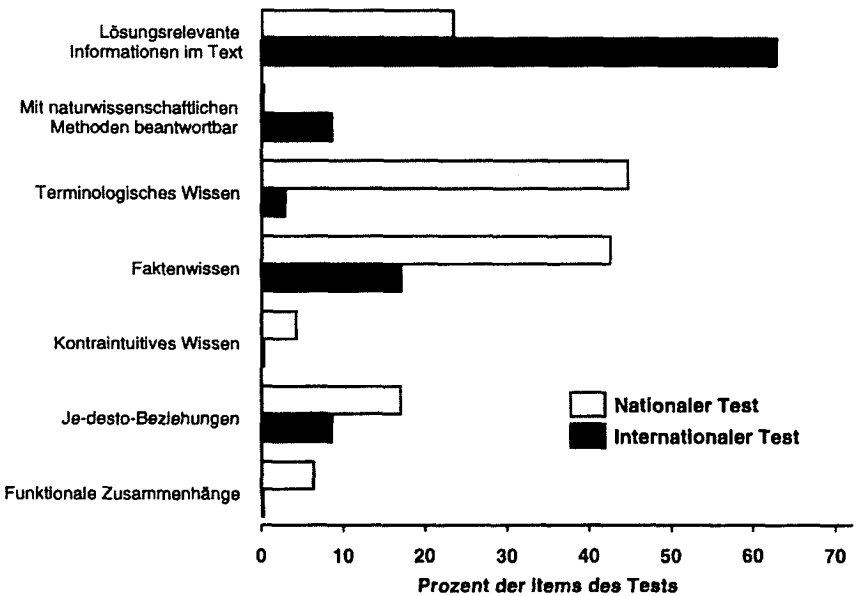
Die Abbildung 2 zeigt einen deutlichen Unterschied zwischen den beiden Tests in der Anforderung, die im Vorspann und im Aufgabentext gegebenen lösungsrelevanten Hinweise zu verarbeiten. Kommt diese Anforderung im nationalen Naturwissenschaftstest nur bei etwa jedem zwölften Item zum Tragen, so ist dies im internationalen Test bei etwa 45 Prozent aller Items der Fall. Auch in diesem Befund spiegeln sich die unterschiedlichen bzw. einander ergänzenden Konstruktionsprinzipien des internationalen und nationalen Tests wider.

Große Unterschiede bestehen auch bezüglich der Kategorien „Etwas ausrechnen“ und „Divergent denken“. Für die anderen Kategorien finden sich, wie auch konzeptionell beabsichtigt, keine bedeutsamen Unterschiede zwischen dem internationalen und dem nationalen Test.

3.3 Merkmale der für das Lösen der Aufgaben erforderlichen Wissensbasis

Unter diesem Aspekt wird für die einzelnen Aufgaben analysiert, welche Wissensbasis für eine erfolgreiche Lösung notwendig ist. Diese Merkmale wurden getrennt kodiert, um nicht die Prozesskomponenten mit der Verfügbarkeit von Wissens-elementen zu vermengen. Mit den Aufgabenmerkmalen dieser Kategorie werden vor allem die aus didaktischer Sicht sinnvoll zu differenzierenden Wissensinhalte aufgeschlüsselt.

Abbildung 3:
Verteilung der Merkmale der für das Lösen der Aufgaben erforderlichen Wissensbasis im nationalen und internationalen Naturwissenschaftstest.



Insgesamt lässt sich aus der Abbildung 3 ablesen, dass der internationale Test andere Ansprüche an die Wissensbasis stellt als der nationale Test. Über 60 Prozent aller Items des internationalen Tests erfordern die Verarbeitung von lösungsrelevanten Informationen im Aufgabentext. Diese setzt freilich ein Verständnis der jeweiligen naturwissenschaftlichen Konzepte und Prozesse voraus.

Demgegenüber wird im nationalen Test bei etwa 45 Prozent aller Items die Kenntnis eines nicht näher im Aufgabentext erläuterten naturwissenschaftli-

chen Begriffs vorausgesetzt. Bei den nationalen Aufgaben muss in höheren Maße als bei den internationalen Items Fakten- und Konzeptwissen (z.B. Geschwindigkeit als Quotient aus Weg und Zeit) präsent sein, um eine Aufgabe lösen zu können. Auch das Wissen über kontraintuitive Sachverhalte sowie die Kenntnis quantitativer Beziehungen zwischen Größen wird im nationalen Test häufiger vorausgesetzt.

Wenn auch diese Analyseergebnisse die unterschiedliche Ausrichtung der internationalen und nationalen Testkonzeption belegen, so sollten doch trotz aller Unterschiedlichkeit die Übereinstimmungen bzw. der gemeinsame Kern beider Testformen nicht übersehen werden.

3.4 Implikationen

Neben der Funktion dieser Itemmerkmale, die *Schwierigkeit* der Testaufgaben zu erklären (hierauf wird im Folgenden eingegangen), sollten sie auch Anhaltspunkte liefern, welche kognitive *Fähigkeiten* mit den Testaufgaben angesprochen werden.

Die eingangs beschriebenen kognitiven *Kompetenzen* entsprechen - mit leichten Abwandlungen - bestimmten Merkmalen der hier dargestellten Aufgabenanalyse. So entsprechen das Ziehen von Schlussfolgerungen, die Verwendung mentaler Modelle und das Grafikverständnis dem 3., 4. und 6. Aufgabenmerkmal der ‚kognitiven Anforderungen‘ (siehe Abbildung 2). Die Fähigkeit zum Verbalisieren entspricht dem 7. und 8. ‚formalen Aufgabenmerkmal‘ (siehe Abbildung 1) und das Abrufen und Anwenden von Faktenwissen dem 3. und 4. Merkmal der Wissensbasis (siehe Abbildung 3).

Für viele der übrigen Aufgabenmerkmale hat es aus theoretischen Gründen wenig Sinn, spezielle kognitive Kompetenzen anzunehmen. Für einige Aufgabenmerkmale war entweder die Anzahl entsprechender Items zu gering, um für alle Schüler Skalenwerte berechnen zu können, oder empirische Analysen haben ergeben, dass entsprechende Kompetenzdimensionen statistisch schwer von anderen Kompetenzen abzugrenzen sind.

4 Die Schwierigkeit der Items als Funktion der Aufgabenmerkmale: Methode

Im Folgenden wird dargestellt, wie die oben beschriebenen Aufgabenmerkmale mit der *Schwierigkeit* der Items zusammenhängen. Diese Zusammenhänge wurden mit Hilfe einer Regressionsanalyse berechnet. Bei dieser Analyse stellt die Itemschwierigkeit (in Form des Itemparameters des Rasch-Modells) die Kriteriumsvariable dar. Prädiktoren sind die Aufgabenmerkmale (kodiert als Dummy-Variablen).

Allerdings sind nicht alle Aufgabenmerkmale in gleicher Weise für die Einbeziehung in ein Schwierigkeitsmodell geeignet. Ein Grund ist, dass bestimmte Aufgabenmerkmale gehäuft gleichzeitig auftreten können, ihr Ein-

fluss auf die Aufgabenschwierigkeit untereinander also konfundiert sein kann. Ein zweiter Grund für die Reduzierung der Prädiktoren ist der, dass nicht alle Kategorien, die zur Beschreibung eines Items sinnvoll sind, auch zur Charakterisierung der Itemschwierigkeit taugen. Dazu gehören beispielsweise die Kategorien „Auswahlantworten“ (formale Merkmale), „Etwas aus der Wissensbasis abrufen“ (kognitive Merkmale) oder „Alle lösungsrelevanten Informationen im Text“ (erforderliche Wissensbasis). Ein dritter Grund, nicht alle Beschreibungskategorien zu übernehmen, ist die inhaltliche Ähnlichkeit mancher Kategorien. Eine gemeinsame Berücksichtigung in demselben Prädiktionsmodell würde dazu führen, dass Prädiktoren identische Varianzanteile des Kriteriums aufklären würden. Dazu gehören die Paare „Naturwissenschaftliche Bezeichnungen“ und „Nennung eines Begriffs“, „Etwas ausrechnen“ und „Aufgabe enthält Zahlen“ sowie „Aufgabe enthält Grafik“ und „Repräsentation ändern“. Daher werden nur die erstgenannten Merkmale in die Regressionsgleichung aufgenommen. Es ergeben sich 15 Prädiktoren zur Vorhersage der Itemschwierigkeiten (s. Tabelle 3).

Tabelle 3:

Regressionsgewichte und Standardfehler der Aufgabenmerkmale zur Vorhersage der Itemschwierigkeiten (B: Regressionsgewicht; S: Standardfehler)

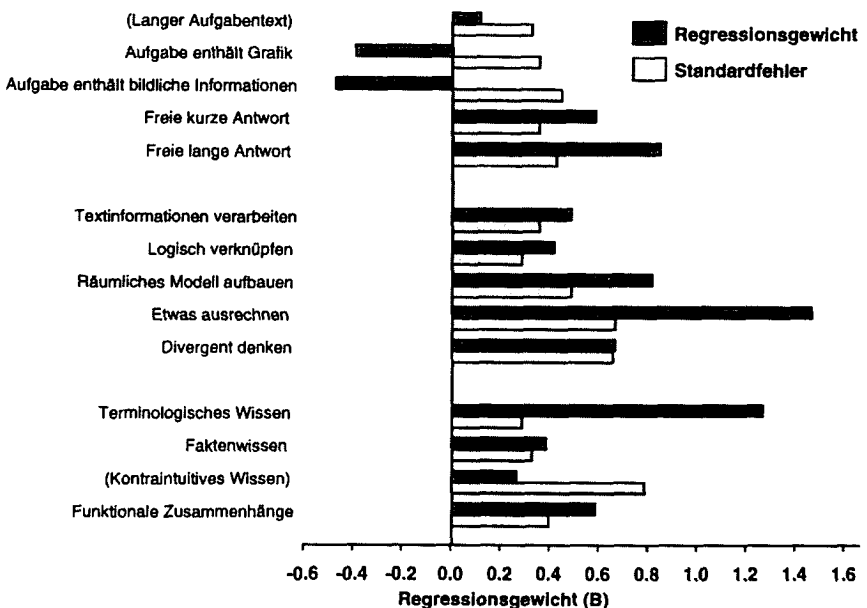
Aufgabenmerkmale	Einbezug aller Prädiktoren		Einbezug der Prädiktoren mit B>S	
	B	S	B	S
Formale Aufgabenmerkmale				
Langer Aufgabentext	0.12	0.33		
Aufgabe enthält Grafik	-0.39	0.36	-0.41	0.35
Aufgabe enthält bildliche Informationen	-0.47	0.45	-0.50	0.43
Freie kurze Antwort	0.59	0.36	0.62	0.35
Freie lange Antwort	0.85	0.43	0.85	0.42
Kognitive Anforderungen				
Textinformationen verarbeiten	0.49	0.36	0.53	0.33
Logisch verknüpfen	0.42	0.29	0.46	0.27
Räumliches Modell aufbauen	0.82	0.49	0.85	0.46
Etwas ausrechnen	1.47	0.67	1.39	0.64
Divergent denken	0.67	0.66	0.68	0.64
Merkmale mit Wissensbasis				
Terminologisches Wissen	1.27	0.29	1.30	0.28
Faktenwissen	0.39	0.33	0.34	0.31
Kontraintuitives Wissen	0.27	0.79		
Funktionale Zusammenhänge	0.59	0.40	0.54	0.38
Multiple R²	.452		.450	

5 Ergebnisse

Insgesamt wurden zwei Regressionsanalysen gerechnet. Die erste mit allen 15 einbezogenen Aufgabenmerkmalen als Prädiktoren führte zu einem aufgeklärten Anteil an der Varianz der Aufgabenschwierigkeiten von 45,2 Prozent (multiples R^2). Jedoch war bei einigen Kategorien der Standardfehler (Spalte S in Tabelle 3) größer als der Absolutwert des Regressionsgewichtes (Spalte B). Für eine zweite Analyse wurden diese Kategorien eliminiert, was zu marginalen Änderungen der Gewichte der verbliebenen Prädiktoren und einer minimalen Reduzierung der aufgeklärten Varianz führte (s. Tabelle 3).

Die Vorhersage der Aufgabenschwierigkeit durch die Aufgabenmerkmale führt zu einem aussagekräftigen Befund, der im Folgenden näher betrachtet werden soll. Die Abbildung 4 zeigt die Ergebnisse der Tabelle 3 in grafischer Form.

Abbildung 4:
Vorhersage der Aufgabenschwierigkeit durch Aufgabenmerkmale
(Regressionsgewichte und Standardfehler der Prädiktoren für die Itemschwierigkeit;
 $R^2 = 45,0$; bei den eingeklammerten Prädiktoren ist der Standardfehler größer
als das Regressionsgewicht).



Vor einer Diskussion dieser Ergebnisse sollen im Folgenden die relativ großen Standardfehler und die Größe der aufgeklärten Varianz von 45 Prozent kommentiert werden. Hierfür können drei Gründe verantwortlich sein:

- Erstens kann die Annahme eines additiven Zusammenwirkens der Schwierigkeitskomponenten, wie sie in einem Regressionsmodell getrof-

fen wird, unangemessen sein. Diese Voraussetzung ist zumindest auf der auf einzelne Schüler bezogenen Datenebene nicht erfüllt. So kann zum Beispiel ein Schüler eine Aufgabe, die eine bestimmte Wissensbasis *und* eine bestimmte Fähigkeit voraussetzt, nur dann lösen, wenn er über beide Komponenten verfügt. Fehlt das Wissen, enthält seine (Nicht-)Lösung keine Information über seine Fähigkeit und umgekehrt. Da sich die durchgeführten Regressionen aber stets auf die für die Gesamtpopulation ermittelten Rasch-Parameter beziehen, ist die Modellverletzung sicher weniger gravierend.

- Zweitens wurden manche Kategorien nur bei sehr wenigen Items signiert, was den relativ großen Standardfehler der zugehörigen Regressionsgewichte erklärt.
- Drittens ist denkbar, dass die untersuchte Schülerpopulation in dem Sinne heterogen ist, als einzelne Aufgabenkomponenten in verschiedenen Schülergruppen unterschiedliche Schwierigkeiten aufweisen.

Angesichts dieser einschränkenden Faktoren liefert die Schwierigkeitsanalyse ein sehr anschauliches Ergebnis.

6. Diskussion und Ausblick

Bei den Aufgabenmerkmalen fällt besonders auf, dass die Vorgabe einer Grafik bzw. eines stilisierten Bildes (etwa einer Schaltskizze) die Lösung der Aufgabe erleichtert (negatives Gewicht). Dies weist darauf hin, dass bei der Vorgabe nicht-verbaler Elemente die Anforderungen an das Textverständnis reduziert worden sind und damit das Item leichter wird. Während die Länge des Aufgabentextes die Aufgabenschwierigkeit nicht nennenswert beeinflusst, ergibt sich aus den hohen Gewichten für die Formulierung verbaler Statements, dass die aktive Verbalisierung im Rahmen der Aufgabenbeantwortung sehr wohl einen Schwierigkeitsfaktor für die Schülerinnen und Schüler darstellt.

Von den übrigen kognitiven Anforderungen erweist sich die Bedingung „Etwas ausrechnen“ als besonders aussagekräftig für die Itemschwierigkeit. Relativ schwierig sind weiterhin Aufgaben, in denen es um räumliches Vorstellungsvermögen und um divergentes Denken geht. Wenig Schwierigkeiten bereitet es dagegen, „Elemente der Wissensbasis logisch miteinander zu verknüpfen“ (konvergent zu denken).

Bezüglich der Merkmale der Wissensbasis fällt auf, dass Aufgaben, in denen das Verständnis oder die Nennung eines bestimmten naturwissenschaftlichen Begriffs vorausgesetzt wird, besondere Schwierigkeiten zu bereiten scheinen.

Insgesamt zeigt die Schwierigkeitsanalyse ein plausibles Muster von kognitiven Hürden oder Hilfestellungen bei der Aufgabenbearbeitung. Im Gegenzug stellen die Ergebnisse eine Art Validierung des entwickelten Rasters von

Aufgabenmerkmalen dar. Die zur Beschreibung der Gemeinsamkeit und Unterschiedlichkeit von nationalem und internationalem Test herangezogenen Aufgabenmerkmale haben eine nachweisbare psychologische Qualität und korrespondieren mit Eigenschaften der Aufgabenbearbeitung.

Der Befund, dass die Länge des Aufgabentextes offenbar keinen signifikanten Einfluss auf die Itemschwierigkeit hat, mag zunächst verwundern. Ein Grund könnte darin liegen, dass es in der Regel für die erfolgreiche Lösung eines Items nicht erforderlich ist, einen langen, mehreren Items vorangestellten Aufgabenvorspann im Gedächtnis zu behalten. Werden bestimmte Teile des Vorspanns für die Lösung eines Items gebraucht, so werden diese im Itemtext wiederholt. Ein langer Aufgabentext ist also letztlich nur über die dazu erforderliche Lesezeit lösungsbestimmend.

Die Reichweite der hier zu beschreibenden Befunde ist jedoch begrenzt: Für zukünftige Testkonstruktionen (z.B. für PISA 2003 und PISA 2006) sind die Ergebnisse von großer Bedeutung. In ihrer inhaltlichen Aussagekraft sind sie jedoch nicht ohne Weiteres generalisierbar. Bestimmte Aufgabenmerkmale, die sich in *diesem* Test als besonders leicht oder schwierig erweisen, können in einem anderen Test von deutlich anderer Schwierigkeit sein. Ein Beispiel ist etwa das Vorhandensein einer Grafik in der Aufgabe. Dieses Merkmal erleichtert in unserem Aufgabenmaterial die Lösung einer Aufgabe (s. unten Tabelle 3), kann aber in einem anderen Test dazu verwendet werden, die Aufgabenschwierigkeit zu *erhöhen*, wenn etwa besonders knifflige Grafiken zur Lösung einer Aufgabe herangezogen werden müssen.

Ausblick. Für die nächste PISA-Erhebung 2003, bei der der internationale Naturwissenschaftstest wiederum durch einen nationalen Test ergänzt werden wird, ist die Konstruktion eines vollständigen Facettendesigns geplant, das kognitive Komponenten und Inhaltsbereiche der naturwissenschaftlichen Fachrichtungen systematisch miteinander kombiniert.

Hierfür wird der Ansatz von PISA 2000, bei dem fünf kognitive Teilkompetenzen differenziert wurden, überarbeitet und stärker formalisiert. Ähnlich wie bei der ersten PISA-Runde sollen anhand der Testaufgaben bestimmte Kompetenzen erfasst werden, die sich auf eine Reihe von kognitiven Leistungen beziehen. Diese werden nicht inhaltsfrei, sondern bereichsspezifisch erfasst. Aufgrund der Anbindung der kognitiven Prozesse an die jeweiligen Inhalte kann damit auf die getrennte Erfassung der Wissensbasis (Faktenwissen abrufen und anwenden) bei der PISA 2003-Erhebung verzichtet werden. Um eine optimierte Anbindung an die internationale PISA-Konzeption zu erreichen (vgl. z.B. OECD, 1999; Baumert et al., 2001), werden weitere kognitive Kompetenzen in das Testdesign integriert. Diese sind „Divergentes Denken“, „Bewertung/Evaluation“ und „Umgang mit Zahlen“, wobei die beiden erstgenannten Dimensionen Prozesskomponenten des internationalen PISA-Designs darstellen. Diese sieben kognitiven Kompetenzen stellen die erste Facette im facettentheoretischen Testdesign dar.

Die zweite Facette bilden die zentralen Konzepte („big ideas“) der jeweiligen naturwissenschaftlichen Fachrichtung. Es werden drei globale systembezogene Betrachtungsweisen unterschieden, nämlich Systeme mit kontinuierlichen Variablen, diskontinuierliche Systeme und dynamische Systeme. Obwohl die naturwissenschaftlichen Disziplinen Biologie, Chemie und Physik zu allen drei Betrachtungsweisen Beiträge aufweisen, bestehen deutliche Korrespondenzen, so dass auf der Ebene der zentralen Konzepte folgende Zuordnung getroffen wird:

- Biologie: dynamische Systeme;
- Chemie: diskontinuierliche Systeme;
- Physik: kontinuierliche Systeme.

Diesen zentralen Konzepten lassen sich nun weiterhin Basiskonzepte - quasi als inhaltliche Ausgestaltung der zentralen Konzepte - im jeweiligen Kontext des Fachbereichs zuordnen, z.B. die Basiskonzepte „Kohlenstoffkreislauf“ für das Fach Biologie, „Aggregatzustände“ für das Fach Chemie oder „Energieumwandlungen“ für das Fach Physik. Im facettentheoretischen Testdesign werden nun die Facetten der kognitiven Kompetenzen gekreuzt mit einer Facette naturwissenschaftlicher Inhaltsbereiche (Basiskonzepte der jeweiligen Fächer). Zu jedem Basiskonzept wird demnach ein Itemset mit jeweils sieben Items (für jede kognitive Komponente ein Item) erstellt.

Dieses Testdesign optimiert die Testaufgaben insofern, als damit die kognitiven Kompetenzen 'unabhängig' von den Inhalten (weil stets für dieselben Inhalte erhoben) und inhaltspezifische Fähigkeiten 'unabhängig' von den Prozessmerkmalen skaliert werden können. Dabei kann geprüft werden, ob es fachspezifische Prozeßkompetenzen gibt (z.B. zur Anwendung mentaler Modelle im Biologieunterricht) oder ob die Interaktion von Inhalten und kognitiven Prozessen vernachlässigbar ist.

Die Orientierung der Testkonstruktion an einem Facettendesign dient also primär der validen Erfassung von Schülerkompetenzen. Gleichwohl ermöglicht sie auch eine Schwierigkeitsanalyse der Aufgaben im zuvor dargestellten Sinne. Da die Skalenwerte in einem Facettendesign auf den aggregierten Antworten aller Items mit demselben Merkmal beruhen, geben Skalenmittelwerte für bestimmte Untergruppen (z.B. Jungen und Mädchen) direkt Auskunft über den Schwierigkeitsunterschied dieses Merkmals in diesen Gruppen. Insofern erwarten wir uns von den Ergebnissen der zweiten PISA-Erhebung psychometrisch besser abgesicherte Ergebnisse über fachspezifische und fachübergreifende Stärken und Schwächen deutscher Schülerinnen und Schüler in den verschiedenen Schulformen und sozio-demografisch relevanten Teilpopulationen.

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 und Budrich.
- Duit, R., Häußler, P., & Prenzel, M. (2001), Schulleistung im Bereich der naturwissenschaftlichen Bildung. In F. E. Weinert, *Leistungen in Schulen*. Weinheim, Basel: Beltz Verlag.
- Gräber, W., & Bolte, C. (Hrsg.) (1997), *Scientific literacy. An international symposium*. Kiel: IPN.
- OECD (1999), *Measuring student knowledge and skills: A new framework for assessment*. Paris: OECD.
- OECD (2001), *Lernen für das Leben. Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000*. Paris: OECD.
- Prenzel, M., Carstensen, C., Rost, J. & Senkbeil, M. (in Vorbereitung), Validierung der kognitiven Komponenten an der Ergänzungstichprobe PISA 2000.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001), Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In: Baumert, J., Klie-me, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (191-248). Opladen: Leske und Budrich.
- Thurstone, L.L. (1938), *Primary mental abilities*. Chicago: University of Chicago Press.

Anschrift der Autoren:

Manfred Prenzel, Peter Häußler, Jürgen Rost, Martin Senkbeil
Leibniz-Institut für die Pädagogik der Naturwissenschaften
Olshausenstraße 62, 24098 Kiel