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: Praktische Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenzen

Autor: Stebler, Rita / Reusser, Kurt / Ramseier, Erich

DOI: https://doi.org/10.5169/seals-786229

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: 11.07.2025

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

Praktische Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenzen

Erträge aus dem TIMSS-Experimentiertest[1]

Rita Stebler, Kurt Reusser & Erich Ramseier

Beim TIMSS-Experimentiertests wurden in 44 Deutschschweizer Klassen der siebten Jahrgangsstufe (N=396) mathematische und naturwissenschaftliche Handlungsprobleme zur Leistungsmessung eingesetzt. Die Jugendlichen mussten Experimente planen und durchführen, Ergebnisse protokollieren, Trends aus den Daten herauslesen, Schlussfolgerungen ziehen und diese anhand mathematischer und naturwissenschaftlicher Fachbegriffe erklären. Im internationalen Vergleich haben die Deutschschweizer Siebtklässler ausgezeichnet abgeschnitten. Die Leistungen der Mädchen und der Knaben unterscheiden sich nicht. Klassenübergreifende Leistungsvergleiche mit Experimentieraufgaben sind aufgrund ungelöster, messtheoretischer Probleme zur Zeit noch nicht vorbehaltlos zu empfehlen. Der Einsatz praktischer Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenzen beim individuellen Problemlösen und in interaktiven Lehr-Lern-Umgebungen entspricht den Leitzielen unserer Lehrpläne.

Der TIMSS-Experimentiertest - auf dem Weg zu einer neuen Prüfungskultur

TIMSS ist das Kürzel für *Third International Mathematics and Science Study*. In dieser auf der Sekundarstufe I und II[2] durchgeführten Untersuchung wurden weltweit die Mathematik- und Naturwissenschaftsleistungen sowie deren Rahmenbedingungen verglichen. Auf der Sekundarstufe I bestand die TIMSS-Erhebung aus einem schriftlichen Leistungstest, an dem sich 41 Länder beteiligten, und einem Experimentiertest, der in 21 Ländern[3] durchgeführt wurde.

Experimentiertest ist die freie Uebersetzung des englischen Fachbegriffs *Performance Assessment*. Er wurde in TIMSS für jenen Testteil verwendet, bei dem die Schülerinnen und Schüler durch praktisches Handeln, Beobachten und Nachdenken anspruchsvolle mathematische und naturwissenschaftliche Probleme lösen mussten[4]. Durch den Experimentiertest soll der vorwiegend auf Mehrfach- (multiple choice) und Kurzantworten beruhende schriftliche TIMSS-Leistungstest ergänzt und seine Aussagekraft mit Bezug auf das eigenständige Lösen praktischer Anwendungsprobleme erhöht werden (Robitaille & Garden, 1996).

Verstehen und Problemlösen - die Leitideen einer neuen, authentischeren Prüfungskultur

Im Zuge von TIMSS wurden in unseren Schulen erstmals in grösserem Stil Experimentieraufgaben zur Leistungsmessung eingesetzt. Dieses aktuelle, *verstehens- und anwendungsorientierte Prüfungsformat* stammt aus den USA. Dort besteht gegenwärtig ein reges öffentliches und wissenschaftliches Interesse an neuen Formen der Leistungsmessung. Es entspringt einem tiefen Unbehagen gegenüber der eigenen, langjährigen Testpraxis. In den USA unterziehen sich jedes Jahr viele Schülerinnen und Schüler obligatorischen Schulleistungstests. Dabei werden hauptsächlich Aufgaben mit Auswahl- (*multiple choice*) oder Kurzantworten eingesetzt. Mit solchen Aufgaben können Schulleistungen zwar ökonomisch erfasst werden. Es besteht jedoch die Gefahr, dass sie vorwiegend jene Inhalte abdecken, die sich leicht in solche Testfragen umformen, unter Zeitdruck abrufen und zweifelsfrei beurteilen lassen (Shepard, 1992). Obschon man grundsätzlich auch mit Aufgaben im multiple-choice Format gründliches Verstehen und komplexes Problemlösen erfassen kann (Wolf, 1994), zielen die meisten Fragen in standardisierten Schulleistungstests auf isoliertes Faktenwissen und Routineverfahren (Resnick & Resnick, 1992). Sie sind einem behavioristischen Lernbegriff verpflichtet (Gage & Berliner, 1992) und vermitteln den Schülerinnen und Schülerin insofern ein falsches Bild von erwünschtem Lern- und Problemlöseverhalten, als sie den Eindruck erwecken, es gäbe pro Frage nur eine richtige Antwort, die man auswendig lernen und bei Bedarf abrufen müsse. Die Grenzen herkömmlicher Schulleistungstests zeigen sich auch darin, dass sie nur ansatzweise erfassen können, ob der Testnehmer einen Argumentationsgang folgerichtig aufbauen, einen komplexen Sachverhalt versprachlichen oder ein Handlungsproblem eigenständig lösen kann.

Lange Zeit wurde nicht erkannt, dass solche *Prüfungsaufgaben unerwünschte Wirkungen auf die schulischen Lehr-Lern-Kulturen* haben können. Um bei den staatlichen Leistungsvergleichen gut abzuschneiden und Lohneinbussen zu vermeiden, richten viele Lehrpersonen ihren Unterricht inhaltlich und didaktisch auf die dort eingesetzten Aufgaben aus (Silver, 1992). Sie vermitteln ihren Schülerinnen und Schülern ein umfassendes Faktenwissen, das sich im Schulleistungstest zwar auszahlt, in Alltag und Beruf aber wenig taugt. Gründliches Verstehen, vernetztes Denken, kreatives Problemlösen und kompetente Selbst-

steuerung, im englischen Sprachraum als 'higher order thinking'(Baker, 1990)[5] oder 'mindfulness' (Langer, 1989) bezeichnet, bleiben in testorientierten Lehr-Lern-Kulturen häufig auf der Strecke (Reusser & Stebler, 1997; Shavelson & Baxter, 1992). Damit sind grundlegende Bildungsziele gefährdet. Andere Lehrpersonen leisten passiven Widerstand, indem sie den Test nicht ordnungsgemäss durchführen, am Testtag die leistungsschwachen Kinder beurlauben oder die Schülerantworten eigenhändig aufbessern. Für die spezifische Testvorbereitung und die Durchführung werden im Schnitt 100 Unterrichtsstunden eingesetzt (Smith, 1991). Das ist etwa ein Zehntel der jährlichen Unterrichtszeit. Unmut herrscht in den USA auch über den Umgang der Tagespresse mit den Ergebnissen von Schulleistungstests. Es werden ohne ausreichende Erklärungen Rangreihen der Schulen und Klassen veröffentlicht, die nicht selten dazu dienen, die Lehrpersonen zu verunglimpfen und Reformversuche abzuwürgen (Gage & Berliner, 1992)[6].

«Creating tests worth taking» (Wiggins, 1992) ist das Leitmotiv der Gegenbewegung: *Eine gute Testaufgabe ist eine komplexe Anwendungsaufgabe* (Shepard, 1989) wäre die dazu passende lern- und denkpsychologische Präzisierung. Sie ergibt sich aus der aktuellen Auffassung von multiplen Intelligenzen (Gardner, 1994) und entspricht einem erweiterten kognitionspsychologischen Lernbegriff. Ihm zufolge ist Lernen nicht nur eine aktive, konstruktive, kumulative und zielbezogene, sondern auch eine situierte, selbstregulierte und interaktive Tätigkeit (Stebler, Reusser, & Pauli, 1994).

Die Konzeption von Lernen, Denken und Problemlösen als *situierte* Tätigkeit (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991) geht wesentlich auf die Ergebnisse der Expertiseforschung (Chi, Feltovich, & Glaser, 1981) und die ethnologischen Untersuchungen zur Alltags- und Strassenmathematik zurück (Nuñes, Schliemann, & Carraher, 1993). Sie haben gezeigt, dass der Erwerb und Gebrauch von Sach- und Problemlösewissen stärker an bestimmte Kontexte und Bereiche gebunden ist, als in einem auf die direkte Vermittlung eines systematisierten, abstrakt-begrifflichen Fachwissens ausgerichteten Schulunterricht gewöhnlich angenommen wird. Abstrakt vermitteltes Schulwissen kann in Anwendungssituationen oft nicht spontan abgerufen werden. In einem Kontext oder Fachbereich erworbene Kenntnisse und Fertigkeiten lassen sich ohne besondere Vorkehrungen nur bedingt auf andere Kontexte oder Fächer übertragen. Diese Erkenntnisse haben zu den didaktischen Empfehlungen geführt, Wissen und Problemlöseverfahren vorzugsweise in konkreten Situationen zu vermitteln und dann sukzessive von kontextuellen Bindungen zu befreien (abstrahiertes Wissen; Adams, 1989) oder die Vermittlung abstrakten Wissens mit Anwendungsbeispielen zu kombinieren (Anderson, Reder, & Simon, 1996).

Dass Lernen und Problemlösen insofern *selbstgesteuerte* Tätigkeiten sind, als sie von der Schülerin oder dem Schüler eigenständig in Gang gesetzt, überwacht, den Erfordernissen der Situation laufend angepasst, beurteilt und verantwortet werden, wurde bereits in der Problemlöse- und Denkpsychologie nach der Jahrhundertwende erörtert (Dewey, 1933/1960; Kerschensteiner, 1926). Beim Verebben des Behaviorismus bzw. als Folge der kognitiven Wende in der Psychologie wurde dieser Faden von der Metakognitionsforschung wieder auf-

genommen (Flavell, 1983). Die von Dewey in der Analyse des Denkprozesses beschriebenen Phasen *analysieren, vermuten, prüfen* und *verifizieren* erscheinen bei Brown (1983) als Kontrollstrategien, werden von Sternberg (1992) als Metakomponenten der Intelligenz behandelt und gelten als Grundlagen des wissenschaftlichen Arbeitens (Ruiz-Primo & Shavelson, 1996).

Experimentieraufgaben lösen - formale und materiale Kompetenzen orchestrieren

Durch den Einbezug anspruchsvoller Anwendungsaufgaben in die Schulleistungstests will man in den USA zum einen Verstehens- und *Problemlöseleistungen kontextualisiert* und umfassender *messen* (Baxter & Shavelson, 1994), zum anderen die *Unterrichtsgestaltung* und die Lern- und Denkprozesse der Schülerinnen und Schüler stärker *in Richtung 'higher order thinking' lenken* (Resnick & Resnick, 1992; Shepard, 1992). Im Hinblick auf diese Ziele werden u.a. authentische, lebensweltliche, bedeutsame, interessante, herausfordernde und anspruchsvolle Experimentieraufgaben empfohlen (Shavelson & Baxter, 1992). Man hat dabei komplexe, in fachlichen Kontexten verankerte Handlungsprobleme vor Augen, die relevante Bildungsinhalte verkörpern, mehrere Lösungen zulassen und Lern- und Problemlöseaktivitäten auslösen, die sich mit den in Abbildung 1 dargestellten Attributen charakterisieren lassen.

Abbildung 1: Experimentieraufgaben sollen authentisches Problemlösen messen und fördern. Die Abbildung zeigt einige, in der Fachliteratur aufgeführte Facetten dieses schillernden Begriffs.

Je nach Experimentieraufgabe müssen die Schülerinnen und Schüler ein greifbares Produkt herstellen oder Objekte bezüglich bestimmter Merkmale vergleichen, die Bestandteile eines Ganzen identifizieren, Objekte klassifizieren, einen Vorgang über längere Zeit systematisch beobachten und protokollieren sowie durch logisches Denken und vernünftiges Urteilen begründete Schlussfolgerungen ziehen (Shavelson, et al. 1997). Es wird erwartet, dass sie sich wie *Forscher* (Shavelson, 1996) und *Fachexperten* (Baxter, Elder, & Glaser, 1996) verhalten, die Hypothesen generieren, planen, prüfen, überwachen, Daten protokollieren, beurteilen und die Ergebnisse anhand ihres Fachwissens erklären.

Im Kontext von TIMSS werden diese und weitere aufgabenübergreifende Tätigkeiten als *Leistungserwartungen* (performance expectations) bezeichnet. Die Funktion dieses messtheoretischen Konstruktes «is to describe, in a non-hierarchical scheme, the many kinds of performances or behaviours that a given test item or block of content might be expected to elicit from students' (Robitaille, et al., 1993, p. 44). Diese Umschreibung verweist auf die Tatsache, dass man kognitive Kompetenzen nicht direkt, sondern nur über Verhaltensäusserungen erfassen kann. Wenn man also Denken und Problemlösen messen will, muss man als erstes jenes Verhalten beschreiben, von dem man annimmt, dass es Rückschlüsse auf die zu erfassenden Kompetenzen zulässt. Als nächstes muss man Testaufgaben entwickeln, die das gewünschte Verhalten auslösen. Schliesslich müssen Experten beurteilen, wie sehr die Aufgaben und die durch sie provozierten Verhaltensäusserungen inhaltlich mit dem vorher beschriebenen Verhalten übereinstimmen (Ingenkamp, 1985; Gage & Berliner, 1992).

Auf diese Weise haben Experten bei TIMSS vorwiegend induktiv ein auf die aufgabenübergreifenden, *formalen Kompetenzen bezogenes Kategoriensystem* entwickelt. Es wurde in allen Testteilen angewendet (siehe Ramseier, 1997). Beim Experimentiertest wurde für jede Teilaufgabe eines komplexen Handlungsproblems vorhergesagt, welches Verhalten sie auslösen und damit welche formale Kompetenz sie messen soll. Wenn man die Teilaufgaben pro Kategorie der Leistungserwartungen aufsummiert, werden ergänzend zu den Fachleistungen auch aufgabenübergreifende Aussagen in bezug auf die Kompetenzen der Jugendlichen beim Problemlösen, beim Erforschen und beim Anwenden von Routineverfahren möglich.

Nach diesen Ausführungen können wir den einleitend genannten Zweck des TIMSS-Experimentiertests dahingehend präzisieren, dass ausgehend vom aktuellen kognitionspsychologischen Lernbegriff, das Zusammenspiel formaler und materialer Kompetenzen beim Lösen kontextualisierter praktischer Anwendungsprobleme erfasst und so ein anderer Aspekt der Schulleistungen gemessen werden soll als beim schriftlichen TIMSS-Leistungstest (Moser, et al., 1997).

In der Fortsetzung interessieren die Leistungen der Deutschschweizer Siebtklässler im internationalen Vergleich mit Bezug auf die vom Experimentiertest erfassten (a) materialen und (b) formalen Aspekte des handelnden Problemlösens. Anschliessend (c) wird am Beispiel der Experimentieraufgabe 'Tablette' aufgezeigt, wie sich die Lösungen von Experimentieraufgaben zur differenzierten Bestimmung des Lernstandes nutzen lassen.

Methode

Stichprobe

Der TIMSS-Experimentiertest wurde in 21 Ländern bei 14-Jährigen durchgeführt. In der Deutschschweiz sind Jugendliche dieses Alters im siebten Schuljahr, in den meisten anderen Ländern im achten oder neunten. Für die Deutschschweiz wurde eine repräsentative Stichprobe von 44 Klassen (12 Realklassen, 25 Sekundarklassen und 7 Gymnasialklassen) gebildet. Die Klassen wurden zufällig aus der Stichprobe des schriftlichen Leistungstests von TIMSS ausgewählt. Pro Klasse wurden ebenfalls nach Zufallsprinzip 9 Schülerinnen und Schüler einbezogen (N=396; Knaben 49%).

Materialien und Durchführung

Der Experimentiertest besteht aus sechs mathematischen und sechs naturwissenschaftlichen Problemstellungen mit insgesamt 65 Teilaufgaben (vgl. Abb. 2, S. 34) Er ist als Werkstatt bzw. Circuit konzipiert. Die Schülerinnen und Schüler besuchten nach Rotationsplan drei Posten, an denen sie drei bis fünf kleine Experimente planten, durchführten und ihr Vorgehen und ihre Ergebnisse in einem Testheft protokollierten. Je nach Komplexität der Experimentieraufgabe standen 15 oder 30 Minuten Bearbeitungszeit zur Verfügung. Jede der 12 Experimentieraufgaben wurde pro Klasse drei Jugendlichen vorgelegt. Dies ergab ein Total von 132 Schülerlösungen (44 mal 3) pro Experimentieraufgabe.

Der Test wurde von trainierten Studierenden mit Unterrichtserfahrung durchgeführt. Sie reisten zu den Klassen, stellten die Werkstatt auf, überwachten die Jugendlichen beim Experimentieren und brachten die ausgefüllten Testhefte zurück. Pro Klasse wurde ein Halbtag eingesetzt. Die reine Testdauer betrug 120 Minuten.

Kodierung

Die Kodierung der Schülerantworten erfolgte nach dem inhaltsanalytischen Kategoriensystem von TIMSS. Es wurde ein aus zwei Ziffern bestehender synthetischer Code verwendet. Durch die erste Ziffer wurde die Qualität (z.B. ansatzweise, teilweise oder ganz richtig), durch die zweite die Spezifität (z.B. Lösung mit Fachbegriff, Lösung mit Alltagswissen) der Schülerantwort erfasst. Neun Studierende übernahmen die Kodierarbeit. Pro Experimentieraufgabe wurden 10% der Testhefte gegenkodiert. Bei der Qualität der Antworten stimmten die Kodierer in 96% ihrer Urteile exakt überein, bei der Spezifität der Antworten in 92% ihrer Urteile (Harmon, et al., 1997, A-24).

Mathematik	Inhalte	Leistungserwartungen
Würfel (6 Teilaufg.) Finde heraus, was passiert, wenn Du nach einer bestimmten Regel die gewürfelten Zahlen veränderst.	Ganze Zahlen Wahrscheinlichkeit	Routineverfahren anwenden Muster erkennen
Taschenrechner (7 Teilaufg.) Verwende den Taschenrechner, um ein Zahlenmuster zu erforschen und die fehlenden Zahlen herauszufinden.	Ganze Zahlen Bedeutung der Operationen	Routineverfahren anwenden Strategie entwickeln und beschreiben Vorhersagen
Falten und Schneiden (4 Teilaufg.) Falte und schneide Papierblätter, um jene Formen herzustellen, die den abgebildeten Mustern entsprechen.	Geometrie Symmetrie	Problemlösen Vorhersagen
Möbeltransport (8 Teilaufg.) Finde heraus, wie gross Möbel sein dürfen, damit sie im Korridor (Gang) um die Ecke transportiert werden können.	Geometrie Proportionalität	Routineverfahren Problemlösen
Verpackungen (3 Teilaufg.) Entwirf verschiedene Schachteln, in die genau 4 Tischtennisbälle passen.	Geometrie Flächen und Körper	Routineverfahren Problemlösen
Plastilin (8 Teilaufg.) Verwende die Waage, um Plastilinklumpen von unterschiedlichem Gewicht herzustellen. Wäge so genau wie möglich.	Masse & Einheiten Proportionalität	Routineverfahren Problemlösen Strategie entwickeln und beschreiben
Naturwissenschaften	**Inhalte**	**Leistungserwartungen**
Puls (4 Teilaufg.) Finde heraus, wie sich Dein Puls verändert, wenn Du 5 Minuten lang die Stufe auf- und absteigst.	Biologie Blutkreislauf	Untersuchung durchführen Daten sammeln, ordnen, darstellen Daten interpretieren Fachbegriffe anwenden
Magnete (2 Teilaufg.) Verwende die Gegenstände aus dem Sack, um herauszufinden, ob Magnet A oder B stärker ist.	Physik Magnetismus	Untersuchung durchführen Daten interpretieren Schlussfolgern Fachbegriffe anwenden
Batterien (4 Teilaufg.) Finde heraus, welche Batterien geladen und welche entladen (leer) sind.	Physik Elektrizität	Untersuchung durchführen Daten interpretieren Schlussfolgern Fachbegriffe anwenden
Gummiband (7 Teilaufg.) Finde heraus, wie sich die Länge des Gummibandes verändert, wenn mehr und mehr Unterlegscheiben angehängt werden.	Physik Elastizität	Untersuchung durchführen Daten sammeln, ordnen, darstellen Daten interpretieren Fachbegriffe anwenden
Tabletten (6 Teilaufg.) Untersuche, wie schnell sich Tabletten bei unterschiedlicher Wassertemperatur auflösen.	Physik Kinetik	Untersuchung planen und durchführen, Material brauchen Daten sammeln, ordnen, darstellen Schlussfolgern, Fachbegriffe anw.
Schatten (6 Teilaufg.) Finde heraus, wie sich die Grösse des Schattens verändert, wenn Du die Karte bewegst.	Physik Licht	Untersuchung planen und durchführen, Material brauchen Daten sammeln, ordnen, darstellen Schlussfolgern, Fachbegriffe anw.

Abbildung 2: TIMSS-Experimentieraufgaben - Inhalte und Leistungserwartungen.

Ergebnisse

Die hier dargestellten Ergebnisse beruhen auf den Punktwerten für die Lösungsqualität. Aufgeführt wird mehrheitlich die durchschnittliche Lösungsqualität in Prozent der maximal möglichen Punktzahl pro Fach, Experimentieraufgabe oder Leistungserwartung. Ausgewählte Angaben zur Spezifität der Schülerantworten werden in die Beschreibung der Ergebnisse eingeflochten.

Fachleistungen

Land	Schuljahr	Alter	**Mathematik**		**Naturwissenschaften**	
Singapur	8	14.5	70	(1.7)	72	(1.8)
Schweiz	**7**	**14.1**	**66**	**(1.5)**	**65**	**(1.0)**
Schweden	7	13.9	65	(1.3)	63	(1.5)
Schottland	9	13.7	61	(2.2)	64	(1.5)
Norwegen	7	13.9	65	(1.1)	58	(0.8)
Tschechien	8	14.4	62	(1.7)	60	(1.3)
Kanada	8	14.1	62	(1.3)	59	(1.3)
Neuseeland	8.5 - 9.5	14.0	62	(1.3)	58	(1.5)
Spanien	8	14.7	52	(1.1)	56	(1.0)
Iran	8	14.6	54	(1.7)	50	(2.8)
Portugal	8	14.6	48	(1.3)	47	(1.2)
Zypern	8	13.8	44	(1.2)	49	(1.0)
Australien	8 oder 9	14.3	66	(1.5)	63	(1.1)
England	9	14.0	64	(1.0)	71	(0.9)
Niederlande	8	14.3	62	(1.5)	58	(1.4)
USA	8	14.2	54	(1.4)	55	(1.4)
Kolumbien	8	15.8	37	(2.5)	42	(1.4)
Rumänien	8	14.6	66	(2.0)	57	(2.0)
Slowenien	8	14.7	64	(1.0)	58	(1.4)
Internationaler Durchschnitt			**59**	**(0.3)**	**58**	**(0.4)**

Tabelle 1: TIMSS-Experimentiertest: Prozentsatz der richtigen Lösungen in Mathematik und Naturwissenschaften.

Anmerkungen. Quelle: IEA Third International Mathematics and Science Study (TIMSS), 1994-95. Die Standardfehler sind in Klammern aufgeführt. Die Länder unter dem Doppelstrich haben bei der Auswahl der Schüler die Vorgaben von TIMSS nicht erfüllt. Ihre Stichprobe ist kein genaues Abbild der Schülerschaft des Landes. Hongkong und Israel wurden wegen ungeeigneter Stichproben aus dem internationalen Vergleich ausgeschieden.

Wie Tabelle 1 zeigt, haben die Deutschschweizer Siebtklässler im TIMSS-Experimentiertest ausgezeichnet abgeschnitten. In Mathematik erzielten nur die Achtklässler aus Singapur deutlich höhere Testwerte. In den Naturwissenschaften trifft das für die Jugendlichen aus Singapur und England zu.

Besonders erfreulich ist, dass die Deutschschweizer Jugendlichen dieses Spitzenresultat erreichten, obschon sie zum Testzeitpunkt ein Schuljahr weniger hinter sich hatten als Gleichaltrige anderer Länder. Es ist sehr wichtig zu beachten, dass im Experimentiertest weltweit die 14-Jährigen getestet wurden. Wegen des vergleichsweise späten Einschulungstermins sind Jugendliche dieses Alters in der Deutschschweiz im siebten Schuljahr, in den meisten anderen Ländern im achten oder neunten. Die Testwerte vieler Länder unterscheiden sich nicht überzufällig. Die Bildung von Rangreihen ist daher zu unterlassen. Die Mädchen und die Knaben erbrachten in allen Ländern einschliesslich der Deutschschweiz vergleichbare Leistungen. Bei einzelnen der 12 Experimentieraufgaben zeigten sich in wenigen Ländern - nicht aber in der Deutschschweiz - geschlechtsspezifische Leistungsunterschiede ($p < .05$).

Die Abbildung 3 (siehe S. 37) zeigt die Leistungen der Deutschschweizer Siebtklässler bei *den einzelnen Experimentieraufgaben.* Zum Vergleich ist auch der Prozentsatz richtiger Lösungen der Achtklässler aus Singapur und der internationale Durchschnitt aufgeführt. Die Leistungen unserer Schülerinnen und Schüler liegen bei 11 der 12 Experimentieraufgaben über dem internationalen Durchschnitt. Mit Ausnahme der Experimentieraufgabe 'Verpackungen' sind die Unterschiede signifikant ($p < .05$). Die positivste Abweichung besteht bei der Problemstellung *Plastilin*. Hier haben die Deutschschweizer Jugendlichen sowohl beim Herstellen der vier Plastilinklumpen (Wägen im Grammbereich) als auch beim Beschreiben ihres Vorgehens sehr gute Leistungen erzielt. Der hohe Prozentwert bei der Problemstellung *Falten/Schneiden* ist damit zu erklären, dass das Anfertigen von Scherenschnitte bei uns eine gut geübte Tätigkeit ist.

Bei der Problemstellung *Möbeltransport* sind der schweizerische und der internationale Prozentwert identisch. Ursache dafür sind die vielen Falschlösungen bei jenen zwei der acht Teilaufgaben, bei denen die Schülerinnen und Schüler Möbelmodelle und Möbelmasse realen Möbelstücken zuordnen mussten. Ob unsere Jugendlichen sich nur mit Möbelmassen nicht auskennen oder generell eine vage Vorstellung von Längenmassen haben, kann anhand der vorliegenden Daten nicht beurteilt werden.

Dass der Prozentwert der Experimentieraufgabe *Verpackungen* nicht deutlicher über dem internationalen Durchschnitt liegt, ist wesentlich darauf zurückzuführen, dass viele der Deutschschweizer Siebtklässler bei der dritten Teilaufgabe eine Schachtel anstatt des verlangten Faltplanes hergestellt haben. Möglicherweise haben sie die Anleitung ungenau gelesen, sich vom Material zum Basteln verleiten lassen oder einfach geprüft, ob ihr Faltplan einer vier Bälle fassenden Schachtel entspricht.

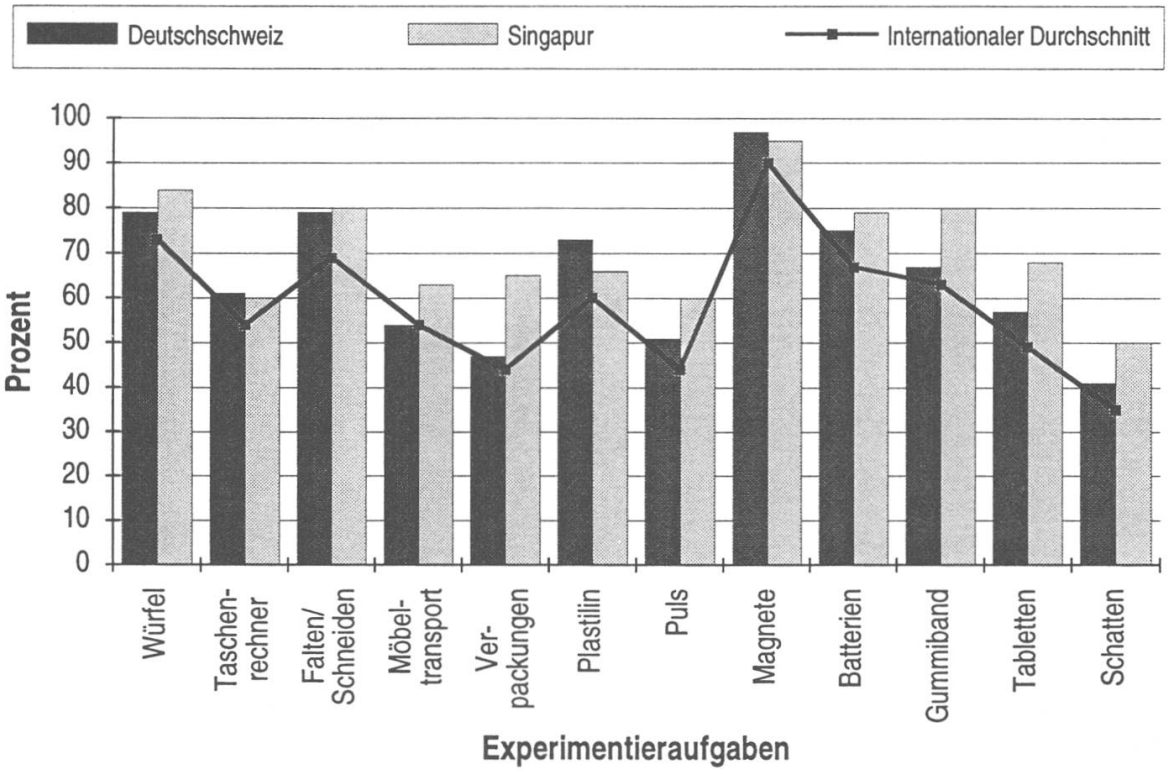

Abbildung 3: Prozentsatz der richtigen Lösungen pro Aufgabe.

Wie der Ländervergleich zeigt, sind die Leistungsunterschiede zwischen den Jugendlichen aus der Deutschschweiz und Singapur bei den Experimentieraufgaben Verpackungen, Gummiband und Tabletten am grössten. Bei der Problemstellung *Gummiband* konnten nur wenige unserer Schülerinnen und Schüler ihre Messergebnisse in ein Balken- oder Säulendiagramm umsetzen. Bei der Experimentieraufgabe *Tabletten* fehlte ihnen der Fachbegriff 'Kinetik'.

Die Auswertung nach *Schultypen* zeigt, dass die Gymnasiasten und die Sekundarschüler bei 11 der 12 Experimentieraufgaben signifikant bessere Leistungen erbrachten als die Realschüler. Obschon die Testwerte der Gymnasiasten durchwegs etwas höher sind als jene der Sekundarschüler, ist der Unterschied zwischen Gymnasiasten und Sekundarschülern nur bei der Problemstellung 'Gummiband' überzufällig.

Formale Kompetenzen

Um die Testleistungen der Jugendlichen hinsichtlich der aufgabenübergreifenden Kompetenzen zu vergleichen, ordnete TIMSS jede Teilaufgabe einer Kategorie der Leistungserwartungen zu. Für die folgenden Auswertungen wurden die Teilaufgaben ausgewählter Kategorien zu fünf Oberkategorien zusammengezogen (Tab. 2, S. 38).

In *Mathematik* wurde zwischen Routineverfahren (13 Teilaufgaben) und Problemlösen (21 Teilaufgaben) unterschieden. Alle Länder einschliesslich der Deutschschweiz erzielten bei den Routineverfahren höhere Testwerte als beim

Land	Mathematik				Naturwissenschaften					
	Routine-verfahren		Problem lösen		Routine verfahren		Problemlösen/ Fachwissen		Erforschen	
Singapur	80	1.3)	62	(2.3)	75	(1.8)	59	(3.0)	74	(1.9)
Schweiz	**76**	**(1.8)**	**60**	**(1.8)**	**63**	**(1.4)**	**55**	**(1.6)**	**70**	**(1.3)**
Schweden	73	(1.3)	60	(1.6)	59	(1.9)	56	(2.3)	67	(1.5)
Schottland	75	(1.7)	52	(2.3)	69	(1.8)	48	(2.1)	65	(1.5)
Norwegen	75	(1.2)	58	(1.3)	57	(1.2)	48	(1.6)	63	(1.1)
Tschechien	73	(1.6)	56	(1.7)	57	(2.0)	53	(2.2)	65	(1.6)
Kanada	74	(1.4)	54	(1.3)	64	(2.2)	50	(1.6)	60	(1.4)
Neuseeland	72	(1.1)	55	(1.6)	65	(2.1)	47	(1.6)	57	(1.6)
Spanien	66	(1.4)	46	(1.3)	45	(1.8)	39	(1.6)	57	(1.2)
Iran	61	(1.8)	49	(1.8)	53	(3.4)	61	(2.0)	56	(2.7)
Portugal	66	(1.2)	36	(1.6)	47	(1.4)	32	(1.8)	45	(1.4)
Zypern	58	(1.3)	38	(1.4)	48	(1.7)	37	(1.9)	50	(1.1)
Australien	75	(1.4)	61	(1.9)	67	(1.9)	54	(2.0)	66	(1.1)
England	77	(1.1)	54	(1.3)	77	(1.4)	49	(2.0)	73	(1.0)
Niederlande	77	(1.7)	54	(1.5)	63	(1.7)	39	(1.9)	57	(1.4)
USA	64	(1.6)	49	(1.4)	61	(2.2)	43	(1.5)	55	(1.4)
Kolumbien	49	(2.7)	30	(2.7)	35	(2.4)	32	(2.2)	41	(1.5)
Rumänien	74	(1.9)	60	(2.4)	53	(2.5)	48	(3.3)	61	(2.2
Slowenien	72	(1.2)	57	(1.1)	60	(1.3)	48	(1.5)	59	(1.3)
Internationaler Durchschnitt	**70**	**(0.4)**	**52**	**(0.4)**	59	(0.4)	47	(0.5)	60	(0.4)

Tabelle 2: Prozentsatz der richtigen Lösungen nach Fach und Leistungserwartung.

Anmerkungen. Quelle: IEA Third International Mathematics and Science Study (TIMSS), 1994-95. Die Standardfehler sind in Klammern dargestellt. Die Länder unter dem Doppelstrich haben bei der Auswahl der Schüler die Vorgaben von TIMSS verletzt. Ihre Stichprobe ist kein genaues Abbild der Schülerschaft des Landes. Mathematik: Routineverfahren (13 Teilaufg.), Problemlösen (21 Teilaufg.); Naturwissenschaften: Routineverfahren (7 Teilaufg.), Problemlösen/Fachwissen (16 Teilaufg.), Erforschen (12 Teilaufg.).

Problemlösen. Der Unterschied zwischen den Jugendlichen aus Singapur und der Deutschschweiz ist bei den Routineverfahren etwas grösser als beim Problemlösen. Beim Problemlösen liegen die Testwerte der Schülerinnen und Schüler aus der Deutschschweiz nicht nur näher bei Singapur, sondern auch deutlicher über dem internationalen Durchschnitt als bei den Routineverfahren.

In den *Naturwissenschaften* wurden drei Teilbereiche unterschieden: Erforschen (16 Teilaufgaben), Routineverfahren (7 Teilaufgaben) sowie Problemlösen und Anwenden von Fachwissen (12 Items). Die höchsten Werte hatten die Deutschschweizer Siebtklässler beim Erforschen, die zweithöchsten bei den Routineverfahren und die tiefsten beim Problemlösen und Anwenden von Fachwissen. Auch hier ist der Unterschied zu den Achtklässlern aus Singapur bei den Routineverfahren am grössten. Und wiederum zeigen sich beim Problemlösen, aber auch beim Erforschen positivere Abweichungen bezogen auf den internationalen Durchschnitt als bei den Routineverfahren.

Aus diesen Ergebnissen lässt sich vorsichtig schliessen, dass die Achtklässler aus Singapur die Deutschschweizer Siebtklässler im Experimentiertest unter anderem deshalb überrundet haben, weil sie ein Schuljahr länger Zeit hatten, sich Routineverfahren anzueignen. Umgekehrt kann man sagen, dass die Deutschschweizer Siebtklässler beim Problemlösen und beim Erforschen trotz des fehlenden Schuljahres und des etwas tieferen Durchschnittsalters mit dem Spitzenreiter Singapur gut mithalten können. Weiter wird deutlich, dass sich die Verstehens- und Problemlöseorientierung unserer Lehrpläne im internationalen Leistungsvergleich auszahlt.

Der Lernstand bei der Experimentieraufgabe 'Tabletten'

Welche Testergebnisse primär interessieren, hängt vom Zweck der Leistungsmessung ab. Während bei schul- und länderübergreifenden Leistungsvergleichen die soziale Bezugsnorm und die Durchschnittswerte im Vordergund stehen, sind zur Bestimmung des Lernstandes im Hinblick auf eine adaptive Unterrichtsgestaltung die sachbezogene Bezugsnorm und die Häufigkeiten aufschlussreicher. Am Beispiel der Tablettenaufgabe (Abb. 4, S. 40) wollen wir einen differenzierten Eindruck über den Aufbau und die Kodierung der TIMSS-Experimentieraufgaben vermitteln.

Die Experimentieraufgabe Tablette stammt aus dem Physikunterricht (Abb. 2). Sie ist in fünf Teilaufgaben gegliedert[7]. Die Schülerinnen und Schüler erhalten kaltes und heisses Wasser, mehrere Becher, sechs Brausetabletten, einen Thermometer und eine Uhr mit Sekundenzeiger. Sie müssen ein naturwissenschaftliches Experiment durchführen, um begründete Aussagen über die Wirkung der Wassertemperatur auf die Geschwindigkeit zu machen, mit der sich Tabletten auflösen.

TABLETTEN

An dieser Station brauchst Du:

Heisses und kaltes Wasser
Mehrere Becher
Einige Tabletten
Einen Löffel
Eine Uhr mit einem Sekundenzeiger
Ein Thermometer
Ein 30 cm-Lineal

Lies ALLE Anweisungen genau durch!

Deine Aufgabe:
Untersuche, wie schnell sich Tabletten bei unterschiedlichen Temperaturen auflösen.

Folgendes sollst Du tun:

- Plane eine Untersuchung, um herauszufinden, welche Wirkungen unterschiedliche Temperaturen auf die Geschwindigkeit haben, mit der sich Tabletten auflösen.

1. Schreibe Deinen Plan hier auf. Dein Plan sollte enthalten,
   - was Du messen willst.
   - wie viele Messungen Du machen willst.
   - wie Du Deine Messungen in einer Tabelle darstellen willst.

Wie schnell sich Tabletten bei unterschiedlicher Temperatur auflösen.
Zwischen 4-6 Messungen will ich machen.
Ich werde es in einer Tabelle, wo jeweils die Testergebnisse untereinander dargestellt ist zeigen.

2. Führe Deine Tests mit den Tabletten durch. Zeichne eine Tabelle und schreibe alle Deine Messungen auf.

	Temperatur	*Zeit der Auflösung*	
	25C°	*1 min 3 sek*	
	29C°	*56 sek*	
	33C°	*51 sek*	
	36C°	*47 sek*	

3. Welche Wirkung haben aufgrund Deiner Untersuchung verschiedene Temperaturen auf die Geschwindigkeit, mit der sich Tabletten auflösen?

Um so heisser das Wasser ist desto schneller lösen sich die Tabletten auf

4. Warum denkst Du, haben unterschiedliche Temperaturen diese Wirkung? Schreibe Deine Er--klärung auf.

Beim warmen Wasser bewegen sich die kleinen Teilchen * schneller, damit kann sich alles besser auflösen und verteilen. Durch die ständige Bewegung der Moleküle, die viel Platz brauchen, wird die Tablette aufgelöst.
**** (Moleküle)***

5. Wenn Du Deinen Plan ändern musstest, beschreibe Deine Änderungen und begründe, warum Du sie gemacht hast. Wenn Du Deinen Plan nicht ändern musstest, schreibe «Keine Änderung».

Keine Änderung

Leere den Inhalt Deiner Becher in das Waschbecken, trockne die Becher ab und hinterlasse alles so, wie Du es vorgefunden hast!

Abbildung 4: Testheft der Experimentieraufgabe 'Tablette' mit Schülerlösung.
Anmerkungen. Das Originaltestheft bestand aus drei A4 Seiten.

1. Bei der ersten Teilaufgabe müssen die Schülerinnen und Schüler einen *Untersuchungsplan* aufstellen.

Entscheidend ist, dass sie (i) die Untersuchung beschreiben und (ii) angeben, welche Variablen sie messen und (iii) welche sie konstant halten. Rund zwei Drittel der Schülerinnen und Schüler erfüllten diese Kriterien ganz oder teilweise (Abb. 5). Teil- und Falschlösungen entstanden dadurch, dass die Probanden ihr geplantes Vorgehen ungenau beschrieben, nur eine der beiden relevanten Grössen (Wassertemperatur und Zeitdauer) nannten oder nicht angaben, dass sie mehr als einen Versuch vorsahen.

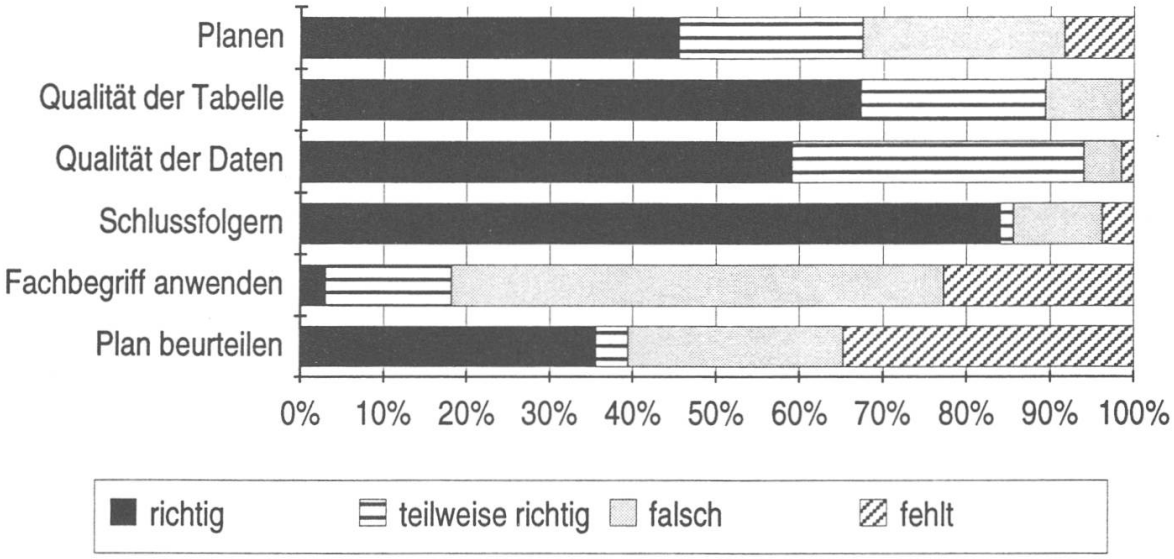

Abbildung 5: Tablettenaufgabe - Lösungsqualität der Teilaufgaben.

2. Anschliessend mussten die Schülerinnen und Schüler die *Untersuchung durchführen.*

Sie mussten eine Tabelle erstellen, die Wassertemperatur und die Auflösezeit der Tablette messen und die beiden Werte eintragen. Die Anzahl der Versuche war frei. Beurteilt wurden hier zwei Aspekte, die Qualität der Tabelle und die Qualität der Daten. Für die *Qualität der Tabelle* waren drei Kriterien massgebend. Die Tabelle musste (i) mindestens zwei Einträge enthalten. Bei jedem Eintrag mussten (ii) Wassertemperatur und Auflösezeit der Tablette gekoppelt sein. Die (iii) Kolonnen und Messdaten mussten beschriftet sein. Die Schülerinnen und Schüler erbrachten hier sehr gute Leistungen. Die Teil- oder Falschlösungen sind mehrheitlich darauf zurückzuführen, dass die Probanden entweder keine Tabelle erstellt oder die Kolonnentitel und Dimensionen weggelassen hatten. Hinsichtlich der *Qualität der Daten* mussten für eine vollständig richtige Lösung (i) mindestens drei Messwerte (ii) in plausibler Grösse ausgewiesen werden, die zeigen (iii), dass Tabletten sich mit steigender Wassertemperatur schneller auflösen. Auch diese Kriterien wurde von den meisten Schülerinnen und Schülern erreicht. Wer hier Punkte verlor, hatte in der Regel weniger als

drei Versuche protokolliert. Ob die betreffenden Probanden noch nicht wissen, dass Experimente mehrere Messungen erfordern, oder wie die Versuchsleiter annahmen, Hemmungen hatten, die Tabletten zu vergeuden, muss offen bleiben.

3. Welche *Wirkungen* haben aufgrund Deiner Untersuchung verschiedene Temperaturen auf die Geschwindigkeit, mit der sich Tabletten auflösen?

Mit der Teilfrage drei wurde die Kompetenz der Schülerinnen und Schüler geprüft, Muster in Daten zu erkennen und *Schlussfolgerungen* zu ziehen. Beurteilt wurde, (i) ob die Schlussfolgerung mit den protokollierten Daten übereinstimmt und (ii) ob die Beziehung zwischen der Wassertemperatur und der Auflösedauer der Tablette beschrieben ist. Diese Teilaufgabe wurde am besten gelöst. Vereinzelt fehlten die Schlussfolgerungen.

4. Bei der vierten Teilfrage mussten die Untersuchungsergebnisse anhand des *Fachbegriffs* 'Kinetik' erklärt werden.

Es wurde die Antwort erwartet, dass (i) mit steigender Wassertemperatur die kinetische Energie der Moleküle zunimmt und damit auch deren mittlere Geschwindigkeit, wodurch (ii) die Tablette heftiger attackiert und schneller zertrümmert wird. Unsere Siebtklässler waren mit dieser Frage überfordert, da der Fachbegriff 'Kinetik' in der Regel erst später unterrichtet wird. Die meisten wiederholten ihre Schlussfolgerung oder versuchten eine Erklärung aus ihrem Alltagswissen herzuleiten:

- *Wenn es z.B. schneit und es 30° warm wäre, würde der Schnee gleich wieder schmelzen. ... Auch beim Kochen ist es so ... Mit kaltem Wasser kann man nicht kochen.*
- *Weil Flüssiges aufweicht oder auflöst.*
- *Wenn man Geschirr mit kaltem Wasser wäscht, hat man mehr Mühe zum Trocknen als mit heissem Wasser.*
- *Ich denke, es ist wie bei der Schokolade, je tiefer die Temperatur, um so länger braucht auch sie, bis sie verlaufen ist. Die Kälte hat eine geringere Energie als die Wärme, sie braucht darum auch länger, bis etwas schmilzt oder vergeht.*
- *Wenn das Wasser wärmer ist, könnte es sein, dass sich der Sauerstoff, der sich in den Tabletten befindet, ausdehnt und deswegen die Tablette schneller aufgelöst wird.*
- *Ich glaube, heisses Wasser ist aggressiver gegenüber Tabletten.*
- *Im Wasser hat es ganz viele kleine 'Teilchen', die mit zunehmender Wärme wirrer um sich irren (Ausdehnung des Wassers), so dass sie die Tablette eigentlich auseinanderhäkeln.*

5. Bei der fünften Teilfrage mussten die Schülerinnen und Schüler ihren Untersuchungsplan *beurteilen*.

Kodiert wurde, ob die Antwort mit dem Plan übereinstimmte (die Aussage: keine Änderung wurde nur akzeptiert, wenn der Plan vollständig war) und (ii) ob allfällige Änderungen begründet wurden. Mit rund 40% ist der Anteil der richti-

gen Lösungen vergleichsweise tief, liegt aber über dem internationalen Durchschnitt. Fast alle richtigen Lösungen beruhen darauf, dass die Jugendlichen vollständige Pläne als richtig beurteilten. Erstaunlich ist, dass rund ein Viertel der Probanden ihren unvollständigen Plan während des Experimentierens verändert hatte, die Änderungen aber nicht berichtete. Ein weiteres Drittel gab kein Urteil ab.

Zusammenfassend erbrachten die Deutschschweizer Siebtklässler beim Planen und Durchführen der Untersuchung sowie beim Schlussfolgern mehrheitlich sehr gute Leistungen. Beim Fachbegriff 'Kinetik' und beim Beurteilen der Untersuchung besteht Förderbedarf.

Diskussion

Der Experimentiertest kam bei den Jugendlichen und den Lehrpersonen gleichermassen gut an. Für beide Seiten war es eine interessante und willkommene Erfahrung. Die zeitlichen Vorgaben waren realistisch und die Experimentieraufgaben und Testmaterialien der Schulstufe angemessen. Aufgrund der hohen organisatorischen Anforderungen erwies es sich als gute Entscheidung, pro Klasse zwei Testpersonen mit Unterrichtserfahrung einzusetzen. Die Ergebnisse des Experimentiertests zeigen, dass die Deutschschweizer Siebtklässler sowohl in der Mathematik als auch in den Naturwissenschaften an der internationalen Spitze mithalten können, und dies, obschon sie zum Testzeitpunkt ein Schuljahr weniger absolviert hatten als Gleichaltrige in den meisten anderen Ländern. Die Leistungen der Mädchen und der Knaben unterscheiden sich nicht. In der Mathematik stimmen die Ergebnisse des Experimentiertests gut mit jenen des schriftlichen TIMSS-Leistungstests (siehe Moser, et al., 1997) überein. In den Naturwissenschaften haben die Deutschschweizer Siebtklässler im Experimentiertest besser abgeschnitten als im schriftlichen Leistungstest (siehe Ramseier in diesem Heft).

Der TIMSS-Experimentiertest gemessen an acht Gütekriterien

Experimentiertests sind ein handlungsbezogenes Prüfungsformat, das als Ergänzung zu den herkömmlichen, standardisierten Schulleistungstests empfohlen wird. Standardisierte Tests müssen Schulleistungen ökonomisch, zuverlässig (reliabel) und so konstant messen, dass die Vergleichbarkeit der Ergebnisse über Jahre hinweg gewährleistet ist. Diese *Gütekriterien* gelten auch für Experimentiertests, sind aber weniger zentral als die Validität. Linn, Baker und Dunbar (1991) haben die klassischen Kriterien der Inhalts-, Vorhersage- und Konstruktgültigkeit für Experimentiertests auf acht Punkte erweitert. Wir werden diese Punkte mit Bezug auf den TIMSS-Experimentiertest und die Ergeb-

nisse der Deutschschweizer Siebtklässler erläutern und die Diskussion mit einigen Überlegungen zum Einsatz von Experimentieraufgaben im Unterricht abschliessen.

1. Die möglichen ***Wirkungen*** auf die Inhalte und die didaktische Gestaltung des Unterrichts müssen bedacht werden.

Dem Kriterium der Inhaltsgültigkeit entsprechend sind gute Experimentieraufgaben auf die in den Lehrplänen festgehaltenen formalen und materialen Ziele bezogen. Sie prüfen Inhalte, die in qualitativ hochstehendem, verstehensorientiertem Unterricht aufgebaut und geübt werden. Die Leistungsbeurteilung ist transparent, die Kriterien für die Beurteilung der Lösungen werden offengelegt.

Die TIMSS-Experimentieraufgaben sind mit Ausnahme der noch nicht unterrichteten Fachbegriffe gut auf unsere *Lehrpläne* bezogen und entsprechen unseren Lehr-Lern-Kulturen. Dies soll am Beispiel der Naturwisenschaften genauer erläutert werden. Im naturwissenschaftlichen Unterricht der Sekundarstufe I sind das Erforschen, Verstehen und Beschreiben von Phänomenen aus der Lebenswelt der Jugendlichen zentrale Anliegen.

> «Die Schülerinnen und Schüler kennen ihre natürliche Umwelt aus unmittelbarem Erleben und Beobachten. Sie sammeln und ordnen Objekte und Fakten aus Natur, Experimenten und anderen Quellen. ... Aufgrund der Auseinandersetzung mit der Natur und der vom Menschen veränderten Umwelt sind sie in der Lage, wesentliche Erscheinungen mit klaren Begriffen zu benennen. Dieses grundlegende Orientierungswissen hilft ihnen, an der Diskussion über aktuelle Probleme teilzunehmen. Zudem ermöglicht es ihnen den Einstieg in wissenschaftliches Denken.» (Lehrplan für die Volksschule des Kantons Zürich, 1991, S. 29).

Der Erwerb abstrakter Konzepte und Fachausdrücke wird bei uns auf dieser Schulstufe noch nicht als so wichtig erachtet. Diese Vorgabe der Lehrpläne wurde den Schweizer Jugendlichen im schriftlichen TIMSS-Leistungstest offenbar «zum Verhängnis». Sie verloren vor allem bei jenen Aufgaben Punkte, die Fachbegriffe oder Fachausdrücke verlangten (Ramseier, 1997; Stebler, Reusser, & Ramseier, 1997). Dieselbe Tendenz zeigt sich auch bei jenen Teilfragen des Experimentiertests, bei denen die Schlussfolgerungen aus den Daten anhand der Fachkonzepte 'Kinetik', 'Blutkreislauf' und 'Stromkreis' erläutert werden mussten. Anders als im schriftlichen Leistungstest konnten die Schülerinnen und Schüler diese Ausrutscher im Experimentiertest mit sehr guten Leistungen beim Sammeln, Organisieren und Beschreiben von Daten sowie beim Schlussfolgern kompensieren.

2. Die Experimentieraufgaben müssen für alle Probanden gleichermassen ***fair*** sein.

Da Experimentieraufgaben kontextspezifisch sind, ist eine angemessene thematische Vielfalt unabdingbar, um die Voraussetzungen der *Knaben* und der

Mädchen sowie Jugendlicher verschiedener *Ethnien* ausgewogen zu berücksichtigen. Ferner müssen Kodiersysteme entwickelt werden, welche die *intendierten Kompetenzen* der Prüflinge erfassen.

Da keine geschlechtsspezifischen Leistungsunterschiede gefunden wurden, kann der Experimentiertest in dieser Hinsicht als fair gelten. Dass Knaben und Mädchen ähnliche Leistungen erbrachten, dürfte weniger am Testformat als an den Inhalten der Aufgaben liegen. Bei Experimentieraufgaben zum Thema 'Elektrizität' beispielsweise, sind die Leistungen der Knaben üblicherweise deutlich besser als jene der Mädchen (Jovanovic, Solano-Flores, & Shavelson, 1994). Bei der Aufgabe 'Batterien' im TIMSS-Experimentiertest trat dieser Effekt jedoch nicht auf.

Die Ergebnisse des Experimentiertests stimmen weitgehend mit jenen des schriftlichen TIMSS-Leistungstests überein. In der *Mathematik* erbrachten die Knaben und die Mädchen auch dort ähnliche Leistungen (Keller, 1997). Dies entspricht dem aktuellen Forschungsstand. Bei Jugendlichen der Sekundarstufe I zeigen sich, sofern nicht Hochbegabte untersucht werden, in der Regel keine bedeutsamen geschlechtsspezifischen Leistungsunterschiede (Friedman, 1994; Frost, Hyde, & Fennema, 1994; Hyde, Fennema, & Lamon, 1990; Tartre & Fennema, 1995). In den *Naturwissenschaften* erzielten die Knaben der siebten Jahrgangsstufe im schriftlichen TIMSS-Leistungstest signifikant bessere Ergebnisse als die Mädchen, weil ihre Testwerte bei den Physik- und den Chemieaufgaben, nicht aber bei den Geographie-, Biologie- und Umweltaufgaben, deutlich höher waren. Allerdings ist zu beachten, dass die Effektgrösse des Unterschiedes so schwach ist (Keller, 1997), dass kein unterrichtspraktisch und bildungspolitisch relevanter Tatbestand vorliegt.

Das *Kodiersystem* von TIMSS ermöglichte eine *reliable* und effiziente Beurteilung der Schülerlösungen. Es wurde deutlich, dass die präzise Formulierung der Beurteilungskriterien besonders bei Fragen nach Fachkonzepten entscheidend ist. Damit hier nicht Scheinwissen erfasst wird, darf nicht bloss auf den Fachausdruck abgestellt werden, sondern es müssen die relevanten Beziehungen des Fachbegriffs operationalisiert werden.

3. Die ***limitierte Generalisierbarkeit*** der Testergebnisse muss berücksichtigt werden.

Die Leistungen beim Lösen von Experimentieraufgaben hängen stark von der je spezifischen Aufgabe und dem Testzeitpunkt ab (Ruiz-Primo & Shavelson, 1996). Es lassen sich daraus nur bedingt Schlüsse über Leistungen bei anderen Themen, in anderen Fachbereichen und bei anderen Aufgabentypen ableiten (Shavelson, et al. 1997). Wer Experimentieraufgaben bei klassenübergreifenden Leistungsvergleichen oder zur Selektion einsetzt, muss sich der limitierten Generalisierbarkeit der Ergebnisse bewusst sein.

Ruiz-Primo und Shavelson (1996) zufolge braucht es zwischen 8 und 23 Experimentieraufgaben, um zuverlässige Angaben über den Leistungsstand eines Individuums zu machen. Bei Vergleichen auf Klassen- und Schulebene

kann man mit weniger Aufgaben auskommen. Im TIMSS-Experimentiertest löste ein Proband drei bis fünf Handlungsprobleme (mit je zwei bis acht Teilaufgaben). Insgesamt wurden 12 Experimentieraufgaben eingesetzt. Die von TIMSS gewählte Anzahl Testaufgaben ist für *Leistungsvergleiche auf der Systemebene* angemessen, sofern sie im Rahmen eines messtechnisch validen Designs eingesetzt werden. Über die Leistungen einzelner Schüler oder Schülerinnen lassen sich auf der Basis von drei bis fünf Experimentieraufgaben aber keine Angaben machen.

4. Das ***Anspruchsniveau*** der Experimentieraufgaben muss bestimmt werden.

Um das Anspruchsniveau (cognitive complexity) einer Aufgabe zu bestimmen, ist eine Aufgabenanalyse erforderlich, bei der das zur Aufgabenlösung erforderliche Sach- und Problemlösewissen spezifiziert und die Wechselwirkungen zwischen Aufgabe und Testnehmer berücksichtigt werden.

Bei TIMSS wurden Informationen über das Anspruchsniveau der Experimentieraufgaben durch Expertenurteile und einen Pilottest eingeholt. Zur Beschreibung der erforderlichen Denk- und Problemlösekompetenzen wurde das Konstrukt der *Leistungserwartung* verwendet. Die Kategorien wurden vorwiegend induktiv gewonnen. Dies mag mit ein Grund dafür sein, dass sich die bei TIMSS verwendeten Begriffe Routineverfahren und Problemlösen nicht vollständig mit unserem kognitionspsychologischen Verständnis decken.[8]

5. Die Experimentieraufgaben müssen sich auf ***bedeutsame Inhalte*** des Faches beziehen.

Es müssen mit Hilfe von Experten Inhalte gewählt werden, die für das entsprechende Fach relevant sind und einen hohen Bildungswert haben.

Diese Anforderung ist unseres Erachtens mit Bezug auf die formalen und die materialen Aspekte des Experimentiertests erfüllt. Die Anwendungsprobleme verlangen Routinen und Strategien, die für eigenständiges praktisches Problemlösen zentral sind. In materialer Hinsicht sind sie *exemplarisch* für grundlegende mathematische und naturwissenschaftliche Sachverhalte (z.B. Blutkreislauf, Wahrscheinlichkeit, Strahlensatz).

6. Die Experimentieraufgaben müssen den zu testenden ***Bereich gut abdecken***.

Im Gegensatz zu herkömmlichen Schulleistungstests, wo es vor allem darauf ankommt, dass eine relevante Auswahl zentraler Fachkonzepte abgedeckt ist *(content sampling)*, wird bei Experimentiertests eher darauf geachtet, dass die für wissenschaftliches Arbeiten wichtigen Problemlöse- und Experimentiertätigkeiten berücksichtigt werden (*process sampling*).

Dies trifft für den TIMSS-Experimentiertest zu. In formaler Hinsicht lösen die Experimentieraufgaben ein reichhaltiges *Problemlöse- und Experimentierverhalten* aus. Hinsichtlich des Fachwissens ist das Spektrum dagegen sehr eng. Fünf Physik- und eine Biologieaufgabe dürften den Bereich der Naturwissenschaften, der im schriftlichen TIMSS-Leistungstest die Physik, Biologie, Chemie, Geographie und Umweltwissenschaften einschloss, kaum abdecken. Es ist daher fraglich, ob die Ergebnisse der beiden Testteile in materialer Hinsicht vergleichbar sind.

7. Die ***Experimentieraufgaben*** müssen für die Probanden interessant, bedeutsam und lehrreich sein.

Eine gute Testaufgabe ist eine interessante und anspruchsvolle Anwendungsaufgabe, lautet die Devise. Es ist bekannt, dass sich manche Jugendliche bei Leistungstests nicht voll einsetzen, weil die Aufgaben unattraktiv sind. Auch bei der TIMSS Erhebung war es für die Schülerinnen und Schüler vermutlich spannender, selbständig kleine Experimente durchzuführen als Fragen mit Mehrfach- und Kurzantworten zu bearbeiten, deren Lösungen keinen Einfluss auf die Noten hatten.

8. ***Aufwand und Ertrag*** müssen in einem vertretbaren Verhältnis stehen.

Die Entwicklung und Durchführung von Experimentiertests ist verglichen mit standardisierten Leistungstests sehr kosten-, personal- und zeitaufwendig. Der Einsatz von Experimentiertests ist vorerst nur dann angezeigt, wenn Kompetenzen gemessen und gefördert werden sollen, für die sich ökonomischere Testinstrumente nicht eignen.

Zur Zeit erarbeiten in den USA verschiedene Forschergruppen Handreichungen zur Entwicklung ökonomischer und reliabler Experimentiertests (Shavelson, et al. 1997). In England sind Experimentieraufgaben in Prüfungen und Unterricht etabliert (Gage & Berliner, 1992; Gee & Clackson, 1992). Im deutschsprachigen Raum hat dieses Testformat noch nicht Fuss gefasst. Dies mag damit zusammenhängen, dass man hier keine Erfahrungen mit landesweiten Leistungsvergleichen hat und somit kein Bedarf besteht, die unerwünschten Nebenwirkungen zu kurieren. Solange die bislang angesprochenen messtechnischen Probleme vor allem mit Bezug auf die Generalisierbarkeit und Konstruktvalidität (Leistungserwartungen) nicht befriedigend gelöst sind und für unseren Sprachraum noch keine Wegleitungen zur Konstruktion, Durchführung und Auswertung vorliegen, können wir den Einsatz von Experimentieraufgaben zu klassenübergreifenden Leistungsvergleichen nur mit Vorbehalten empfehlen. Für den Unterricht sind sie unseres Erachtens vorzüglich geeignet.

Ein Hauptziel unserer Lehrpläne ist der Erwerb *eines flexibel anwendbaren Sach- und* Problemlösewissens. Experimentieraufgaben sind ein erfolgversprechendes didaktisches Arrangement, um Alltags- und Fachwissen zu vernetzen und positive Lernübertragung anzuregen. Allerdings muss gewährleistet werden, dass das Experimentieren nicht zum Manipulieren verkommt. Entscheidend ist, dass die Experimentieraufgaben als produktive Lernaufgaben in thematische Unterrichtseinheiten eingebettet werden und die Schülerinnen und Schüler erfahren, dass es beim Experimentieren nicht um spektakuläre Effekte geht, sondern darum (Schauble, Glaser, Duschl, Schulze, & John, 1995),

- die Beziehungen zwischen Ursachen und Wirkungen systematisch zu erforschen,
- ausgewählte Variablen begründet zu variieren,
- sinnvolle Muster zu entdecken,
- Vergleiche und Beurteilungen auf mehreren Dimensionen vorzunehmen und
- Beziehungen zwischen Experiment und Realität herzustellen.

Experimentieraufgaben eignen sich sowohl für individuelles als auch für kooperatives Lernen. Beim paar- und gruppenweisen Experimentieren können die Schülerinnen und Schüler ihr sachbezogenes Alltags- und Schulwissen aktivieren, vergleichen, diskutieren, reflektieren und integrieren. Es können, um einige Schülerantworten zur Tablettenaufgabe aufzunehmen, Beziehungen hergestellt werden zwischen dem Schmelzen der Schokolade, dem Verdampfen des Spaghettiwassers, dem Abtrocknen des heissen Geschirrs, dem Überkochen der Milch, sowie dem Abkühlen des Badewassers und dem gleichzeitigen Steigen der Raumtemperatur. Dabei können die Schülerinnen und Schüler feststellen, dass Stoffe ihren Zustand ändern (Aggregatzustand), dass ein Wärmefluss stattfindet (Wärmeausbreitung) und dass mit steigender Temperatur die Energie und in der Regel auch das Volumen zunimmt, was auf das Verhalten der Elementarteilchen zurückzuführen ist (Kinetik). Die Lehrperson kann diese Erfahrungen und Erkenntnisse aufnehmen, Vergleiche anregen, Begründungen verlangen, zusammen mit den Schülerinnen und Schülern auf den Begriff bringen und den naturwissenschaftlichen Fachausdruck einführen (Brown, Campione, Webber, & McGilly, 1992; Collins, Brown, & Newman, 1989; Hatano & Inagaki, 1991; Labudde, 1993). Bei diesem Tun entsteht *abstrahiertes Wissen*, das in der Regel besser vernetzt und leichter transferierbar ist, als das von situativen und phänomenhaften Bezügen entblösste, abstrakte Wissen, das im Unterricht oft direkt zu vermitteln versucht wird. Gleichzeitig werden die Schülerinnen und Schüler mit *den Grundfertigkeiten des wissenschaftlichen Arbeitens* vertraut. Sie lernen, wie man Hypothesen aufstellt, prüft, überarbeitet, beurteilt, verwirft oder annimmt.

Im weiteren können die Schülerinnen und Schüler beim Experimentieren erleben, dass mathematische und naturwissenschaftliche Gesetze *keine unumstösslichen Satzungen* sind. Was die momentan beste Erklärung für ein Phänomen ist, wird in der Forschungsgemeinschaft ausgehandelt. Auch die gemeinsam akzeptierte und fortan gültige Erklärung lässt Fragen offen und kann bei Bedarf

revidiert werden (Labudde, 1996). «Mathematik ist eine Wissenschaft mit langer Tradition, sie erfährt aber auch dauernd Wandlungen, Ergänzungen und Weiterentwicklungen» (Lehrplan für die Volksschule des Kantons Zürich, 1991, S. 255). Guter Unterricht sensibilisiert die Schülerinnen und Schüler für solche Veränderungen und legt so einen Grundstein für lebenslanges Lernen.

Durch die Teilnahme am TIMSS-Experimentiertest haben wir eine interessante, prozessorientierte und anspruchsvolle Art der Leistungsmessung kennengelernt, die unseren Lehrplänen entspricht, durch einen erweiterten kognitionspsychologischen Lernbegriff begründet ist und anderen Facetten der Intelligenz Rechnung trägt als viele herkömmliche Schulleistungstests. Aufgrund der noch nicht befriedigend gelösten, messtechnischen Probleme und im Kontext unserer Lehr-Lern- und Prüfungskultur sehen wir den Nutzen von Experimentieraufgaben primär im unterrichtlichen Einsatz zur integrierten Förderung formaler und materialer Kompetenzen.

Fussnoten

1 Die vorliegende Untersuchung ist eingebettet in das Forschungsprojekt 'Schule, Leistung und Persönlichkeit' und gehört zum Nationalen Forschungsprogramm 33. Sie wurde unterstützt vom Schweizerischen Nationalfonds zur Förderung der wissenschaftlichen Forschung und von der Stiftung für wissenschaftliche Forschung an der Universität Zürich. An der Durchführung und Auswertung der Studie war neben dem Autorenteam auch Susanne Leibundgut massgeblich beteiligt. Studierende der Pädagogik an den Universitäten Zürich und Bern halfen bei der Erhebung, Kodierung und Eingabe der Daten. Diesen Personen, aber auch den Lehrkräften sowie den Schülerinnen und Schülern, die am Experimentiertest teilgenommen haben, möchten wir an dieser Stelle herzlich danken.

2 Die Auswertung des schriftlichen Leistungstests auf der Sekundarstufe I liegt als Buch vor (Moser, Ramseier, Keller, & Huber, 1997). Die Ergebnisse für die Sekundarstufe II sind noch nicht veröffentlicht.

3 Die Schweiz ist das einzige Land im deutschen Sprachraum, das am TIMSS-Experimentiertest teilgenommen hat.

4 Die Bezeichnung 'Performance Assessment' wird uneinheilich verwendet. Manche Leute verstehen darunter alle Formen der Leistungsmessung, die nicht auf dem multiple-choice Format beruhen.

5 Das Schlagwort 'higher order thinking' wird definitionsgemäss für kognitive Aktivitäten verwendet, bei denen Informationen nicht bloss abgerufen, sondern in irgendeiner Form transformiert werden müssen (Baker, 1990).

6 Um die Absurdität dieser Art von Berichterstattung zu verdeutlichen, schlagen Gage und Berliner (1992) vor, dass man einmal die Sterberate aller Spitäler eines Staates rangieren und der Leserschaft die Hintergrundinformationen vorenthalten könnte.

7 Da bei der zweiten Teilaufgabe zwei Aspekte beurteilt wurden, enthält die Abbildung sechs Balken.

8 Die Erläuterung dieser Aussage würde den Rahmen dieses Textes sprengen.

Literatur

Adams, M. J. (1989). Thinking skills curricula: Their promise and progress. *Educational Psychologist, 24*(1), 25-77.

Anderson, J. P., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11.

Baker, E. L. (1990). Developing comprehensive assessments of higher order thinking. In G. Kulm (Ed.), *Assessing higer order thinking in mathematics* (pp. 7-20). Washington: American Association for the Advancement of Science.

Baxter, G. P., Elder, A. D., & Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational Psychologist, 31*(2), 133-140.

Baxter, G. P., & Shavelson, R. J. (1994). Science performance assessment: Bechmarks and surrogates. *International Journal of Educational Research, 21*(3), 279-298.

Brown, A. L. (1983). Metakognition, Handlungskontrolle, Selbststeuerung und andere, noch geheimnisvollere Mechanismen. In F. E. Weinert & R. H. Kluwe (Hrsg.), *Metakognition, Motivation und Lernen* (S. 60-109). Stuttgart: Kohlhammer.

Brown, A. L., Campione, J. C., Webber, L. S., & McGilly, K. (1992). Interactive learning environments: A new look at assessment and instruction. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments. Alternative views of aptitude, achievement and instruction* (pp. 121-211). Boston: Kluwer Academic Publishers.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction. Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dewey, J. (1933/ 1960). *How we think: a restatement of the relation of reflective thinking to the educative process.* Lexington, Mass: Heath.

Erziehungsdirektion des Kantons Zürich (1991) (Hrsg.). *Lehrplan für die Volksschule des Kantons Zürich.* Lehrmittelverlag des Kantons Zürich.

Flavell, J. H. (1983). Annahmen zum Begriff Metakognition sowie zur Entwicklung von Metakogntion. In F. E. Weinert & R. H. Kluwe (Hrsg.), *Metakognition, Motivation und Lernen* (S. 223-231). Stuttgart: Kohlhammer.

Friedman, L. (1994). Meta-analytic contributions to the study of gender differences in mathematics: The relationship of mathematical and spatial skills. *International Journal of Educational Research, 21*(361-371).

Frost, L. A., Hyde, J. S., & Fennema, E. (1994). Gender, mathematics performance, and mathematics-related attitudes and affect: A meta-analytic synthesis. *International Journal of Educational Research, 21,* 373-385.

Gage, N. L., & Berliner, D. C. (1992). *Educational Psychology* (5. Ed.). Boston: Houghton Mifflin Company.

Gardner, H. (1994). Abschied vom IQ. Die Rahmen-Theorie der vielfachen Intelligenzen. Stuttgart: Klett-Cotta.

Gee, B., & Clackson, S. G. (1992). The origin of practical work in the English school science curriculum. *School Science Review, 73*(265), 79-83.

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.* Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Hatano, G., & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 331-348). Washington, DC: American Psychological Association.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Sex differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155.

Ingenkamp, K. (1985). *Lehrbuch der pädagogischen Diagnostik.* Weinheim: Beltz.

Jovanovic, J., Solano-Flores, G. & Shavelson, R.J. (1994). Performance-based assessments. Will gender differences in science achievement be eliminated? *Education and Urban Society, 26,* 352-366.

Keller, C. (1997). Geschlechterdifferenzen: Trägt die Schule dazu bei? In U. Moser, E. Ramseier, C. Keller, & M. Huber, (Hrsg.). *Schule auf dem Prüfstand. Eine Evaluation der Sekundarstufe I auf der Grundlage der «Third International Mathematics and Science Study»* (S.137-179). Zürich: Rüegger.

Kerschensteiner, G. (1926). *Theorie der Bildung.* Leipzig: Teubner.

Labudde, P. (1993). *Erlebniswelt Physik.* Bonn: Dümmler.

Labudde, P. (1996). Genetisch-sokratisch-exemplarisches Lernen im Lichte der neueren Wissenschaftstheorie. *Beiträge zur Lehrerbildung, 14*(2), 170-174.

Langer, E. (1989). *Mindfulness.* Reading, MA: Addison-Wesley.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

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.

Nuñes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics.* Cambridge: Cambridge University Press.

Ramseier, E. (1997). Leistungsprofil und Unterricht. Eine Analyse der schweizerischen Leistungen im naturwissenschaftlichen Test von TIMSS *[in diesem Heft].*

Reusser, K. & Stebler, R. (1997). Every word problem has a solution - The social reality of mathematical modeling in schools. *Learning and Instruction, 7*(4), 309-327.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments. Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Robitaille, D.F., Schmidt, W.H., Raizen, S., Mc Knight, C., Britton, E., & Nicol, C. (1993). *Curriculum Frameworks for Mathematics and Science.* TIMSS Monograph No. 1. Vancouver: Pacific Educational Press.

Robitaille, D. F. & Garden, R. A. (1996). TIMSS Monograph No. 2: Research Questions & Study Design. Vancouver: Pacific Educational Press.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Rhetoric and reality in science performance assessment: An update. *Journal of Research in Science Teaching, 33*(10), 1045-1063.

Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences, 4*(2), 131-166.

Shavelson, R. J. (1996). On the romance of science curriculum and assessment-reform in the United-States. In D. K. Sharpes & A. L. Leino (Eds.), *The dynamic concept of curriculum: Invited papers to honour the memory of Paul Hellgren. Reserach Bulletin 90* (pp. 57-75). Helinki, Finland: University of Helsinki.

Shavelson, R. J., & Baxter, G. P. (1992). Linking assessment with instruction. In F. F. Oser, A. Dick & J-P. Patry (Eds.), *Effective and responsible teaching* (pp. 80-90). San Francisco: Jossey-Bass Publishers.

Shavelson, R. J., Solano-Flores, G., & Ruiz-Primo, M. A. (1997). *Toward a science performance assessment technology.* Paper presented at the 7th EARLI Conference Athens, Greece, 26-30 August.

Shepard, L. A. (1989). Why we need better assessments. *Educational Leadership, 46*(7), 4-9.

Shepard, L. A. (1992). Commentary: What policy makers who mandate tests should know about the new psychology of intellectual ability and learning. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments. Alternative views of aptitude, achievement and instruction* (pp. 301-328). Boston: Kluwer Academic Publishers.

Silver, E. A. (1992). Assessment and mathematics education reform in the United States. *International Journal of Educational Research, 17*(5), 489-502).

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

Solano-Flores, G., & Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical, and logistical issues. *Educational Measurement: Issues and Practice, 16*(2), 16-25.

Stebler, R., Reusser, K., & Pauli, C. (1994). Interaktive Lehr-Lern-Umgebungen: Didaktische Arrangements im Dienste des gründlichen Verstehens. In K. Reusser & M. Reusser-Weyeneth (Hrsg.), *Verstehen. Psychologischer Prozess und didaktische Aufgabe* (S. 227-259). Bern: Huber.

Stebler, R., Reusser, K., & Ramseier, E. (1997). Spitzenleistungen der Schweizer Siebtklässler im TIMSS-Experimentiertest. *Schweizer Lehrerinnen- und Lehrerzeitung - SLZ, 10,* 18-21.

Sternberg, R. J. (1992). CAT: A program of comprehensive abilities testing. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments. Alternative views of aptitude, achievement and instruction* (pp. 213-274). Boston: Kluwer Academic Publishers.

Tartre, L. A., & Fennema, E. (1995). Mathematics achievement and gender: A longidudial study of selected cognitive and affective variables [grades 6-12]. *Educational Studies in Mathematics, 28*(3), 195-198.

Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership, 49*(8), 26-33.

Wolf, R. M. (1994). Performance assessment in IEA studies. *International Journal of Educational Research, 21*(3), 239-245.

Des travaux pratiques pour développer l'ensemble des compétences des élèves. Résultats des tests de l'étude TIMSS

Résumé

Les expériences réalisées dans le cadre de l'étude TIMSS dans 44 classes de Suisse alémanique se sont appuyées sur des tests pratiques qui ont permis d'évaluer les prestations des élèves de 7e année en sciences et en mathématiques: il s'est agi pour eux de concevoir et de réaliser des expériences, d'en formuler les résultats, puis de dégager de ces données certaines tendances, pour en tirer des conclusions et les analyser à l'aide des outils conceptuels spécifiques de ces disciplines. Au niveau international, les résultats de ces élèves ont été remarquables, sans qu'il y ait à cet égard à distinguer entre filles et garçons. S'agissant maintenant d'une comparaison entre différentes classes, il existe encore certaines difficultés non résolues concernant les critères théoriques d'appréciation, et il convient donc de limiter pour l'instant la confiance que l'on peut placer dans ces tests d'évaluation comparative. L'introduction de ce type d'expérimentations pratiques visant à développer les aptitudes pratiques et théoriques des élèves en les mettant en situation de devoir résoudre un problème donné est entièrement compatible avec les grandes lignes de nos plans d'études.

Applicazioni pratiche per lo stimolo integrato di competenze formali e materiali - risultati dal test sperimentale TIMSS

Riassunto

Il test TIMSS è stato applicato in 44 classi della svizzera tedesca del VII anno scolastico (N=396) per problemi a carattere matematico e scientifico. I giovani dovevano pianificare e realizzare delle sperimentazioni, descrivere i risultati, dedurne i trends e spiegarli utilizzando la terminologia matematica e scientifica. Nel confronto internazionale gli allievi svizzeri tedeschi hanno ottenuto risultati brillanti. Non si sono riscontrate differenze nel rendimento delle ragazze e dei ragazzi. Confronti tra le classi sono attualmente possibili solo con le dovute riserve a causa di problemi teorici di misurazione. L'attuazione dei compiti pratici per lo stimolo delle competenze formali e materiali nella risoluzione individuale di problemi e in un ambiente di insegnamento-apprendimento interattivo corrisponde agli obiettivi dei nostri piani di studio.

Practical Tasks for Integrated Furtherance of Formal and Material Competence - Results from the TIMSS-Experiment Test

Summary

For the TIMSS-experiment test, mathematical and scientific action problems were used to measure the achievement in 44 German-Swiss classes of the seventh year level. The students had to plan and carry out experiments, write notes of the results, gather tendencies, and draw conclusions and explain these using mathematical and scientific technical terms. In an international comparison, the German-Swiss seventh-formers come off very well. The achievements of the girls and the boys do not differentiate. Presently, comprehensive comparisons of achievement with experiment tasks cannot be recommended without reservations because of unsolved, measuring-theoretical problems. The use of practical tasks for integrated furtherance of formal and material competence in individual problem solving and in interactive teaching-learning environments corresponds to the central ideas of our curricula.