Zeitschrift: Bildungsforschung und Bildungspraxis : schweizerische Zeitschrift für Erziehungswissenschaft = Éducation et recherche : revue suisse des sciences de l'éducation = Educazione e ricerca : rivista svizzera di scienze dell'educazione

Herausgeber: Schweizerische Gesellschaft für Bildungsforschung

Band: 20 (1998)

Heft: 1

Artikel: Leistungsprofil und Unterricht

Autor: Ramseier, Erich

DOI: https://doi.org/10.5169/seals-786228

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 13.12.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Leistungsprofil und Unterricht

Eine Analyse der schweizerischen Leistungen im naturwissenschaftlichen Test von TIMSS

Erich Ramseier

Die Schweiz schneidet beim internationalen Vergleich von TIMSS in den Naturwissenschaften deutlich schlechter ab als in der Mathematik. Diese Diskrepanz gab Anlass zur vorliegenden Analyse der schweizerischen Testleistungen in den Naturwissenschaften. Untersucht wurde, welche Aufgaben für die schweizerischen Schülerinnen und Schüler besonders schwierig oder aber besonders leicht sind. Als Beitrag dazu schätzten Lehrpersonen die fachliche und terminologische Schwierigkeit der Aufgaben ein. Es zeigt sich, dass in der Schweiz Aufgaben relativ gut bewältigt werden, die hohe kognitive Anforderungen stellen, bei denen es also um das Entdecken und Anwenden von naturwissenschaftlichen Prinzipien geht. Aufgaben, die spezielle fachliche Inhalte testen, werden in der Schweiz dagegen relativ schlecht gelöst und zwar besonders dann, wenn sie hohe begrifflich-terminologische Ansprüche stellen. Dieses Leistungsprofil stimmt mit den Merkmalen und Schwerpunkten des naturwissenschaftlichen Unterrichts in der Schweiz überein. Auf der Grundlage der Aufgabenanalyse lassen sich verschiedene Teilergebnisse von TIMSS besser verstehen.

1 Anstoss zur Untersuchung und Fragestellung

1.1 Ausgangslage: Unterschiedliche Ergebnisse in Mathematik und Naturwissenschaften

Im neusten internationalen Vergleich von Schulleistungen hat die Schweiz im sechsten bis achten Schuljahr in den Naturwissenschaften deutlich schlechter abgeschnitten als in der Mathematik (Moser et al. 1997). Dies ist, was die Sekundarstufe I betrifft, wohl das Auffälligste am Ergebnis der Schweiz in der

Third International Mathematics and Science Study (TIMSS)[1]. Ein Grund dafür ist sicher, dass in der Schweiz die Unterrichtszeit in Mathematik im internationalen Vergleich eher lang, in den Naturwissenschaften eher kurz ist (Moser et al. 1997; Ramseier 1997). In der Schweiz liegen zudem die Naturwissenschaften gemessen am Stellenwert, den ihr die Schülerinnen und Schüler zubilligen, weiter hinter der Mathematik zurück als in anderen Ländern. Das unterschiedliche Abschneiden in den beiden Fächern ist somit plausibel. Dass sich die Leistungen eines Landes in den beiden Fächern genau entsprechen, ist schliesslich nicht zu erwarten. Der Zusammenhang zwischen den Leistungen in Mathematik und Naturwissenschaften ist auch bei den Schülerinnen und Schülern längst nicht perfekt ($r = .57$, im siebten Schuljahr der Schweiz).

Das Ergebnis ist dennoch unerwartet. Die Schweiz hat nämlich vier Jahre früher, beim zweiten *International Assessment of Educational Progress* (IAEP, Lapointe et al. 1992a; 1992b; Moser 1992) in den Naturwissenschaften genauso wie in der Mathematik hinter den ostasiatischen Ländern das beste Ergebnis erzielt. Dafür, dass die Schweiz bei TIMSS nicht mehr einen so brillanten Rang erreicht hat wie bei IAEP, gibt es gute Gründe, die in der Anlage dieser Untersuchungen liegen (Ramseier 1997). Wie ist es aber zu verstehen, dass das schweizerische Ergebnis in den Naturwissenschaften und in Mathematik bei IAEP etwa gleich war, bei TIMSS hingegen deutlich verschieden ist? Diese Frage ist der Ausgangspunkt der hier beschriebenen Analyse[2]. Eine anfänglich unangenehme Diskrepanz erwies sich dabei als Hinweis auf inhaltlich interessante Gegebenheiten.

Wenn die Testergebnisse in einem Fach nach vier Jahren deutlich anders ausfallen, kann dies natürlich daran liegen, dass sich der Unterricht in dieser Zeit verändert hat. Es sind aber keine massiven didaktischen Neuerungen auszumachen, die sich in dieser kurzen Zeit so stark auf die Leistungen hätten auswirken können. Vor einer solchen inhaltlichen Interpretation muss auf alle Fälle geprüft werden, ob beide Male das gleiche gemessen wurde.

Tatsächlich unterscheiden sich die naturwissenschaftlichen Tests von IAEP und TIMSS. Der IAEP-Test ist deutlich leichter. Seine Aufgaben werden im Mittel aller Länder, die an beiden Tests teilnahmen, zu 69% richtig gelöst, jene des TIMSS-Tests dagegen nur zu 58% (Moser et al. 1997: 49). In einer Inhaltsanalyse wurden zudem bei TIMSS 32% der Fachbegriffe als eher schwierig oder sehr schwierig eingestuft, bei IAEP dagegen nur 19% (Ramseier 1997: 14).

1.2 Die Besonderheiten des naturwissenschaftlichen Unterrichts in der Schweiz

Die Unterschiede zwischen den naturwissenschaftlichen Tests von TIMSS und IAEP können die Leistungsunterschiede nur dann erklären, wenn die Anforderungen des IAEP-Tests dem schweizerischen Unterricht besser entsprechen als jene des TIMSS-Tests. Wie sieht der naturwissenschaftliche Unterricht in der Schweiz aus? Wir haben bereits festgestellt, dass er zeitlich kürzer ist als in anderen Ländern. Von daher ist es plausibel, dass das Ergebnis im leichteren

Test besser ist. Der IAEP-Test, der einfachere Begriffe und Bezeichnungen verwendet, scheint auch mit den im folgenden kurz skizzierten inhaltlichen Prioritäten des schweizerischen Unterrichts besser übereinzustimmen.

Besonders auf der Primarstufe ist die Behandlung naturkundlicher Themen in vielen Kantonen in übergreifende Fächer wie «Heimatkunde» oder «Realien» integriert (Moser et al. 1997). Ziel dieses Unterrichts ist weniger, Fachwissen zu vermitteln, sondern vielmehr, das Kind darin zu unterstützen, sich selbst und die Welt, in der es lebt, besser verstehen zu lernen. Auch auf der Sekundarstufe I stehen mehr «Orientierungswissen» sowie die Anwendung einfacher naturwissenschaftlicher Methoden, das Beobachten und Experimentieren und nicht systematisches Fachwissen im Vordergrund. Moser et al. (1997: 72 ff.) illustrieren das mit Zitaten aus mehreren Lehrplänen der deutschschweizerischen Kantone. Sie führen auch Belege dafür an, dass andere Länder, zum Beispiel die USA, andere Schwerpunkte setzen und schon früh Fachbegriffe und abstrakte Themen wie «Atome, Ionen, Moleküle» behandeln. Laut der Einschätzung tschechischer und schweizerischer Experten steht auch in Tschechien der Erwerb von Faktenwissen im Vordergrund des naturwissenschaftlichen Unterrichts. Dabei werden schon in unteren Klassen relativ komplizierte wissenschaftliche Begriffe eingeführt (Ramseier 1997).

1.3 Fragestellung

Die skizzierten Prioritäten des naturwissenschaftlichen Unterrichts in der Schweiz und die Unterschiede zwischen den Ergebnissen in den naturwissenschaftlichen Tests von TIMSS und IAEP passen zueinander. Aus ihnen lässt sich folgern, dass schweizerische Schülerinnen und Schüler über relativ wenig fachliches Detailwissen und insbesondere über weniger fachspezifische, komplexe Begriffe und Bezeichnungen verfügen. Im Verstehen von naturwissenschaftlichen Zusammenhängen und der Anwendung von naturwissenschaftlichen Methoden sollten sie keinen Rückstand haben, jedenfalls wenn es um Phänomene aus dem unmittelbaren Erlebnisbereich der Kinder geht. Diese Folgerungen sind plausibel. Wenn sie zutreffen, so müssen sie sich auch im Leistungsprofil *innerhalb* des TIMSS-Tests zeigen. Dies gibt uns eine Möglichkeit, diese Aussagen zu überprüfen. Wenn ein Faktor die Leistung der Schweiz im TIMSS-Test beeinflusst, sollte sich dies in konsistenter Weise zeigen. Die Schweiz muss dann in jenen Aufgaben relativ gut abschneiden,

- die insgesamt eher leicht sind,
- die nicht spezifisches Fachwissen und die Kenntnis schwieriger Begriffe und Bezeichnungen voraussetzen,
- bei denen es um naturwissenschaftliche Methoden, Problemlösen, das Verständnis von Zusammenhängen und deren Anwendung auf konkrete Situationen geht.

Mit einer Analyse der spezifischen Schwierigkeit der einzelnen Aufgaben für die Schweiz können wir untersuchen, ob sich die Prioritäten des Unterrichts und das Leistungsprofil inhaltlich entsprechen.

2 Vorgehen

2.1 Analysekriterium und Analysemethode

Die oben vorgestellten Hypothesen beziehen sich auf Unterschiede bzw. Zusammenhänge zwischen Aufgaben. Im Kern geht es darum, herauszufinden, welche Aufgaben für die Schweiz *relativ* leicht bzw. schwierig sind. Die Einfachheit bzw. Schwierigkeit einer Aufgabe kommt im Anteil richtiger bzw. falscher Lösungen zum Ausdruck (Rost 1996: 95). Da es vertrauter ist, mit der Aufgabenschwierigkeit zu argumentieren, verwenden wir im folgenden diese Grösse. Das primäre Aufgabenmerkmal für unsere Analyse ist somit der Prozentsatz falscher Lösungen in einem Land, die *nationale Aufgabenschwierigkeit*. Diese nationale Aufgabenschwierigkeit wird hier für alle 146 Aufgaben[3] anhand der Leistungen im höheren der beiden von TIMSS getesteten Schuljahre definiert. Den Durchschnitt der nationalen Aufgabenschwierigkeiten einer Aufgabe bezeichnen wir als *internationale Aufgabenschwierigkeit.*

Bei unserer Fragestellung geht es weniger darum abzuklären, warum welche Aufgaben schwierig sind. Es sind selbstverständlich weitgehend dieselben Aufgaben, die für die Schweiz oder für irgendein anderes Land schwierig sind. Entsprechend korrelieren die schweizerische Aufgabenschwierigkeit und die internationale Aufgabenschwierigkeit hoch ($r = .92$). Es geht spezifischer darum, welche Aufgaben für die Schweiz, verglichen mit anderen Ländern, *besonders* schwierig sind. Dies können wir an der Differenz zwischen schweizerischer und internationaler Aufgabenschwierigkeit ablesen. Diese Differenz bezeichnen wir als *relative Aufgabenschwierigkeit für die Schweiz.*

Ungewohnt an unserer Analyse ist, dass wir Beziehungen und Unterschiede bei Merkmalen von Aufgaben und nicht von Personen untersuchen. Signifikanztests prüfen dann, ob gefundene Zusammenhänge auf ein Universum von Aufgaben verallgemeinert werden können, bzw. ob die Zusammenhänge zwischen den Aufgabenmerkmalen stärker sind, als was wir zufällig erwarten. Personen und Länder spielen nur indirekt eine Rolle, indem sie die Ausprägung bestimmter Aufgabenmerkmale bestimmen, z.B. die nationale Aufgabenschwierigkeit. Wir können davon ausgehen, dass unsere Aussagen für die Schweiz und andere Länder gültig sind, da der Berechnung der Aufgabenmerkmale, soweit es um Lösungshäufigkeiten geht, repräsentative nationale Stichproben zugrunde liegen.

Welche Aufgabenmerkmale die relative Aufgabenschwierigkeit für die Schweiz wesentlich beeinflussen, bestimmen wir mittels einer multiplen Regression. Die Hypothesen, die am Ende des vorigen Abschnitts formuliert wurden, geben an, welche Aufgabenmerkmale als Prädiktoren einzubeziehen sind. Es sind dies die kognitiven Anforderungen der Aufgaben, die begrifflichen und thematischen Anforderungen sowie die internationale Aufgabenschwierigkeit. Zusätzlich beziehen wir das Aufgabenformat und das Ausmass der Behandlung im Unterricht ein. Diese Faktoren werden in den folgenden Abschnitten vorgestellt.

2.2 *Begriffliche und thematische Anforderungen*

Nach unseren Hypothesen erwarten wir, dass Aufgaben, die spezifisches Fachwissen und insbesondere die Kenntnis schwieriger Begriffe und Bezeichnungen voraussetzen, für die Schweiz besonders schwierig sind. Um diese Aspekte der Aufgaben zu erfassen, wurden vier Lehrpersonen aus Bern und Umgebung gebeten, die Testaufgaben nach ihren begrifflichen und ihren thematischen Anforderungen auf einer fünf- bzw. vierstufigen Skala einzustufen. Die Kategorien der *thematischen Anforderungen* variieren von «aus dem unmittelbaren Alltag, Erfahrungsbereich» bis zu «Fachwissen, das wohl nur in der Schule erworben wird – falls überhaupt», jene der *begrifflichen Anforderungen* zwischen «einfache, allgemeinverständliche Begriffe» und «spezifische schwierige Fachbegriffe als Fremdworte»[4]. Die eher grobe Erfassung dieser Merkmale, insbesondere die Beschränkung auf Urteile aus der Region Bern, ist vertretbar, da es um eine praxisorientierte Erfassung von Aufgabenmerkmalen geht. Ihr Nutzen ist dann gezeigt, wenn die Besonderheiten des schweizerischen Testergebnisses damit besser verstanden werden können.

Bei diesen Einschätzungen geht es uns primär um die begrifflichen Anforderungen. Es wurden aber zwei Beurteilungen verlangt, um überprüfen zu können, ob die Lehrpersonen zwischen den begrifflichen und den inhaltlichen Anforderungen unterscheiden und ob diese Aspekte bei den Aufgaben nicht eng kovariieren. Wenn die Unterscheidung gelingt, so können die begrifflichen Anforderungen – als von den inhaltlichen getrennt – mehr der mangelnden terminologischen als der mangelnden inhaltlichen Vertrautheit des Begriffs zugeordnet werden.

Zu allen Aufgaben liegen je acht Einschätzungen vor, nämlich jeweils die Urteile von vier Personen zu zwei Aspekten. Wie gut stimmen die Lehrpersonen in ihrer Beurteilung überein? Unterscheiden sie konsistent zwischen den beiden Dimensionen? Für die Qualitätsprüfung liegt strukturell eine ähnliche Situation vor, wie wenn zwei Likert-Skalen aus je vier Items gebildet werden. Dort wird die Zuordnung der Items zu verschiedenen inhaltlichen Dimensionen faktorenanalytisch anhand der Antworten von Personen geprüft. Anschliessend wird pro Skala die Reliabilität geschätzt. Analog werden hier die zwei mal acht Einschätzungen der Aufgaben analysiert. Die Faktorenanalyse ergibt zwei Faktoren, die, wie erwartet, den beiden unterschiedlichen Aspekten der Anforderungen entsprechen (vgl. Tabelle 1). Die Reliabilität der beiden Aufgabenmerkmale (Mittelwerte der jeweiligen vier Urteile) ist bei den begrifflichen Anforderungen gut, bei den thematischen Anforderungen befriedigend. Insgesamt kann gesagt werden, dass über die Beurteilungen durch die Lehrpersonen die beiden Dimensionen der begrifflich-terminologischen und der inhaltlichen Anforderungen zuverlässig erfasst werden und dass sich die beiden Dimensionen deutlich voneinander abheben ($r = .51$).

unrotierte Faktoren:	Eigen-wert	% erklärte Varianz	rotierte Faktoren	Eigenwert	% erklärte Varianz	Ladungen der einzelnen Urteile	
						zugehörig	andere
1	3.635	45.4	begriffliche Anforderungen	2.648	33.1	.74-.85	.01-.49
2	1.132	14.1	thematische Anforderungen	2.119	26.5	.58-.80	.09-.37
3	0.866	10.8	**Skalen (Mittelwerte der Urteile)**	**Mittelwert**	**Standard-abw.**	**Reliabilität (Cronbachs α)**	
4	0.690	8.6	begriffliche Anforderungen	2.26	0.97	.81	
			thematische Anforderungen	2.90	0.68	.69	

Hauptkomponentenmethode, Varimax-Rotation

Tabelle 1: Einschätzung der Aufgaben nach begrifflichen und thematischen Anforderungen (Faktoren und Reliabilitätsanalyse der Urteile)
Anzahl der Aufgaben (N) = 135

2.3 Kognitive Anforderungen

Nach unseren Hypothesen erwarten wir unter anderem, dass die Schweiz bei Aufgaben relativ gut abschneidet, bei denen es um naturwissenschaftliche Methoden, Problemlösen, das Verständnis von Zusammenhängen und deren Anwendung auf konkrete Situationen geht. Im Gegensatz dazu stehen Aufgaben, die das blosse Reproduzieren von Wissen prüfen. Bei diesem Aspekt geht es somit um die kognitiven Prozesse, die für die Lösung der Aufgabe erforderlich sind und nicht um deren spezifischen Inhalte. Damit wird die alte Unterscheidung zwischen formaler und materialer Bildung angesprochen (z.B. Reusser & Reusser, 1994). Die neuere kognitionswissenschaftlich orientierte pädagogische Psychologie zielt auf die Förderung solcher formaler Kompetenzen. Sie hat aber gezeigt, dass es nicht um einen Gegensatz zu materialen Kompetenzen, sondern um eine Ergänzung geht: Verstehen und Problemlösen setzen eine gut strukturierte, bereichsspezifische Wissensstruktur voraus und sind auf konkrete Kontexte bezogen (z.B. Resnick 1989; Stebler et al. in diesem Heft; Weinert & Schrader 1997). Ohne Wissensstruktur kann kein fachliches Problem gelöst, geschweige denn ein natürliches Phänomen verstanden werden.

Auch wenn formale und materiale Kompetenzen zur Lösung anspruchsvoller Probleme zusammenspielen müssen, macht es Sinn, diese beiden Aspekte zu unterscheiden. Wenn wir hier Aufgaben analysieren, geht es nicht direkt um die formalen Kompetenzen von Personen, sondern um die Ansprüche, welche

die Aufgaben diesbezüglich stellen. Wir sprechen daher von den *kognitiven Anforderungen* der Aufgaben. Zwischen den Anforderungen der Aufgabe und den zur Lösung eingesetzten kognitiven Prozessen besteht kein eindeutiger Zusammenhang. Je nach Vorkenntnissen und Entwicklungsstand verlangt die gleiche Aufgabe von den einen Lernenden komplexes Schlussfolgern, während andere bereits eine entsprechende Routine erworben haben und sie zur Lösung einsetzen können. Das ändert allerdings nichts daran, dass einige Aufgaben ausschliesslich das Abrufen von Einzelwissen testen, während andere in der Regel auch komplexere kognitive Verarbeitungsprozesse erfordern. Die kognitiven Anforderungen sind damit ein unverzichtbares Merkmal der Aufgabenanalyse.

Die Definition der kognitiven Anforderungen ist bei Moser et al. (1997) und Ramseier (1997) dargestellt. Sie stützt sich auf die Dimension der Leistungserwartungen (performance expectations) des Kategoriensystems, welches TIMSS sowohl zur Auswahl und Einordnung der Testaufgaben als auch zur inhaltsanalytischen Beschreibung von Lehrplänen und Lehrmitteln verwendet (Robitaille et al. 1993). Die Leistungserwartungen in den Naturwissenschaften sind bei TIMSS in 20 Kategorien, gebündelt in fünf Hauptkategorien, unterteilt. Die Testaufgaben wurden von den Vertretern der an TIMSS beteiligten Länder während der Testkonstruktion per Expertenurteil diesen Kategorien zugeteilt. Diese Zuordnung der Aufgaben zu genau einer Kategorie ist sicher in manchen Fällen diskutabel. Wir können dennoch von ihr ausgehen, da allfällige Ungenauigkeiten verhindern würden, dass überhaupt Effekte nachweisbar wären.

Die Aufgaben verteilen sich sehr ungleich auf die Kategorien des TIMSS-Rasters. Um eine auch zahlenmässig sinnvolle Verteilung der Aufgaben zu erhalten, wird die erste Kategorie der kognitiven Anforderungen mit der Kategorie «Einzelwissen» (understanding simple information) von TIMSS gleichgesetzt, die zweite mit der Kategorie «Komplexes Wissen» (understanding complex information). In der dritten Kategorie der kognitiven Anforderungen, «Wissen entdecken und anwenden», werden alle übrigen Kategorien von TIMSS zusammengefasst. Die Bezeichnungen der am stärksten belegten dieser TIMSS-Kategorien charakterisieren gut, worum es bei diesen Aufgaben geht: «applying scientific principles to develop explanations», «applying scientific principles to solve quantitative problems», «designing investigations» und «interpreting data».

Durch unsere Zusammenfassung der TIMSS-Kategorien entsteht eine dreistufige Dimension zur Charakterisierung der Testaufgaben, die eine offensichtliche Zunahme an Komplexität der verlangten kognitiven Prozesse widerspiegelt: vom Reproduzieren einfachen Wissens über das Reproduzieren komplexen Wissens zum Entdecken und Anwenden von Wissen. Wir können daher, im Gegensatz zu den vielfältigeren Leistungserwartungen bei TIMSS, von einer hierarchisch geordneten, d.h. ordinalen Dimension ausgehen. Mit dieser hierarchischen Dimension sind wir - wenn auch nur mit einer groben Einteilung - wieder in der Nähe der Bloomschen Taxonomie von Lernzielen, von der TIMSS und die früheren IEA-Studien ausgingen (Robitaille et al. 1993: 41). Auch diese Bloomsche Taxonomie geht von einer Hierarchie aus. Dort beschreiben die

Kategorien Kenntnisse, Verständnis, Anwenden, Analyse und Synthese immer höhere Ebenen kognitiver Tätigkeit (Gage & Berliner 1977: 46 f.).

Es ist wichtig, die Dimension der kognitiven Anforderungen von der generellen Schwierigkeit der Aufgaben zu unterscheiden. In einem einfachen Modell können wir davon ausgehen, dass sich die Schwierigkeit der Aufgabe aus verschiedenen Komponenten ergibt[5]. Eine solche Komponente bilden die kognitiven Anforderungen. Daneben hat es aber weitere Komponenten, wie z.B. die begriffliche oder inhaltliche Komplexität. «Wissen zu reproduzieren ist als kognitiver Prozess einfach strukturiert; wenn aber nach schwierigen Fremdwörtern und Fachbegriffen gefragt wird, können Schülerinnen und Schüler nur selten die richtige Antwort geben: Bei gleichartigen kognitiven Anforderungen variiert die Schwierigkeit der Aufgaben, z.B. je nach terminologischer Komplexität, beträchtlich» (Moser et al. 1997: 62 f.).

2.4 Internationale Aufgabenschwierigkeit

Mit den kognitiven, thematischen und begrifflichen Anforderungen haben wir spezifische Komponenten der Aufgabenschwierigkeit erfasst. Sicher ist deren Messung nicht perfekt, und es wird weitere, nicht gemessene Komponenten geben. All diese Komponenten, soweit sie nicht spezifisch schweizerisch sind, kommen in der internationalen Aufgabenschwierigkeit zum Ausdruck. Wir berücksichtigen auch dieses Merkmal in der Aufgabenanalyse, um unsere Hypothese zu testen, dass die Schweiz bei generell schwierigen Aufgaben besonders Mühe hat.

2.5 Behandlung im Unterricht

Das internationale Projekt TIMSS hat alle Länder eine Beurteilung der Angemessenheit der Testaufgaben vornehmen lassen. Zu beurteilen war, ob die Mehrheit der Schülerinnen und Schüler dem *Thema* der Aufgabe vor dem Testzeitpunkt im Unterricht begegnet sei. Diese Beurteilung erlaubt es jedem Land, einen internationalen Vergleich allein auf der Basis jener Aufgaben vorzunehmen, die es selbst als seinem Unterricht angemessen einstuft (vgl. Beaton et al. 1996a: B-1).

Für die Schweiz schätzten Lehrpersonen aus neun Kantonen aus allen Sprachregionen diese Angemessenheit, bezogen auf die jeweiligen Lehrpläne, ein (Moser et al. 1997). Die Beurteilung fällt je nach Kanton und Schultyp unterschiedlich aus. Unsere Aufgabenanalyse bezieht sich auf die Schweiz als Ganzes. Es muss daher ein gesamtschweizerisches Mass konstruiert werden, das wiedergibt, wie weit die Themen der Aufgaben bis zum oberen von TIMSS getesteten Schuljahr behandelt worden sind. Dazu wird im folgenden die mittlere Einstufung einer Aufgabe verwendet. Bei dieser Mittelung wurden die regionalen Einstufungen - ihrerseits Mittelwerte zwischen Kantonen - entsprechend ihrer Vertretung in der Stichprobe gewichtet. Ein Wert von z.B. 40 zeigt somit an,

dass das zur Aufgabe gehörende Thema über alle Beurteilungen hinweg zu 40% als behandelt eingeschätzt wurde. Diese Grösse wird kurz als *«Behandlung im Unterricht»* bezeichnet. Der Mittelwert liegt bei 63%. Fast ein Drittel der Aufgaben wird mehrheitlich als nicht behandelt eingestuft.

Unsere Ausgangshypothesen (vgl. Ende von Abschnitt 1.3) beziehen sich nicht auf das Ausmass der Behandlung im Unterricht. Es ist aber zu vermuten, dass die relative Schwierigkeit der Aufgaben dadurch beeinflusst wird. Diese Variable ist für sich von Interesse und könnte die uns interessierenden Zusammenhänge tangieren. Sie soll deshalb in die Aufgabenanalyse einbezogen werden.

2.6 Format der Aufgaben

Der schriftliche TIMSS-Leistungstest setzt sich aus Aufgaben mit drei unterschiedlichen Formaten zusammen. Aus Gründen der Effizienz besteht er hauptsächlich aus Multiple-Choice-Aufgaben. Gerade um komplexere Leistungen zu erfassen, enthält er aber auch Aufgaben, die eine kurze Antwort, z.B. in Form der richtigen Zahl, oder aber eine ausführliche Antwort, z.B. die Beschreibung eines Lösungsablaufs oder eines Zusammenhangs in eigenen Worten, verlangen. Das Aufgabenformat interessiert uns gemäss unseren Hypothesen nicht direkt. Innerhalb des schriftlichen Tests bestimmt es aber einen Aspekt des Handlungskontextes, mit dem Schülerinnen und Schüler je nach der Prüfungspraxis in ihrem Land unterschiedlich vertraut sind. Da in der Schweiz Multiple-Choice-Aufgaben wenig üblich sind, könnte die relative Schwierigkeit der Mulitple-Choice-Aufgaben für die Schweiz tendenziell höher sein und so die Aufgabenanalyse beeinflussen. Es muss daher geprüft werden, ob das Aufgabenformat unsere Analyse beeinflussen kann.

Es zeigt sich, dass die Mulitple-Choice-Aufgaben generell leichter sind als die offenen: Im internationalen Mittel werden sie nur zu 41% falsch gelöst, die offenen dagegen zu 52%. Unsere Vermutung bestätigt sich nur tendenziell: Die *relative* Schwierigkeit für die Schweiz ist bei den Multiple-Choice-Aufgaben um 2.7 Prozentpunkte höher als bei den offenen Aufgaben (Unterschied knapp nicht signifikant, $p = .06$).

Die Multiple-Choice-Aufgaben wurden im Test gezielt durch andere Formate ergänzt, um komplexere Leistungen besser zu erfassen. Es ist deshalb zu vermuten, dass die Aufgaben mit offenen Antworten so gewählt wurden, dass sie vermehrt höhere kognitive Anforderungen erfordern. Tatsächlich stellen die Aufgaben mit offenen Antworten signifikant ($p < .0001$) und bedeutsam höhere kognitive Anforderungen als die Multiple-Choice-Aufgaben[6]. Wenn Aufgaben mit niedrigen kognitiven Anforderungen vermehrt im Multiple-Choice-Format gestellt wurden und dieses den schweizerischen Schülerinnen und Schülern tendenziell etwas Mühe macht (die relative Schwierigkeit also hoch ist), so erwarten wir einen negativen Zusammenhang zwischen kognitiven Anforderungen und der relativen Schwierigkeit für die Schweiz, der nichts mit den Anforderungen zu tun hat, sondern nur mit dem Aufgabenformat. Wir müssen daher diesen

formalen Aspekt der Aufgaben einbeziehen, um unzulässige Interpretationen zu vermeiden.

3 Ergebnisse

3.1 Multiple Regressionsanalyse

Wir haben insgesamt sechs Aufgabenmerkmale, von denen wir erwarten, dass sie zur relativen Aufgabenschwierigkeit für die Schweiz beitragen: Kognitive, begriffliche und thematische Anforderungen, internationale Schwierigkeit, Behandlung im Unterricht und Aufgabenformat. Sie alle beziehen wir als Prädiktoren in eine multiple Regressionsanalyse mit der relativen Aufgabenschwierigkeit für die Schweiz als abhängiger Variablen ein. Damit können wir nicht nur feststellen, ob ein solches Aufgabenmerkmal mit dem Kriterium zusammenhängt (Korrelationskoeffizient r), sondern auch, ob es sich als eigenständige Determinante der relativen Schwierigkeit für die Schweiz erweist, d. h. ob es auch dann zur Vorhersage des Kriteriums beiträgt, wenn alle übrigen Merkmale konstant gehalten werden (standardisierter Regressionskoeffizient β).

Eine erste Analyse mit allen sechs Prädiktoren zeigt, dass die thematischen Anforderungen ($r = .27$), die Behandlung im Unterricht ($r = -.26$) und das Aufgabenformat ($r = .16$) zwar mit der relativen Aufgabenschwierigkeit für die Schweiz korrelieren, dass sie aber keinerlei eigenständigen Erklärungswert haben, der über jenen der anderen Prädiktoren hinausgeht (für alle drei Variablen ist $|\beta| < .04$ und $p > .69$). Damit ist gezeigt, dass das Aufgabenformat in unserer Analyse als Störvariable keine Rolle spielt. Vor allem erweist es sich, dass die Behandlung im Unterricht und die thematischen Anforderungen nicht spezifisch zur Erklärung der relativen Schwierigkeit für die Schweiz beitragen.

In die zweite, in Tabelle 2 dargestellte Regressionsanalyse sind nur noch die relevanten Prädiktoren aufgenommen worden. Sie zeigt ein hochsignifikantes Ergebnis; zusammen erklären die Determinanten 25% der Varianz in der relativen Schwierigkeit der Aufgaben für die Schweiz. Wie sich am standardisierten Regressionskoeffizienten ablesen lässt, haben die kognitiven Anforderungen den stärksten Einfluss auf die relative Aufgabenschwierigkeit für die Schweiz. Auch die begrifflichen Anforderungen erweisen sich klar als eigenständige Determinante der relativen Schwierigkeit für die Schweiz. Dies obwohl die Variable ja nur auf der Einschätzung von Lehrpersonen aus der Region Bern beruht. Die internationale Aufgabenschwierigkeit erfasst weitere, nicht näher spezifizierte Komponenten der Aufgabenschwierigkeit, die aber von den beiden genannten spezifischen unabhängig sind. Überlegt man sich, was denn neben den kognitiven und begrifflich-terminologischen Anforderungen noch plausible Komponenten der Aufgabenschwierigkeit sein können, so kommen am ehesten fachlich-inhaltliche in Frage. Per regional verankertem Expertenurteil haben wir zwar solche thematischen Anforderungen erfasst. Dies gelang aber offenbar

schlechter als durch eine Messung via Lösungshäufigkeiten, d.h. über die nach kognitiven und begrifflichen Anforderungen kontrollierte internationale Aufgabenschwierigkeit. Verschiedene Varianten der Regressionsanalyse verdeutlichen dieses Zusammenspiel der Prädiktoren und zeigen die Robustheit der Ergebnisse gegenüber technischen Variationen (Ramseier 1997).

unabhängige Variable	Standard. Regressionskoeffizient (β)	Standardfehler von β	Semipart. Korrelation	Korrelations koeffizient	t	Signifikanz von t
kognitive Anforderungen	-.37	0.08	-.35	-.36	-4.85	< .001
internationale Schwierigkeit	.24	0.08	.23	.19	3.21	.002
begriffliche Anforderungen	.22	0.08	.22	.33	2.97	.003

Multipler Korrelationskoeffizient:
R = .50; Varianzanalyse zur Regressionsgleichung: F = 15.93; df = 3/142; p < .0001.
(Outlier, Multikollinearität und Korrelation zwischen Residuen und abhängiger Variable stören die Regressionsanalyse nicht.)

Tabelle 2: Multiple Regression der relativen Schwierigkeit für die Schweiz nach Aufgabenmerkmalen

Wir haben die relative Aufgabenschwierigkeit für die Schweiz analysiert. Wir können die gleichen Prädiktoren verwenden, um die relative Aufgabenschwierigkeit für andere Länder zu untersuchen. Wenn es stimmt, dass wir mit unserer Aufgabenanalyse bestimmte Aufgabenmerkmale bzw. ein Leistungsprofil gefunden haben, das die Prioritäten des schweizerischen Unterrichts spiegelt, so muss dieselbe Analyse in Ländern mit anderen Prioritäten zu einem anderen Resultat führen. Wir haben die gleiche Analyse für die USA und für Tschechien gemacht, von denen wir um andere Unterrichtsschwerpunkte wissen. Bei diesen Ländern trägt wie erwartet keiner der Prädiktoren signifikant zur Erklärung der relativen Aufgabenschwierigkeit bei.

Unsere Analyse bezieht sich auf die *relative* Schwierigkeit der Aufgaben für die Schweiz. Es ist wichtig, dies klar von einer Analyse der absoluten Aufgabenschwierigkeit zu unterscheiden. Wenn wir in einer Regressionsanalyse die *absolute* Schwierigkeit für die Schweiz, also den Prozentsatz falscher Lösungen, als Kriterium und die kognitiven, thematischen und begrifflichen Anforderungen sowie die Behandlung im Unterricht als Prädiktoren verwenden, so sieht das Resultat völlig anders aus. Um diese absolute Schwierigkeit für die Schweiz vorauszusagen (R = .47), sind die thematischen Anforderungen (β = .33, p = .001) der wichtigste Prädiktor, die Behandlung im Unterricht (β = -.19, p = .03) und tendenziell die kognitiven Anforderungen (β = .15, p = .05; Vorzeichenwechsel!) spielen auch eine Rolle, während die begrifflichen Anforderungen daneben bedeutungslos sind (β = .03, p = .74). Die thematischen Anforderungen,

also die Einschätzung der Vertrautheit/Schwierigkeit der Thematik der Aufgaben durch die Lehrpersonen, sagt die Lösungshäufigkeit in der Schweiz gut voraus – eine unmittelbare Validierung dieser Einschätzung. Obwohl höhere kognitive Anforderungen den schweizerischen Schülerinnen und Schülern – im Vergleich zu ihren ausländischen Kolleginnen und Kollegen – ja entgegenkommen, machen sie die Aufgaben auch in der Schweiz schwieriger – allerdings weniger als im internationalen Mittel. Interessant ist, dass es bei der Voraussage der absoluten Schwierigkeit – im Gegensatz zur relativen – eine wesentliche Rolle spielt, ob das Thema einer Aufgabe im Unterricht behandelt wurde oder nicht. Zum Glück ist das so, denn sonst müsste man an der Wirksamkeit des Unterrichts zweifeln.

Wenn wir über die Bedeutsamkeit der Prädiktoren reden, müssen wir also klar unterscheiden, ob es *allgemein* um die Erklärung der Schwierigkeit der Aufgaben in der Schweiz geht, oder darum, was die *Besonderheit* der Schwierigkeit für die Schweiz – im Gegensatz zur Mehrheit der anderen Länder – ausmacht. Im ersten Falle analysiert man nicht mehr schweizerische Besonderheiten. Und grundsätzlich sind es dieselben Faktoren, die eine Aufgabe in den verschiedenen Ländern schwierig machen. Diese Faktoren können dann aber in ihrer Bedeutung zwischen den Ländern variieren.

3.2 Internationaler Vergleich und Position der Schweiz

Ein Anlass dieser Aufgabenanalyse war, dass die Schweiz bei einer Rangierung der Länder nach ihren Leistungen in den Naturwissenschaften bei TIMSS weit schlechter abschneidet als in der Mathematik und bei IAEP. Verändert sich diese Rangierung wesentlich, wenn man dabei nur jene Aufgaben berücksichtigt, die für die Schweiz günstige Merkmale aufweisen? Steht die Schweiz im internationalen Vergleich deutlich anders da, wenn man auf Aufgaben abstellt, die ihrem Unterricht entsprechen oder dazu kontrastieren? Zur Klärung dieser Fragen unterteilen wir die Aufgaben in zwei Gruppen: Auf der einen Seite stehen 32 Aufgaben mit hohen kognitiven Anforderungen (Kategorie «Wissen entdecken und anwenden»), die gemessen an den begrifflichen Anforderungen zur leichteren Hälfte gehören. Zur anderenGruppe gehören 33 Aufgaben, die niedrige kognitive Anforderungen stellen und gleichzeitig zur begrifflich schwierigeren Hälfte der Aufgaben zählen. Anders als bei Moser et al. (1997: 30) betrachten wir hier alle 41 Länder, die in die internationale TIMSS-Auswertungen aufgenommen wurden – unabhängig von der Qualität ihrer Stichproben. Während die Schweiz bei einer Rangierung nach dem Prozentsatz richtiger Lösungen in der Gesamtheit der Aufgaben lediglich Rang 25 erreicht (Beaton et al. 1996a: 41), kommt sie bei den Aufgaben mit hohen kognitiven und niedrigen begrifflichen Anforderungen auf Platz 14 – ein Ergebnis, das nicht mehr weit hinter demjenigen in der Mathematik zurückliegt (Rang 8, Beaton et al. 1996b: 41). Ränge sind ein oberflächliches und unstabiles Merkmal. Aussagekräftiger sind Leistungsdifferenzen. Die Differenz in der relativen Aufgabenschwierigkeit für die Schweiz bei den Aufgaben mit hohen kognitiven und niedrigen begrifflichen

Anforderungen einerseits und den Aufgaben mit gegenteiligen Anforderungen andererseits beträgt 8.5 Prozentpunkte. Zum Vergleich: Im internationalen Gesamtmittel werden im 8. Schuljahr etwa 6% mehr Aufgaben richtig gelöst als im 7. Schuljahr. Der Unterschied in der relativen Aufgabenschwierigkeit der beiden Aufgabengruppen ist also als praktisch bedeutsam einzustufen.

Die relative Schwierigkeit der beiden Aufgabengruppen unterscheidet sich auch für manche andere Länder; und zwar weit mehr, als zufällig zu erwarten ist (Ramseier 1997). Insgesamt zehn Länder haben eine statistisch signifikante und bedeutsame Abweichung in die gleiche Richtung wie die Schweiz. Als bedeutsam sehen wir dabei wieder eine Abweichung über 6 Prozentpunkte an. Ausser Neuseeland gehören all diese Länder zum westlichen bzw. nördlichen Europa. Für diese Länder gilt, dass ihre Schülerinnen und Schüler Aufgaben relativ besser bewältigen, die kognitiv anspruchsvoll sind und gleichzeitig eher einfache Begriffe prüfen, als Aufgaben mit gegenteiligen Merkmalen.

Für manche Länder, so die Vereinigten Staaten, Deutschland, Singapur oder Korea, spielt diese Kombination von Aufgabenmerkmalen keine Rolle: Für sie sind die dadurch charakterisierten Teile des TIMSS-Tests ähnlich schwierig.

Auf der anderenSeite steht eine Gruppe von neun Ländern, bei denen die mittlere relative Leichtigkeit der Aufgaben mit hohen begrifflichen und niedrigen kognitiven Anforderungen signifikant und um mindestens 6% über dem Mittelwert in der Gegengruppe liegt. Zu diesen Ländern gehören ausser Hongkong ausschliesslich Länder, die im Gesamttest sehr schwache Leistungen haben, sowie manche osteuropäische Länder. Interessant ist, dass alle osteuropäischen Länder und alle neun Länder mit der schwächsten Gesamtleistung eine Tendenz in diese Richtung haben, die allerdings nicht immer signifikant ist. Hongkong ist ein interessanter Kontrast zur Schweiz, da dieses Land wie die Schweiz in Mathematik weit besser abgeschnitten hat als in den Naturwissenschaften (Rang 4 bzw. 22, vgl. Beaton et al. 1996a: 41; 1992b: 41), dann aber bei naturwissenschaftlichen Aufgaben ein zur Schweiz gegensätzliches Leistungsprofil aufweist. Dies bedeutet zumindest, dass es nicht eine quasi automatische Folge guter mathematischer Kenntnisse ist, wenn die Schweiz bei den kognitiv anspruchsvollen und begrifflich einfachen Aufgaben relativ gute Resultate erzielt.

Im Falle der Schweiz sind wir von bestimmten Prioritäten im Unterricht ausgegangen, die uns zu Erwartungen über Leistungen in verschiedenen Teilen des naturwissenschaftlichen Tests von TIMSS geführt haben. Wir stellen fest, dass andere Länder ähnliche, indifferente oder aber gegenteilige Leistungsprofile zeigen, und dass sich diese Profile sogar z.T. grossen geographischen Regionen zuordnen lassen. Es scheint, dass sich Gruppen von Ländern bilden lassen, die sich durch ähnliche curriculare Prioritäten auszeichnen. Dies ist eine Hypothese, die durch eine Analyse der Curricula dieser Länder überprüft werden sollte.

3.3 Unterschiede nach Fachgebieten

Die Leistungen der Schweiz liegen in Physik und Erdkunde (naturkundlicher Teil der Geographie) über, in der Chemie unter ihrem durchschnittlichen Leistungsniveau, während sie in Biologie etwa diesem mittleren Niveau entsprechen (Beaton et al. 1996a). Das ist gleichbedeutend mit der Aussage, dass es in der relativen Schwierigkeit der Aufgaben für die Schweiz Unterschiede zwischen diesen Fächern gibt. Damit gibt es eine Beziehung zwischen der Aufgabenanalyse und den Fächern. Einerseits könnte es sein, dass das, was wir auf kognitive oder z.B. begriffliche Anforderungen zurückgeführt haben, eigentlich auf Unterschieden zwischen Fächern beruht. Dem ist jedoch nicht so. Wenn wir in die in Abschnitt 3.1 beschriebene Regressionsanalyse auch die Fachzugehörigkeit der Aufgabe einführen, so ändert sich die Bedeutung der ursprünglichen Prädiktoren nicht nennenswert (Ramseier 1997).

Andererseits kann uns die Aufgabenanalyse möglicherweise helfen, die zum Teil überraschenden Unterschiede zwischen den Fachleistungen zu verstehen. Die Themen der Aufgaben aus Physik und Chemie werden nach unserer Experteneinschätzung bis zum TIMSS-Test seltener behandelt als jene aus Erdkunde und Biologie. Aufgrund dieser Unterrichtsschwerpunkte ist es eher unerwartet, dass die Physik und nicht die Biologie zu den für die schweizerischen Schülerinnen und Schüler eher leichten Fächern zählen. In einer Kovarianzanalyse kann man untersuchen, ob die Unterschiede in der relativen Aufgabenschwierigkeit für die Schweiz zurückgehen, wenn man nach den hier besprochenen Prädiktoren der relativen Aufgabenschwierigkeit kontrolliert. Dies ist tatsächlich der Fall (Ramseier 1997). Sieht man sich die Verteilung der verschiedenen Aufgabenmerkmale auf die Fächer an, so werden die Leistungsunterschiede zwischen den Fächern verständlich. In der *Chemie* kumulieren sich die für die Schweiz ungünstigen Aufgabenmerkmale: eher niedrige kognitive Anforderungen, schwierige Begriffe und Themen und hohe internationale Schwierigkeit. In der *Erdkunde* ergibt sich umgekehrt eine der Schweiz entgegenkommende Konstellation: bescheidene thematische und begriffliche, aber hohe kognitive Anforderungen. Kein Wunder, dass das Ergebnis in diesem Fach für die Schweiz am besten ist. In *Biologie* sind zwar die Aufgaben generell leicht, es werden aber vermehrt schwierige Begriffe und Themen geprüft und selten höhere kognitive Anforderungen gestellt. Die Mischung aus für die Schweiz günstigen und ungünstigen Faktoren passt zum durchschnittlichen Ergebnis in diesem Fach. Das unerwartet gute Ergebnis in der *Physik* hat damit zu tun, dass relativ einfache, vertraute Begriffe und Themen geprüft und gleichzeitig recht hohe kognitive Anforderungen gestellt werden.

3.4 TIMSS-Experimentiertest

Der Leistungstest von TIMSS beruht hauptsächlich auf Multiple-Choice-Aufgaben. Diese werden zwar durch eine Reihe von Aufgaben mit offenen Antworten ergänzt. Dennoch kann man mit solchen schriftlichen Tests nicht die ganze

Spanne naturwissenschaftlicher Leistungen messen. Insbesondere wird kaum erfasst, wie eine komplexe, lebensnahe Aufgabe durch eigenständiges Erproben von Lösungswegen gelöst wird. Diese Begrenzung hat gerade in Ländern, bei denen standardisierte Leistungstests eine grosse Rolle spielen, zur Entwicklung alternativer Messverfahren geführt (z.B. Baxter et al. 1996). Zu diesen Alternativen zählen Portfolios, also Dossiers, in denen Produkte und Arbeitsergebnisse der Lernenden (z.B. Projektarbeiten) über eine längere Zeit gesammelt werden. Eine zweite Form sind *performance assessments,* bei denen komplexe Problemstellungen vorgegeben werden, welche die Prüflinge mit Hilfe von Materialien und durch Experimentieren bearbeiten. TIMSS hat diese neue Testform aufgegriffen und in einem eigenen Untersuchungsteil den schriftlichen Leistungstest durch einen solchen Experimentiertest *(performance assessment)* ergänzt. Die Schweiz hat sich daran beteiligt.

Unsere Analyse der Aufgaben des schriftlichen Tests in den Naturwissenschaften zeigt, dass die Schweiz gerade bei höheren kognitiven Anforderungen, z.B. beim Interpretieren von Daten, relativ gute Leistungen zeigt. Aufgrund der Aufgabenanalyse ist daher zu erwarten, dass die Schweiz im TIMSS Experimentiertest nicht nur in Mathematik, sondern auch in den Naturwissenschaften gut abschneidet. Das Ergebnis des Experimentiertests entspricht genau diesen Erwartungen. In beiden Fächern erreicht die Schweiz einen Platz in einer Spitzengruppe, die nur von Singapur und z.T. England übertroffen wird (Harmon et al. 1997; Stebler et al. in diesem Heft). Die Ergebnisse der Aufgabenanalyse und des Experimentiertests sind in hohem Masse konsistent und bestätigen sich gegenseitig.

4 Zusammenfassung und Schlussfolgerungen

Aus der Beschreibung des naturwissenschaftlichen Unterrichts und aus dem Vergleich mit der Untersuchung IAEP lässt sich ableiten, welche TIMSS-Aufgaben für die Schweiz vergleichsweise leicht oder aber schwierig sein müssen. Ob dies zutrifft, wurde in einer Aufgabenanalyse überprüft. Als Kriterium für die spezifische, relative Schwierigkeit der Aufgaben für die Schweiz wurde die Differenz zwischen dem Prozentsatz falscher Lösungen in der Schweiz und im internationalen Mittel verwendet. Die Ergebnisse einer multiplen Regressionsanalyse nach verschiedenen Merkmalen der Aufgaben entsprechen in hohem Masse den Erwartungen. In der Schweiz wurden jene Aufgaben vergleichsweise gut gelöst, die insgesamt eher einfach sind, in denen nicht schwierige Begriffe vorausgesetzt werden, die aber relativ hohe kognitive Anforderungen stellen und nicht nur Wissen abfragen. Wenn das Thema der Aufgabe leicht ist und wenn es in der Schweiz behandelt wurde, so erhöht sich zwar für die Schweiz wie für die anderen Länder die Häufigkeit einer richtigen Lösung; diese Aspekte tragen aber neben den genannten Faktoren nichts zur Aufklärung der *besonderen* Schwierigkeit für die Schweiz bei. Für diese sind somit neben der gene-

rellen Schwierigkeit, gemessen am internationalen Mittel falscher Lösungen, lediglich die kognitiven und die terminologisch-begrifflichen Anforderungen relevante Prädiktoren. In den USA und Tschechien, wo der Unterricht andere Schwerpunkte hat, sind diese Prädiktoren dagegen unwirksam.

Die eben beschriebene Wirkung der Aufgabenmerkmale und das damit hervorgerufene Leistungsprofil passen zur vermuteten Ausgangslage und bestätigen sie damit: Der naturkundliche Unterricht in der Schweiz, mit seinem zeitlich beschränkten Umfang und seinem Schwerpunkt auf dem Verstehen von Zusammenhängen und der Beziehung zwischen Individuum und Natur, bereitet bis zum Testzeitpunkt besser auf eher einfache Aufgaben vor. Vor allem in der Fachterminologie ist ein Rückstand festzustellen, der zu entsprechenden Fehlern im TIMSS-Test führt. Hingegen gelingt es schweizerischen Schülerinnen und Schülern gut, naturwissenschaftliches Wissen anzuwenden, Daten zu interpretieren und Schlüsse zu ziehen. Die Prioritäten im schweizerischen Curriculum haben Konsequenzen, die sich im internationalen Leistungsvergleich zeigen.

Die diagnostizierte Lücke im Bereich von Fachbezeichnungen und Wissen wirft die Frage auf, ob diese in der Schule mehr gefördert werden sollten. Die Antwort ist abhängig von der Priorität der Ziele, die in der Schule angestrebt werden und so letztlich bildungspolitisch bestimmt. Einige Bewertungen und Argumente dazu sollen dennoch genannt werden. Die Kenntnis von Fachbegriffen ist notwendig, um naturwissenschaftliche Phänomene verstehen zu können – Verstehen und Problemlösen setzen eine gut strukturierte, bereichsspezifische Wissensstruktur voraus. Der TIMSS-Test und diese Analyse regen an, im Detail zu prüfen, ob in der Schule der Aufbau solcher Wissensstrukturen im naturwissenschaftlichen Bereich genügend gefördert wird, evtl. auch erst in den Schuljahren nach dem TIMSS-Test. Die Priorität von Verstehen und Anwenden und der Verzicht auf die sehr frühe Einführung abstrakter und dann inhaltsloser Begriffe halte ich in der Tendenz aber für richtig. Man stelle sich vor, die Analyse hätte zum gegenteiligen Ergebnis geführt: «Zwar können schweizerische Kinder viel Wissen abrufen; wenn es aber darum geht, Zusammenhänge zu verstehen und anzuwenden, so scheitern sie.» Eine solche Diagnose wäre offensichtlich alarmierender als das heute festgestellte Manko. «Wichtig ist also, bei allfälligen Korrekturen vorhandene Stärken nicht zu gefährden, sondern auf ihnen aufzubauen» (Moser et al. 1997: 77).

Nicht nur die Leistungen der Schweiz, auch jene vieler anderer Länder fallen deutlich anders aus, je nachdem ob man auf Aufgaben mit hohen kognitiven und niedrigen begrifflichen Anforderungen abstellt oder auf solche mit gegenteiligen Merkmalen. Diese Leistungsvariation bleibt bestehen, wenn man die unterschiedliche Schwierigkeit der Aufgaben in den beiden Gruppen kontrolliert. Die Auswirkung auf die Rangierung der Länder ist beträchtlich. Für internationale Leistungsvergleiche ergibt sich daraus, dass die Konzeption des Leistungstests und die Aufgabenauswahl überaus wichtig sind und deutliche Konsequenzen für das Ergebnis des Vergleichs haben.

Anmerkungen

1 Das schweizerische Projekt ist ein Gemeinschaftsunternehmen des Amtes für Bildungsforschung des Kantons Bern (Projektleitung), des Institut romand de recherches et de documentation pédagogiques, Neuchâtel, des Ufficio studi e ricerche, Bellinzona, der Universität Zürich, Päd. Institut, Fachbereich Päd. Psychologie/Didaktik, sowie der Universität Bern, Höheres Lehramt, und wird vom schweizerischen Nationalfonds, Projekt-Nummer 4033-035819, unterstützt.

2 Eine knappe Zusammenfassung der Analyse enthält bereits Moser et al. (1997: 71 ff.). Alle Details der Analyse sind bei Ramseier (1997) dargestellt.

3 Bei einigen Aufgaben mit offenen Antworten wurden auch Punkte für Teillösungen vergeben. Für die Analyse wurde eine solche Aufgabe in zwei oder drei Teile zerlegt, die wieder mit richtig/falsch bewertet werden, je nachdem ob der erste, zweite und allenfalls der dritte Teilpunkt erreicht wurde. Aus 135 ursprünglichen Aufgaben werden so 146 dichotome Testwerte. Diese werden einfachheitshalber im weiteren als Aufgaben bezeichnet. Bei den Tests in diesem Kapitel ist deshalb die vorausgesetzte Unabhängigkeit der Elemente nicht voll erfüllt.

4 Vgl. die vollständige Kategorienbeschreibungen bei Ramseier (1997).

5 Vgl. das Itemkomponenten-Modell bei Rost (1996). Dieses Modell führt auf der Basis der Item Response Theorie zu technisch weit raffinierteren Analysen der Aufgabenschwierigkeit, die dann aber auch weniger anschaulich sind als die hier geschilderte.

6 Mittelwerte M = 2.6 bzw. M = 1.7. Die drei Anforderungsstufen wurden mit 1 (einfaches Wissen), 2 (komplexes Wissen) und 3 (Entdecken und Anwenden) kodiert.

5 Literatur

Baxter, G. P., Elder, A. D. & Glaser, R. (1996). *Knowledge-Based Cognition and Performance Assessment in the Science Classroom. Educational Psychologist, 31, 133-140.*

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Smith, T. A. & Kelly, D. L. (1996a). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzales, E. J., Kelly, D. L. & Smith, T. A. (1996b). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Gage, N. L. & Berliner, D. C. (1977). *Pädagogische Psychologie.* München: Urban & Schwarzenberg.

Harmon, M., Smith, T. A., Martin, M. O., Kelly, D. L., Beaton, A. E., Mullis, I. V.S., Gonzales, E. J. & Orpwood, G. (1997). *Performance Assessment in IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Lapointe, A. E., Askew, J. M. & Mead, N. A. (1992a). *Learning Mathematics. International Assessment of Educational Progress (IAEP).* Princeton (N.J.): Educational Testing Service.

Lapointe, A. E., Askew, J. M. & Mead, N. A. (1992b). *Learning Science. International Assessment of Educational Progress (IAEP).* Princeton (N.J.): Educational Testing Service.

Moser, U. (1992). *Was wissen 13-jährige? Internationaler Vergleich in Mathematik und Naturwissenschaften.* Bern: Amt für Bildungsforschung der Erziehungsdirektion des Kantons Bern.

Moser, U., Ramseier, E., Keller, C. & Huber, M. (1997). *Schule auf dem Prüfstand. Eine Evaluation der Sekundarstufe I auf der Grundlage der «Third International Mathematics and Science Study».* Zürich: Rüegger.

Ramseier, E. (1997). *Naturwissenschaftliche Leistungen in der Schweiz. Vertiefende Analyse der nationalen Ergebnisse in TIMSS.* Bern: Amt für Bildungsforschung der Erziehungsdirektion des Kantons Bern.

Resnick, L. B. (1989). Knowing, Learning, and Instruction: *Essays in Honor of Robert Glaser.* Hillsdale: Erlbaum.

Reusser, K. & Reusser, M. (1994). Verstehen als psychologischer Prozess und als didaktische Aufgabe: Einführung und Überblick. In: Reusser, K. & Reusser, M. (Hrsg.). *Verstehen. Psychologischer Prozess und didaktische Aufgabe* (S. 9-35). Bern: Huber.

Robitaille, D. F. et al. (1993). *Curriculum Frameworks for Mathematics and Science.* Vancouver: Pacific Educational Press.

Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion.* Bern: Huber.

Stebler, R., Reusser, K. & Ramseier, E. (in diesem Heft) (1998). Praktische Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenzen - Erträge aus dem TIMSS-Experimentiertest. *Bildungsforschung und Bildungspraxis, Heft 1.*

Weinert, F. E. & Schrader, F.-W. (1997). Lernen lernen als psychologisches Problem. In: Weinert, F. E. & Mandl, H. (Hrsg.) (1997). *Enzyklopädie der Psychologie. Psychologie der Erwachsenenbildung* (S. 295-335). Göttingen: Hogrefe.

Tendances et piorités de l'enseignement. Analyse des résultats du travail des élèves en Suisse, sur la base du test conçu par l'étude TIMSS pour les disciplines scientifiques.

Résumé

La comparaison établie au niveau international par l'étude TIMSS montre que dans les écoles suisses le niveau en sciences est nettement inférieur à celui observé en mathématiques. La présente analyse, s'appuyant sur les résultats des

tests, a pour but de comprendre les raisons de cet écart. L'étude a cherché à déterminer quels exercices étaient particulièrement difficiles ou particulièrement faciles pour les élèves. A cet effet, des enseignants ont évalué la difficulté des exercices tant du point de vue de leur contenu que du point de vue terminologique. On a constaté que les élèves venaient relativement facilement à bout des exercices qui sollicitent fortement la capacité d'abstraction, quand il s'agit de découvrir et d'appliquer les principes qui gouvernent l'univers des sciences. Par contre, les exercices relevant d'une spécialité bien déterminée posent plus de problèmes aux élèves de notre pays, surtout lorsqu'ils impliquent des connaissances terminologiques poussées et la maîtrise de concepts bien différenciés. Or, on voit que cette tendance peut très bien s'expliquer par les caractéristiques et les priorités de l'enseignement des sciences en Suisse. Appliquée aux résultats d'autres pays, cette analyse permet d'identifier des groupes présentant des tendances contraires.

Profilo di rendimento e priorità nell'insegnamento. Un'analisi dei risultati ottenuti a livello svizzero con il test TIMSS per le scienze.

Riassunto

La Svizzera ottien, nell'ambito di un confronto internazionale del TIMSS, risultati chiaramente più deboli nelle scienze rispetto a quelli ottenuti in matematica. Questa differenza ha stimolato la seguente analisi a livello svizzero per quel che concerne i risultati del TIMSS nelle scienze. Si sono esaminati quali compiti erano specialmente difficili o specialmente facili per gli allievi svizzeri. Si è potuto dimostrare che in Svizzera i compiti con alta richiesta cognitiva sono risolti relativamente bene; per compiti cognitivamente alti significa che si trattava di scoprire e applicare principi scientifici. Al contrario, compiti nei quali si esaminano soprattutto contenuti disciplinari specifici, vengono risolti in Svizzera relativamente male, soprattutto quando richiedono l'impiego di una terminologia altamente specifica. Questo profilo di rendimento è congruente con le caratteristiche dell'insegnamento delle scienze in Svizzera. Sulla base di queste analisi dei compiti si capiscono meglio alcuni risultati parziali del TIMSS.

Achievement Profile and Priorities in Classes. An Analysis of the Swiss Achievement in the TIMSS Scientific Test.

Summary

In the international TIMSS comparison, Switzerland came off clearly worse in science than in mathematics. This discrepancy gave reason to the present analysis of the Swiss test results which examined what tasks were especially difficult or easy for Swiss students. Therefore, teachers had to assess the subject-related and terminological difficulty of the tasks. The results show that Swiss students were able to cope rather well with tasks demanding high cognitive requirements, tasks which are about discovering and applying scientific principles. Tasks testing special subject-related contents were solved rather badly, especially if they demanded high conceptual-terminological standards. This achievement profile corresponds to the characteristics and main focuses of the science classes in Switzerland. The present analysis facilitates a better understanding of the divergent test achievements of TIMSS.