Was testen Tests?

Objektiv-hermeneutische Analysen am Beispiel von TIMSS und PISA.

Dissertation zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
an der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Potsdam

eingereicht von Wolfram Meyerhöfer,
Helene-Lange-Straße 7, 14469 Potsdam

Dezember 2003

Gutachter:

Prof. Dr. Thomas Jahnke, Universität Potsdam, Institut für Mathematik

PD Dr. Andreas Wernet, Universität Potsdam, Institut für Pädagogik

Prof. Dr. Werner Peschek, Universität Klagenfurt, Institut für Mathematik

Inhaltsverzeichnis

EINLEITUNG

In dieser Arbeit wird der Frage nachgegangen, was in den Mathematikleistungstests von TIMSS[1] und PISA[2] – als exponierte Beispiele für internationale Vergleichsuntersuchungen – gemessen wurde.

Die Frage erscheint zunächst überflüssig, weil dort offensichtlich Mathematikleistungen gemessen wurden. Ein mathematischer Leistungstest legt gewissermaßen per se fest was Mathematikleistungen – oder mathematische Leistungsfähigkeit oder mathematische Leistung(skraft) – sind.
In dieser Arbeit wird untersucht, wie bei diesen Tests „mathematische Leistungsfähigkeit" unter Rückgriff auf mathematikdidaktische Begrifflichkeiten konstruiert ist. Es wird diskutiert, wie mathematische Leistungstests festlegen, was mathematische Leistung ist. Es wird ebenfalls diskutiert, zu welchen verschiedenen Konstrukten von mathematischer Leistung TIMSS und PISA gelangen. In dieser Arbeit wird aber vor allem anhand der Testaufgaben ausgelotet, was die Tests testen:

Die Geltung einer Testaussage wird auf drei Ebenen erzeugt. Die erste Ebene ist die Ebene der Testaufgabe: Mit ihr soll nachvollziehbar eine bestimmte Eigenschaft getestet werden. Erst wenn man weiß, was jede einzelne Aufgabe mißt, kann man die Gesamtheit des Getesteten wissen. An der Testaufgabe wird die Geltung erzeugt, weil dort der Schüler das tut, was ihm einen Testpunkt verschafft. – Der Schüler setzt ein Kreuz oder er schreibt eine Zahl beziehungsweise eine Argumentation hin. Die Geltungserzeugung an der Aufgabe findet ihren materiellen Abschluß, wenn entweder der Scanner das Kreuz an der richtigen Stelle identifiziert oder wenn der menschliche Kodierer einen Punkt vergibt, weil eine richtige Zahl da steht oder weil die Argumentation durch einen Punkt belegt werden kann.
Die zweite Ebene der Geltungserzeugung ist die Ebene des Gesamttests: Die Testaufgaben müssen so konstruiert bzw. verzahnt sein, dass die Aufgabengesamtheit jene Fähigkeitsgesamtheit mißt, welche getestet werden soll.
Die dritte Ebene ist die Ebene der zu testenden Population mit der Möglichkeit einer Gesamtaussage über diese Population bzw. über Individuen aus dieser Population.
Die Geltung der auf jeder Ebene getroffenen Aussage muß jeweils überprüfbar, das heißt in ihrer Genese wiederholbar und nachvollziehbar sein.

In dieser Arbeit wird ausschließlich die Geltungserzeugung an der Aufgabe betrachtet. Hier hat die derzeitige Testerstellung die größten Defizite zu verzeichnen. - Diese Arbeit ist dabei nicht vorrangig ein Plädoyer für sorgfältigere Testerstellung. Sie diskutiert Grundfragen des Testens als Hilfsmittel wissenschaftlichen Arbeitens und als Instrument zur Vergabe von Zukunftschancen. Sie stellt mit der Objektiven Hermeneutik auch eine Methode vor, mit der Testaufgaben auf ihre Meßeigenschaften hin untersucht werden können. Da dabei für TIMSS und PISA erhebliche Meßschärfen, Fehler, Mitmes-

[1] TIMSS: Third International Mathematics and Science Study
[2] PISA: Programme for International Student Assessment

sen von Testfähigkeiten, sprachliche Verwerfungen und Zerstörungen des Mathematischen konstatiert werden, stellt diese Arbeit die Eignung dieser Tests als Meßinstrumente in Frage.

Vor dem Hintergrund dieser Defizite erscheinen die statistischen Probleme eher zweitrangig – wobei auch sie zu mehr Sorgfalt in der wissenschaftlichen Debatte aufrufen. Als Beispiel möchte ich die verbreitete Fehldeutung diskutieren, dass bei PISA die Leistungsfähigkeit des deutschen Mathematikunterrichts gemessen worden sei: Im internationalen PISA-Test wurde die Gruppe der 15jährigen vermessen. Diese Gruppe umfaßt Schüler verschiedener Klassenstufen. Man setzt bei PISA voraus, dass die Beschulung von Schülern einen Lerneffekt hervorruft, sonst könnte man nicht mit Leistungstests Schulsysteme vergleichen wollen. Wenn man das voraussetzt, dann muß ein durchschnittlicher Zehntklässler höhere Leistungen vollbringen als ein durchschnittlicher Neuntklässler. Das bedeutet aber, dass die Durchschnittsleistung der 15jährigen davon abhängt, wie groß der Anteil der verschiedenen Klassenstufen in den verglichenen Populationen ist.

Mit der Frage, wer vermessen wird, wird nun offenbar auch die Antwort auf die Frage verschoben, was gemessen wird: Es stellt sich heraus, dass bei PISA nur zum (geringeren?) Teil die Leistungsfähigkeit des Mathematikunterrichts oder des Bildungssystems gemessen wird, ebenso aber die Schnelligkeit, mit der ein Bildungssystem seine Schüler durchschleust. Dafür gäbe es allerdings sehr viel unvermitteltere Meßverfahren, z.B. stellen Statistiken zu Abgangsalter, Wiederholungen, Überspringen usw. aussagekräftigere und preiswertere Daten zur Verfügung. Außerdem ist bei diesen Daten die Frage, welche Information in den Daten steckt, was also gemessen bzw. beschrieben wird, viel einfacher zu beantworten als bei PISA. Hier überlagert die Messung der Schnelligkeit die Messung der Leistung, es ist aus den Mittelwerten der Testdaten nicht mehr herauszulesen, ob sie durch „echte“ Leistungsunterschiede oder durch Unterschiede in der Klassenverteilung zustande kommen.

Solche und ähnliche Betrachtungen werden in dieser Arbeit ausgespart. Sie reproduzieren die Denkmuster der Tester in immer filigranere statistische Konstrukte mit immer mehr Probanden, die zur Aufrechterhaltung von Modellbedingungen benötigt werden. Der immer gewaltigere statistische Apparat produziert nicht nur immer beeindruckendere technische Werke, sondern auch immer größeres Unbehagen. Das liegt nicht nur an der größeren Fragilität dieser technischen Werke. Das liegt auch an der Unüberschaubarkeit des Weges von den Daten zu den Schlußfolgerungen – selbst offensichtliche Ungereimtheiten werden von vielen Beobachtern nicht benannt, weil sie sich nicht sicher sind, den Apparat wirklich vollständig durchschaut zu haben und weil Irrtümer hier offenbar nicht erlaubt sind. Das Unbehagen schleicht sich auch ein, weil diese technischen Meisterwerke – obwohl in wesentlichen Teilen geheim – umgedeutet werden zum Maßstab schulischen Tuns. Wenn schulisches Handeln nach dem Abschneiden bei Tests beurteilt wird, dann werden Tests eben zum Maß. Das Unbehagen speist sich nicht nur aus dem Geheimen und aus der Entfremdung zwischen Lernen und Leisten. Es speist sich auch aus dem Getesteten selbst, das offensichtlich weit weniger ist als das, was Schule zu erarbeiten sucht.

Hier schließt diese Arbeit an. Sie untersucht den Charakter dieser Entfremdung. Und sie erklärt das Unbehagen, indem sie aufzeigt, wodurch die Aufgaben anderes testen, als sie zu testen vorgeben und als mit ihnen getestet werden soll. Es bleibt dabei nicht bei stoffdidaktisch-textlichen Erkenntnissen. Es offenbart sich auch ein mathematikdidaktischer Habitus, der über Testaufgaben hinausweist. Er ist geprägt durch Zerstörung des Mathematischen bei gleichzeitiger Orientierung auf Fachsprachlichkeit, durch ein Mißlingen der Synthese von Mathematischem und Realem (weil beides in seiner Autonomie und Authentizität nicht ernst genommen wird), durch die Herstellung einer Illusion von Schülernähe und durch Kalkülorientierung statt Orientierung an mathematischer Bildung. Man kann diesen Habitus unter dem Stichwort der „Abkehr von der Sache" zusammenfassen. Auch dieser Habitus dürfte seinen Beitrag zum Unbehagen leisten. Die Rekonstruktion dieses Habitus in den Testaufgaben verweist auf mathematikdidaktische Ideologie, die unreflektiert in Tests transportiert wird.

Der Aufbau dieser Arbeit spiegelt ihre Entstehungsgeschichte. Ich stieß auf das Problem der Testaufgaben, nachdem wir in einer Lehrveranstaltung eine Aufgabe aus einem Mathematiklehrbuch interpretiert hatten, welche Schüler in einer von uns analysierten Schulstunde lösen mußten. Diese Aufgabe wies sprachliche Verwerfungen auf, die im weiteren Unterrichtsverlauf zu Unterrichtsstörungen führten. Als Gegenbeispiel einer gelungenen Aufgabe schlug ich die TIMSS-Aufgabe M7 vor, bei der sich aber in der Interpretation ebenfalls Verwerfungen zwischen manifestem und latentem Aufgabentext zeigten. Ich untersuchte daraufhin mehrere zufällig gewählte TIMSS-Aufgaben, aber erst die fünfte Aufgabe erwies sich als frei von Verwerfungen. In der Zwischenzeit erschien die Debatte zwischen Hagemeister und der TIMSS-Gruppe, welche im ersten Kapitel dieser Arbeit thematisiert ist. Mir wurde klar, dass Hagemeister ähnliche Verwerfungen beschrieb, wie ich sie gefunden hatte. Er hatte aber weder klare Begrifflichkeiten noch theoretisches Rüstzeug, um das Problem genau und nachvollziehbar zu beschreiben (siehe Abschnitt 1.1.). Seine Argumente schienen vom technischen Apparat der TIMSS-Gruppe hinweggefegt zu werden. Allerdings nur auf den ersten Blick, denn die von Hagemeister aufgezeigten Probleme waren so massiv und so deutlich sichtbar, dass es ausgesprochen seltsam war, dass die TIMSS-Gruppe sie nicht nur hinwegrechnen konnte, sondern sie auch theoretisch verleugnete. Da es wenig Sinn gehabt hätte, daraufhin das gesamte Theoriegebäude der TIMSS-Gruppe auf seine Schlüssigkeit zu untersuchen, beschränkte ich mich auf jenen Teil des Theoriegebäudes, der sich mit den Aufgaben befaßte. Daraus entstand das Kapitel 1.2. Es stellte sich heraus, dass die TIMSS-Gruppe den hier untersuchten Theorieteil nicht wirklich bei der Konstruktion oder der Interpretation der Aufgaben benutzt haben kann, sondern diesen Theorieteil nur vorschiebt, um gegenüber Hagemeister den Eindruck zu erwecken, theoretisch abgesichert gearbeitet zu haben. Dieses Prinzip der Zerstörung von Wissenschaftlichkeit findet sich später auch bei PISA wieder.

Im zweiten Kapitel wird eine Einführung in die Methodologie und Praxis objektiv-hermeneutischen Arbeitens gegeben. Diese relativ ausführliche Darstellung ist einerseits nötig, da mit dieser Arbeit die Objektive Hermeneutik neu in die Mathematikdidaktik eingeführt wird. Sie wird andererseits dabei helfen, die Möglichkeiten und Grenzen objektiv-hermeneutischen Arbeitens bei der Untersuchung von Testaufgaben, aber auch von anderen Texten, einzuschätzen.

Das dritte Kapitel trifft grundsätzliche Aussagen zum Testen. Hier wird einerseits der Status und die Rolle von Tests im wissenschaftlichen Erkenntnisprozeß diskutiert. Andererseits werden Tests in ihrer gesellschaftlichen Einbettung betrachtet.

Das vierte Kapitel betrachtet den TIMSS-Test. Es werden all jene Aufgaben interpretiert, die jeder am Test beteiligte Schüler zu lösen hatte. In diesem Kapitel werden ausschließlich die Meßeigenschaften der Aufgaben untersucht.

Im fünften Kapitel wird der PISA-Test betrachtet. Dazu wird aus den veröffentlichten Aufgaben eine kontrastierende Aufgabenauswahl getroffen. Dabei wird die Kontrastierung nach jenen Problempunkten vorgenommen, die sich in den Interpretationen zeigen. Sie gehen über Meßprobleme hinaus. Da TIMSS im Gegensatz zu PISA auf mathematikdidaktischer Grundlage erstellt ist, lohnt sich auch eine nähere Untersuchung jener Probleme, die über reine Meßprobleme hinausgehen. Dies sind insbesondere Probleme im Umgang mit dem Realen bzw. Habitusprobleme.

Im Schlußkapitel finden sich zusammenfassende Betrachtungen zur Meßschärfe, zur Testfähigkeit und zum Raten.

Editorische Hinweise:

Dieser Text bedient sich der alten und der neuen Rechtschreibung.

Bei Zitaten finden sich alle Hervorhebungen im Original.

Alle Texte aus dem Buch *Deutsches PISA-Konsortium (Hrsg.): PISA 2000. (Opladen 2001)* werden zur einfacheren Lesbarkeit als *PISA 2000* zitiert. Der Text der Seiten 15 bis 68 ist *Baumert, Stanat und Demmrich (2001).* Der Text der Seiten 69 bis 140 ist *Artelt, Stanat, Schneider und Schiefele (2001).* Der Text der Seiten 141 bis 191 *ist Klieme, Neubrand und Lüdtke (2001).*

Kapitel 1: Diskussion der bei TIMSS und PISA verwendeten theoretischen Modelle zur Beurteilung von Testaufgaben[3]

Die Frage, was man mit einer Aufgabe mißt, kann in die Frage gewendet werden, welche Leistungen das Individuum erbringen muß, um die Aufgabe zu lösen. Diese Frage ist erstaunlich spärlich beleuchtet, obwohl sie konstitutiv für jeden Test ist. Die TIMSS-Gruppe baut ihre diesbezügliche Argumentation auf

- innerstatistische Absicherungen, deren Grenzen für die Beurteilung des durch die Aufgabe Gemessenen in dieser Arbeit aufgezeigt wird,
- ihrem eigenen Konstrukt der „Fähigkeitsniveaus", das ich separat diskutiere[4] und dessen Unbrauchbarkeit für die Beschreibung von Aufgabenanforderungen und „Fähigkeitsanalysen" sich anhand mehrerer Aufgaben zeigen wird,
- die von Kurt Reusser (1989) beschriebene Theorie der „mentalen Situationsmodelle" auf. Reusser hat ein computerimplementiertes kognitionspsychologisches Modell entwickelt, das beschreibt, wie Kinder einfache Rechenaufgaben lösen. Neu war dabei, dass die Entwicklung eines „mentalen Situationsmodells" beim problemlösenden Individuum in den Mittelpunkt der Betrachtung des kognitiven Prozesses rückt. „Unter einem Situationsmodell wird der semantische Zusammenhang verstanden, den eine Person in ihrem Geist erzeugt, wenn sie einen (narrativen) Text liest. Phänomenologisch gesprochen, handelt es sich beim Situationsmodell um den *intentionalen Gegenstand* bei der *Vergegenwärtigung* eines Handlungs- und Situationszusammenhangs. Das Situationsmodell ist jene personale kognitive Struktur, worauf sich der Verstehensvorgang richtet. Ein Situationsmodell ist das kognitive Korrelat der vom Autor eines Textes *gemeinten* bzw. von einem Leser *verstandenen* Situationsstruktur (...)" (Reusser 1989, S.136)

Bei der folgenden Darstellung gehe ich von einem aussagekräftigen Beispiel aus und lasse die theoretische Betrachtung daraus herauswachsen, was auch die Genese meiner eigenen Erkenntnis aufnimmt:
Kurz nach der publikumswirksamen Veröffentlichung der ersten TIMSS-Ergebnisse meldete sich Volker Hagemeister (1999) in der Zeitschrift „Die Deutsche Schule" zu Wort und kritisierte die TIMSS-Naturwissenschafts-Aufgaben. Er stellte unter der Fragestellung „Was wurde bei TIMSS erhoben?" fachliche Fehler, Irritationsmomente und andere Probleme dar. Baumert u.a. (2000) setzten einen in scharfem Tonfall gehaltenen Beitrag dagegen. Hagemeisters qualitative Einwände wurden hier aber fast nur mit am quantitativen Paradigma orientierten Argumenten „abgebürstet". Außerdem dem von

[3] In diesem Kapitel werden jene theoretischen Modelle diskutiert, die bei TIMSS **und** PISA verwendet wurden, um die Testaufgaben zu beurteilen. Ich setze mich hier explizit mit den Argumenten der TIMSS-Gruppe auseinander, weil nur in der im folgenden beschriebenen Debatte zwischen Hagemeister (1999) und der TIMSS-Gruppe (Baumert u.a. 2000) eine ausführlichere Auseinandersetzung mit den Aufgaben stattfand. Der für PISA erweiterte Theoriekorpus wird in Kapitel 5 diskutiert.

[4] Das Konstrukt der „Fähigkeitsniveaus" ist eine Vorstufe des „Kompetenzstufenmodells", welches bei PISA Verwendung fand und zu dessen Dekonstruktion sich eine Argumentation in Meyerhöfer (2004) findet.

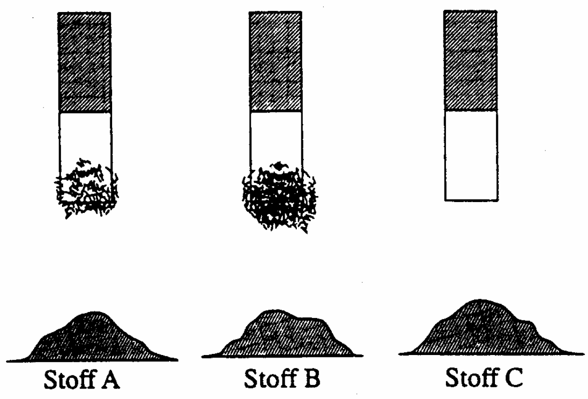
der TIMSS-Gruppe entwickelten Konzept der „Fähigkeitsniveaus” und mit der o.g. Reusserschen Theorie argumentiert.

Ich greife die Debatte zwischen Hagemeister und Baumert u.a. auf, indem ich zunächst eine der dort diskutierten Physikaufgaben interpretiere. Als Fokus der Interpretation dienen dabei noch nicht die Forschungsfragen dieser Arbeit (die ich erst später speziell für Mathematikaufgaben entwickle), sondern lediglich die durch die Debatte vorgegebenen Fragestellungen. In der Interpretation zeigt sich bereits, dass das Konzept der „Fähigkeitsniveaus” die Anforderung der Aufgabe nicht beschreiben kann und eine fruchtbare Debatte über die Eigenschaften dieser Aufgabe nicht ermöglicht.
Es zeigt sich aber auch, dass das von Baumert u.a. vorgestellte mentale Situationsmodell kaum einen Beitrag zur Interpretation der Aufgabe leistet. Dieser Befund leitet die theoretische Auseinandersetzung mit den Möglichkeiten und Grenzen des theoretischen Ansatzes der mentalen Situationsmodelle für Aufgabeninterpretationen ein. Dabei stellt sich heraus, dass Baumert u.a. die Möglichkeiten dieses Ansatzes bei weitem nicht ausschöpfen, dass der Ansatz aber auch schon in seiner Konzeption die latente Textebene vernachlässigt und deshalb für die Interpretation von Testaufgaben weniger aussagekräftig ist als der Ansatz der Objektiven Hermeneutik.

In diesem Kapitel führe ich zunächst eine exemplarische Interpretation der TIMSS-Aufgabe D2 durch, vergleiche sie mit den Methoden von Baumert u.a. und Hagemeister, untersuche dann die Potenzen und die Triftigkeit der Theorie der mentalen Situationsmodelle für Aufgabeninterpretationen und komme danach zur Nutzung der Theorie durch Baumert u.a. für die Interpretation von TIMSS-Aufgaben zurück.

1.1. Die TIMSS-Aufgabe D2: Kurzinterpretation und Diskussion

D2: Jeder der drei abgebildeten Magneten ist in den Stoff unter ihm eingetaucht worden. Welcher Stoff könnte Kaffee sein?



A. Nur A

B. Nur B

C. Nur C

D. Nur A und B

Diese Aufgabe wurde im Streit zwischen Hagemeister (1999) und Baumert u.a. (2000) exemplarisch diskutiert. Dies ist in der sehr breit geführten Debatte um TIMSS der erste (und bis zur noch nicht veröffentlichten Promotion von Reinhard Woschek einzige) Streit, in dem sich explizit, ausführlich und mit öffentlicher Resonanz mit den Testitems auseinandergesetzt wird. Da ferner die Positionen sehr polarisiert und auf eine Pro-Contra-TIMSS-Dualität ausgerichtet sind, eignen sich die Argumentationslinien beider „Seiten", um die Leistungsfähigkeit des von mir verfolgten methodischen Vorgehens zu illustrieren. Ich stelle kurz beide Positionen dar, interpretiere dann die vorliegende Aufgabe und zeige, inwiefern meine Methode die vorgebrachten Argumente stützt, verwirft oder auch besser verstehen lässt.

Hagemeister (1999) wendet gegen diese Aufgabe ein (S.163-165):

- Wenn unter A und B ein Stoffgemisch lag, z.B. Kaffee und Eisenpulver, dann klebt jetzt das Eisen an den Magneten und Kaffee liegt unter A und B.
- Zur Entwicklung der Tests wurden keine Experimente durchgeführt. „Man hätte beim Experimentieren nicht nur bemerkt, dass ein Häufchen Eisenspäne ... an den heute üblichen Dauermagneten vollständig hängenbleibt. Man hätte außerdem z.B. bemerkt, dass die Magnete in dem Stoff, in den sie „eingetaucht worden" sind, Abdrücke hinterlassen hätten." Er referiert auf sorgfältiges Beobachten durch Schüler.
- Es wird lediglich Textverständnis getestet. Ein Schüler: „Naja, wenn man so fragt, dann kann eigentlich „Nur C" richtig sein."
- elaborierte Physikkenntnisse hindern daran, die Lösung schnell zu finden
- fehlende Multiple-Choice-Routine ist ein Handicap. Die Schüler „haben nach einem tieferen Sinn gesucht, weil sie es nicht gewohnt sind, in einer Physikarbeit mit fehlerhaften Texten oder Zeichnungen konfrontiert zu werden."
- Mädchen haben schlechter abgeschnitten, weil sie weniger bereit waren, Halbwissen einzubringen und weil sie häufiger an Ungenauigkeiten der Aufgabenstellungen hängen blieben.

Baumert u.a. (2000) halten dagegen, dass in einer Studie mit der „Methode des stimulated recall" überprüft worden sei, „ob die Aufgabe tatsächlich zur Konstruktion des intendierten subjektiven Situationsmodells auf Schülerseite führt." (S.199) Sie beschreiben das vom Schüler zu konstruierende Modell und ordnen die Aufgabe in das von ihnen konzipierte Konzept der „Fähigkeitsniveaus" ein.

Sie kritisieren Hagemeisters ästhetische und auf fachdidaktisch vorbildlichen Unterricht ausgerichteten Ansprüche an Testaufgaben.

Sie argumentieren, dass die Aufgabe trennscharf ist und dass die von Hagemeister kritisierten Irritationen statistisch nicht nachweisbar sind.

1.1.1. Kurzinterpretation

Ich werde im folgenden eine verkürzte objektiv-hermeneutische Interpretation der Aufgabe darstellen:

Jeder der drei abgebildeten Magnete ...
Hier zeigt sich die Notwendigkeit, zunächst die Abbildung kurz zu interpretieren:

Bildinterpretation (verkürzt):[5]
Man sieht drei schwarz-weiße Rechtecke über drei Stoffhaufen, die mit Stoff A, Stoff B und Stoff C bezeichnet und die alle drei etwa gleich groß sind. Die unterschiedliche Bezeichnung deutet darauf hin, dass es sich um drei verschiedene Stoffe handelt.
Die drei Magnete (Textinformation zu den Rechtecken) befinden sich mittig über den Haufen. Sie sind optisch auf die Haufen orientiert. Am unteren - also ebenfalls „dem Haufen zugewandten" - Ende des linken Magneten (über Stoff A) befinden sich schwarze Punkte (klebt Stoff?), am mittleren Magneten (über Stoff B) befinden sich mehr schwarze Punkte. Am rechten Magneten (über Stoff C) befindet sich nichts.

Jeder der drei abgebildeten Magneten ...
Es wird eine (positive) Aussage über alle drei schematisch dargestellten Magnete getroffen, die auch gleichzeitig als Magnete eingeführt werden - aus der Zeichnung selbst sind sie nicht eindeutig als solche erkennbar. Die Benennung der Anzahl *drei* Magnete verstärkt diese Erkennbarkeit. Es wird also sowohl auf die Zeichnung verwiesen als auch die Zeichnung schon erläutert.
Die Magnete werden trotz dieser Zusammenführung in ihrer je separaten Besonderheit hervorgehoben, denn es heißt nicht nur *Jeder abgebildete Magnet ...,* sondern *Jeder* ***der*** *drei abgebildeten Magnete ...*
Das verweist auf das Vorhandensein dreier „Probleme", die aber auch zusammengehören. In diesem Sinne wird hier nahegelegt, dass es sich um drei identische Magnete handelt.
Mit *abgebildet* wird die Zeichnung charakterisiert. Da Hagemeister den Vorwurf erhebt, dass das Experiment nicht durchgeführt und wiedergegeben wurde, ist zu klären, ob mit *abgebildet* der Status des Experiments als reales oder Gedankenexperiment gekennzeichnet wird. Dazu kontrastiere ich *abgebildet* mit *gezeichnet, gezeigt, dargestellt, skizziert*:
Abgebildet verweist nicht explizit auf einen Anspruch eines wirklichkeitsnahen Abbildes, bei etwas *Gezeichnetem* oder *Gezeigtem* könnte man z.B. eher erwarten, dass dann auch Abdrücke in den Stoffhaufen eingezeichnet sind. *Abgebildet* verweist aber auch nicht explizit darauf, dass hier nur eine grobe Vereinfachung der realen Gegebenheiten vorliegt, *skizziert* würde dies als einziges Verb ohne nochmalige Attributierung fassen.
Aber selbst durch die Verwendung von s*kizzierte Magnete* ließe sich im Sinne eines Ernstnehmens der realen Situation das Nichtvorhandensein von Eindrücken in den Stoffhaufen nicht „heilen". Der Verzicht auf die Darstellung des Eindrucks in Verbindung mit der Unmöglichkeit des sprachlichen Hei-

[5] Ich gebe nur eine verkürzte Bildinterpretation wieder und diskutiere auch das Zusammenspiel von Bild und Text nicht näher, da die vorliegende Kurzinterpretation vorrangig dazu dient, die weitere Argumentation greifbarer zu machen.

lens dieses Defizits konstruiert hier also ein Gedankenexperiment statt eines realen Experiments. Dies steht zunächst im Widerspruch zum manifesten Text, der ja eine reale Experimentiersituation behauptet. Es verbleibt aber immer noch die Möglichkeit, genau diese Nichtrealität des Experiments, also seinen Charakter als Gedankenexperiment, weiter auszubauen und den Status des Experiments damit eindeutig zu machen.

... ist in den Stoff unter ihm eingetaucht worden.
Der Magnet wird *in den Stoff* eingetaucht, also nicht nur an der Oberfläche entlanggestreift. *Den* verstärkt die Zuordnung des jeweiligen Magneten zu dem unter ihm liegenden Stoffhaufen.

Eingetaucht: Es erfolgt kein Verweis darauf, dass noch etwas gemacht wurde, z.B. gerührt oder gepustet bzw. dass etwas am Magneten hängen blieb.
Eingetaucht ist stärker, auch klarer als *getaucht*, hier wird die Irritation durch die in der Abbildung fehlende Tauchspur vergrößert, zumal die Konstruktion der Realität des Experiments durch das *ist ... worden* ja weitergeführt wird. *Eingetaucht* verrät aber auch etwas über die Spezifik des Stoffes: Da es sich - wie die Abbildung zeigt - nicht um eine Flüssigkeit handelt, muß es sich um einen körnigen bzw. pulverigen Stoff handeln.

Er ist in den Stoff *unter ihm* eingetaucht worden. Dies vereindeutigt das experimentelle Vorgehen und die Zuordnung Haufen - Magnet. Das wäre an sich trivial. Latent stellt das aber die Verbindung zwischen dem Stoffhaufen und dem an dem Magneten hängenden Stoff her. Die optische Verbindung zwischen Magnet und Stoffhaufen durch die Ausrichtung des Magneten und durch die Platzierung des an ihm „festklebenden" Stoffes schafft in Verbindung mit dem *in **den** Stoff* und dem *unter ihm* den Zusammenhang zwischen Stoffhaufen und festklebendem Stoff. Hier gibt der Text (inklusive Bild) ein für mich überraschend starkes Signal, dass es sich **nicht** um die Trennung von Stoffgemischen handelt. Inhaltlich wird das dadurch gestützt, dass die Magnete nur eingetaucht, nicht aber im Gemisch bewegt wurden.

Der Magnet *ist* eingetaucht *worden*. Das Experiment liegt nicht so weit zurück wie bei *wurde*. Die experimentellen Folgen sind also noch unmittelbar. Hier wird auf ein **vergangenes, reales** Handeln verwiesen. Der Charakter als reales Experiment wird dadurch gestärkt.

Es handelt sich um eine Passivkonstruktion, die aber nicht personalisiert ist, das verweist auf Wiederholbarkeit. Das Handeln war gerichtet, er *ist* ja z.B. nicht *reingefallen*. Diese Konstruktion betont also den artifiziellen und wissenschaftlich-experimentellen Charakter der Aufgabe. Der Indikativ verweist auf den realen Charakter des Experiments, es ist also kein Gedankenexperiment.

Welcher Stoff ...
Hier deutet sich eine sehr eng geführte Frage an: Man wird keine Zusammenhänge zu erläutern haben, man wird sich nicht über Stoffgruppen auslassen. *Stoff* wird hier thematisch, und zwar genau einer, und der wird punktgenau zu benennen sein: *Welcher*?

Um die Bedeutung des nachfolgenden Fragenteils zu erarbeiten, soll zunächst überlegt werden, wie es wohlgeformt weitergehen könnte:
... ist (nicht) Eisen?

... ist (k)ein Stoffgemisch?

... ist (nicht) magnetisch?

... ist (nicht) Kaffee?

Man sieht schon, dass jede einzelne Frage ein Spektrum an (fachlichem, also eigentlich unterrichtlichem) Diskussions- bzw. Experimentierstoff aufwirft. Hier soll aber lediglich die Bedeutung der gestellten Frage als Testaufgabe diskutiert werden. Dabei fällt zunächst auf, dass nicht *ist* gefragt wird, wie es die bis hier enge Frageführung wohlgeformt erfordern würde. Hier wird *könnte ... sein* gefragt. Schon ein *kann ... sein* hätte die enge Frageführung aufgebrochen. *Könnte ... sein* ist ein noch weiteres Aufbrechen. Dieser Vergleich zeigt, dass an dieser Stelle die bis hier relativ klare Linienführung der Aufgabe verlassen wird. *Könnte ... sein* signalisiert - im Vergleich -, dass man zweifeln kann, dass man genauer nachdenken muß, vielleicht doch lieber erstmal experimentiert. Der hier gesäte Zweifel ist genau der, der in uns die leise Frage erwachen läßt, ob in Kaffee nicht doch irgendwas Magnetisches drin ist, so dass Stoff A der Kaffee ist: Wer hat das schließlich schon ausprobiert? Hier gilt es, den Zweifel des („zu") exakt denkenden Menschen zu übergehen, und das ist eine Komponente von Testfähigkeit. Oberflächlich betrachtet und ohne Kenntnis der Aufgabeninterpretation ist das an dieser Stelle einfach die Fähigkeit, eine Testsituation von einer Lern- bzw. Forschungssituation zu unterscheiden. Die Interpretation zeigt aber gerade, dass der Aufgabentext auf eine Forschungssituation verweist und damit Wissen auf einer wissenschaftlichen, nicht auf einer alltäglichen Ebene abfragt. Damit wird dann nicht mehr nur die Fähigkeit verlangt, diese Situationen zu unterscheiden, sondern die Ebene zu erfassen, auf der die Testfrage „gemeint" ist.

... *Kaffee* ... Der Vergleich mit den anderen Möglichkeiten der Fortführung der Frage zeigt, dass ein nichtmagnetischer Stoff an dieser Stelle den Anforderungen einer eindeutig beantwortbaren Testaufgabe entgegenkommt:
Eine Aufgabenstellung mit Eisen, erst recht mit einem anderen magnetischen Stoff, würde durch das Vorhandensein der Stoffe A und B zu Nichtbeantwortbarkeit führen. Hier würde das Wissen um verschiedene Stärken von Magnetismus thematisiert.
Die Frage nach dem Stoffgemisch wäre nicht beantwortbar, hier liegt keine Experimentieranordnung vor, um das herauszufinden.
Die Frage, welcher Stoff (nicht) magnetisch ist, würde äußerlich lediglich das Verständnis des Begriffs *magnetisch* prüfen, wäre dafür aufgrund der verschiedenen Stärke des Magnetismus der Fälle A und B wiederum problematisch, weil ja auch noch die je verschiedene Stärke des Magnetismus thematisiert wäre. Man könnte hier das Verständnis des Begriffs magnetisch, das Verständnis des Problems von verschiedener Stärke von Magnetismus und die Fähigkeit, über entsprechende Irritationen hinwegzugehen, nicht sauber voneinander trennen.

Mit diesen Gegenbeispielen wird klarer, was die Nutzung von *Kaffee* für die Anforderungsstruktur der Aufgabe bedeutet: Es wird äußerlich nicht die Bedeutung eines Begriffs getestet, sondern ausschließlich das Wissen darum, dass Kaffee nicht magnetisch ist.[6]
Diese Interpretation kann dabei nicht die Frage beantworten, ob es sinnvoll ist zu wissen, dass Kaffee nicht magnetisch ist - der Tester kann mit dem Wissen um die Textbedeutung an dieser Stelle der Interpretation entscheiden, ob er dieses Wissen testen möchte oder ob er nicht eigentlich etwas anderes testen wollte.

Es ist offensichtlich, dass nahezu alle Menschen unseres Kulturkreises auf einer Alltagsebene wissen, dass Kaffee nicht magnetisch ist: Wenn der Kaffee auf dem Küchentisch auskippt, werde ich nicht versuchen, ihn mit dem Magneten einzusammeln. Die gesamte Aufgabenstellung bis zur Stelle *Welcher Stoff ...* arbeitet aber eindeutig nicht auf der Alltagsebene, sondern referiert auf ein wissenschaftliches Experiment. Das bedeutet nun aber, dass hier gar keine Alltagsfrage, sondern eine wissenschaftliche Fragestellung an den Schüler herangetragen wird. Und wissenschaftlicher Habitus bedeutet gerade, mein Küchentischwissen zu hinterfragen, es z.B. mit extrem starken Magneten zu probieren, mit erhitztem oder gefrorenem Kaffeepulver, mit mehr oder weniger stark gemahlenem Kaffee. Das heißt nicht, dass der Schüler die Frage nicht auf der Alltagsebene beantworten kann, es heißt lediglich, dass es **hier** keine Alltagsfrage **ist**. Denkbar ist an dieser Stelle, dass ein durch Physikunterricht geschulter (und oftmals auch bezüglich seiner Alltagserfahrungen in Frage gestellter und „reingelegter") Schüler stärkere Zweifel hat als z.B. ein Grundschüler. Statistisch dürfte dies aber nicht herauszufinden sein (siehe 1.1.3.). Nichtsdestotrotz ist diese Gruppe benachteiligt, sobald das Testresultat für sie persönlich folgenreich wird.

Die Verwendung von *Kaffee* statt *Schwefel* oder irgendeines nichtmagnetischen *Kunststoffs* verdeckt die eben erarbeitete Erkenntnis und suggeriert, hier würde Alltagswissen getestet. Schon die Verwendung eines der letztgenannten Stoffe in der Aufgabenstellung würde den Gedanken, es handele sich um ein „lebenspraktisches", alltagsnahes Experiment, ferner legen.

[6] Ich verwende hier weiter den Begriff Kaffee, obwohl ja eigentlich Kaffeepulver gemeint ist. Man kann diesbezüglich kaum von Irritationspotential sprechen, da Kaffee sprachlich immer sowohl die Flüssigkeit als auch das Kaffeepulver meinen kann. Die Interpretation zeigt, dass die Zeichnung und die Verwendung des Begriffs *Stoff* auf das Pulver verweisen.

1.1.2. Multiple Choice

Bis hier wurde gezeigt, dass die Aufgabenstellung Zweifel induziert, einerseits durch den Bruch zwischen klarer Problemführung und verunklarender Fragestellung, andererseits durch Verwissenschaftlichung einer Alltagsfrage. Dies wird durch Multiple Choice hier verstärkt, da Multiple Choice jedem Zweifel auch seine Antwortalternative bietet.
Das Angebot *D. Nur A und B* zielt auf Schüler, die den Aufgabentext nicht verstanden haben und als Fragestellung ansehen, welcher der drei Stoffe magnetisch sein könnte. Der Tester muß sich also fragen, ob er so etwas testen möchte. Gleichzeitig liegt die Ratewahrscheinlichkeit für Antwortangebote, die komplizierter aussehen als die anderen, höher. Die Gruppe der Ratenden und die Gruppe der falsch Verstehenden sind hier also nicht voneinander trennbar.

Wie würde sich die Teststruktur ändern, wenn nur die Stoffe B und C vorgegeben wären? Die Testintention wäre damit unverstellt zum Ausdruck gebracht – der eine Stoff ist magnetisch, der andere nicht. Es ist offensichtlich, dass die beschriebenen Zweifel in dieser Variante geringer sein werden. Wenn der Stoffhaufen B auch noch in relevantem Maße kleiner wäre als der Stoffhaufen C, so wären Zweifel wohl noch unwahrscheinlicher. Als Antwortangebote blieben – wenn man aus technischen Gründen weiter vier Angebote machen will – aber lediglich:

A. nur B; B. nur C; C. B und C; D. Weder B noch C;

wovon die letzten beiden offensichtlich nicht mehr sinnvoll zu interpretieren sind. Das TIMSS-Design zwingt zur Konstruktion von 4 Antwortalternativen, lediglich aus technischen Gründen des Multiple Choice würde man also auf diese Aufgabenvariante verzichten, die die Testintention unverstellt aufnimmt.

1.1.3. Zusammenfassung und Einordnung in die Debatte von Hagemeister und Baumert u.a.

1. Die Aufgabe birgt Brüche, Resultate verlieren dadurch an Interpretierbarkeit, insbesondere trennt die Aufgabe nicht sicher zwischen Wissenden und Nichtwissenden; Kritik des Fähigkeitsniveaukonzepts

Mit dieser Aufgabe wird inhaltlich das Wissen darum, dass Kaffeepulver nicht magnetisch ist, getestet. Die Fragestellung wird nicht auf einer Alltags-, sondern auf einer wissenschaftlichen Ebene betrachtet. Die Experimentiersituation wird mehrfach als wissenschaftlich-experimentell gekennzeichnet (*... ist eingetaucht worden ...*, *Jeder der drei ...*, Bildkomposition). Der Tester muß hier also entscheiden, ob der Zweck seiner Fragestellung an den Schüler das Abtesten dieser Information auf dieser Ebene ist. Der Aussage von Baumert u.a., es handele sich hier um das Abtesten „lebenspraktischen Wissens”, ist jedenfalls zu widersprechen. Die vermeintlich lebenspraktische Verwendung von Kaffee verdeckt das hier allerdings.

Text und Aufgabe führen zunächst klar und strukturiert in das Problem ein. Bis zur Stelle *Welcher Stoff ...* liegt eine enggeführte Aufgabenstellung vor. Mit dem *könnte Kaffee sein?* wird diese Engfüh-

rung aufgebrochen. Dieser Bruch, der Zweifel aufruft, ist ein doppelter: Er liegt einerseits sprachlich im plötzlichen Aufbruch der Engführung in ein Vages hinein. Er liegt andererseits inhaltlich in der Forderung, eine Aussage über die Laboreigenschaften eines Stoffes zu treffen, den man nur aus dem Alltag kennt. Der hier aufgerufene Zweifel kommt im Labor bzw. im Klassenraum auf und nicht am Küchentisch, wo der Kaffee ausgeschüttet ist. Dieser Zweifel wird in die Testsituation getragen. Die Frage ist dann, ob in Kaffee nicht doch irgendwas Magnetisches drin ist, so dass Stoff A der Kaffee ist: Wer hat das schon ausprobiert, noch dazu mit einem genügend starken Magneten? Unterricht und Wissenschaft lassen uns oft genug erleben, dass man anderes sieht, wenn man genauer hinsieht oder das Experiment verfeinert. Hier gilt es in der Testsituation, den Zweifel des („zu”) exakt denkenden Menschen zu übergehen, und das ist eine Komponente von Testfähigkeit. In der Hagemeister-Arbeit findet sich die eben besprochene Irritation in der Bemerkung einer Schülerin: „Leider weiß ich nicht, ob Kaffee Eisen enthält.” Das ist genau jene wissenschaftliche oder Schulfrage, die man aufgrund der Interpretation erwarten kann; es ist nicht die Frage, die man sich im Alltag stellt, wenn man darüber nachdenkt, ob man den ausgeschütteten Kaffee mit einem Magneten wieder einsammeln kann.
Man kann anhand des Ergebnisses nicht mehr trennen, wer über den Zweifel hinweggegangen ist, wer ihn gar nicht hatte, weil er nicht weit genug dachte, und wer es wirklich ganz genau weiß. Die hier fehlende Trennschärfe ist noch nicht problematisch, denn der tiefer Denkende löst das Problem ja ebenso wie der weniger tief Denkende. Problematisch wird es, weil die Zweiflerin, die sich auf wissenschaftliches Denken einlässt, eben auch zum Nichtlösen oder zur Antwort A gelangen kann - obwohl sie weiß, dass sie den ausgeschütteten Kaffee nicht mit dem Magneten einsammeln kann. Und diese Zweiflerin ist nicht „leistungsschwächer” als derjenige, der das Ganze auf der Küchentischebene beläßt. Sie erbringt eine Leistung in der Auseinandersetzung mit dem Problem und sie verfehlt die Erkenntnis, dass eine andere Ebene der Aufgabenbewältigung gefragt ist. Hier wird auch deutlich, dass die Argumentation von Baumert u.a. - die für diese Aufgabe die „lebenspraktische” Ebene beanspruchen - ein Stück weit darauf hinausläuft, erfolgreich zu antizipieren bzw. zu erraten, was die Tester gerade wollen, ohne es zu sagen.

An dieser Stelle klappt auch das „Gegenrechnen” der TIMSS-Gruppe nicht, die schreibt:

> „Hagemeister vermutet, dass elaborierte Physikkenntnisse zur Konstruktion eines komplexeren, aber für die Aufgabe unangemessenen mentalen Situationsmodells verleiten könnte, gerade weil offen bleibt, um welche Substanzen es sich bei den Alternativen A und B handele. Ähnliches nimmt er für Personen an, in deren Unterricht regelmäßig experimentiert wird und die möglicherweise mit der Trennung von eisenpulverhaltigen Gemischen durch Magnete vertraut sind. Wenn die beiden Einwände stimmen, muß sich dies in einer unbefriedigenden Trennschärfe des Items und positiven biserialen Korrelationen zwischen Distraktoren und Gesamttestwert niederschlagen. Ferner muß sich für Schüler, die einen experimentell ausgerichteten Unterricht genießen, unter Konstanthaltung der sonstigen Testleistung eine höhere Itemschwierigkeit nachweisen lassen. Von beidem kann jedoch keine Rede sein.” (Baumert u.a. 2000, S.200)

> Gegen diese Argumentation ist einzuwenden: Nach Rechnung von Baumert u.a. „differenziert [die Aufgabe] ... zwischen der untersten Kompetenzstufe, die wir als lebenspraktisches Wissen bezeichnet haben, und der zweiten Stufe, die ein Denken in alltagsbezogenen vorwissenschaftlichen Konzepten indiziert". (S.199, auch 200)[7] Die Analyse zeigt aber (wenn man alle Irritationsmomente wegdenkt), dass hier lediglich zwischen dem Vorhandensein und Nichtvorhandensein von Wissen getrennt wird. Vereinfacht gesagt: Wer weiß oder ahnt, dass Kaffee keinesfalls magnetisch ist, der löst die Aufgabe richtig, wer es nicht weiß oder ahnt, der löst sie falsch. Ob damit schon ein Konzept vorliegt, ist fraglich.

Nach vorstehender Interpretation prüft die Aufgabe das Wissen entweder auf Alltagsniveau – das wäre dann die niedrigere postulierte Niveaustufe („Lebenspraktisches Wissen") – oder auf einem wissenschaftlichen Niveau, wo es aber mit der höchsten Niveaustufe IV („Konzeptuelles Verständnis der Schulphysik") wiederum auch nicht korrekt beschrieben ist. Die Interpretation zeigt, dass für den naturwissenschaftlich habitualisierten Schüler ein Faktenwissenproblem vorliegt. Hier zeigt sich jedenfalls die Inhaltsleere des Fähigkeitsniveaukonzepts, denn nicht der Punktwert der Aufgabe erzählt uns etwas darüber, was hier erfaßt wird, ob hier etwas Lebenspraktisches oder Alltagsbezogenes oder Vorwissenschaftliches oder Wissenschaftliches erfragt wird, sondern nur die eingehende fachdidaktische Analyse antwortet darauf.

(1) Warum sollten gerade die „exakten Denker" - noch dazu in statistisch relevanter Größenordnung - diejenigen sein, die ihre Zweifel nicht auch übergehen können? Und wie will man die Zweifler rechnerisch von denen trennen, die die Antwort auf einer Alltagsebene nicht wußten? An dieser Stelle zeigt die genaue Betrachtung der sich aus der Aufgabe ergebenden Möglichkeiten, dass es gerade nicht möglich ist, die beschriebenen Effekte herauszurechnen. Der Zweifler ist vom Nichtwissenden nicht zu trennen, der eher oberflächliche Halbwisser und der „Alltagswisser" nicht von jenem, der es ganz genau weiß.

(2) Hinzu kommt ein weiteres, fundamentales Mißverständnis: Warum sollten denn Schüler mit elaborierten Physikkenntnissen irritiert werden? Sprachregeln gelten und wirken für alle Schüler. Die beschriebenen Brüche treten ebenfalls für alle Schüler auf. Man kann vielleicht annehmen, dass jemand, der nicht über elaborierte Physikkenntnisse verfügt, eher über die latente Aufforderung zu wissenschaftlichem Denken hinweggeht. Man kann aber ebenso annehmen, dass er sie aufnimmt und dann aufgibt. In jedem Fall wird auch diesem Schüler Testfähigkeit helfen, über diese Irritation hinwegzugehen.

[7] TIMSS arbeitet mit einem „Fähigkeitsmodell" (siehe Baumert, Lehmann u.a. 1997, S.78 ff. oder Baumert u.a. 2000, S.108 ff.). Dabei wird die gemeinsame Rasch-Skala der Aufgabenpunktwerte und der Schülerpunktwerte in Abschnitte unterteilt. Den Abschnitten werden dann (indem man sich die in den Abschnitten liegenden Aufgaben ansieht) Aufgabenanforderungen und damit Schülerfähigkeiten zugeordnet. Erst mit diesem Konstrukt werden die Testresultate inhaltlich deutbar. Bereits bei D2 zeigt sich, dass die Zuschreibung nicht funktioniert, weil die „gespaltene" Anforderung nicht beachtet wird. Eine ausführliche Argumentation zur Unmöglichkeit solcher Konstrukte führe ich am Beispiel des Kompetenzstufenmodells von PISA (Meyerhöfer 2004).

(3) Die Eindimensionalität und Inadäquanz des Glaubens, man könne die Häufigkeit von Demonstrations- und Schülerexperimenten mit der Itemschwierigkeit verrechnen und würde dabei irgend eine interpretierbare Aussage erhalten, ist mit dem Gesagten bereits hinreichend belegt.

2. Der fachliche Inhalt wird nicht ernst genommen

Die Abbildung zeigt eine Situation, wie sie nach Durchführung des beschriebenen Experiments nicht gegeben sein kann, weil Abdrücke der Magneten in den Stoffhaufen fehlen. Dieses Defizit ist durch den Text nicht geheilt. Der (manifeste) Inhalt der Aufgabe wird also nicht ernst genommen. Die Verwendung von *eingetaucht* verstärkt diesen Aspekt. Aber auch das bereits beschriebene Springen zwischen wissenschaftlicher und Alltagsbetrachtung des Problems verweist darauf, dass man es nicht wirklich ernst nehmen muß: nicht als wissenschaftliches, aber auch nicht als Alltagsproblem.
Das Experiment ist sprachlich als reales Experiment konstruiert. Eine Konstruktion als Gedankenexperiment könnte den fachlichen Inhalt ernst nehmen und konsequent Wissenschaftlichkeit verfolgen. Denkbar wäre folgender Aufgabentext: *Stell dir vor, jeder der zwei dargestellten Magneten sei in den Stoff unter ihm eingetaucht worden und einer der Stoffe ist Kaffee. Welcher Stoff ist dann Kaffee?*
Der Bruch zur angeblichen Alltäglichkeit ist dann aber separat zu überdenken. Will man Alltagswissen testen, kann man z.B. eine Küchentischsituation aufnehmen.

3. Der Text legt nicht nahe, dass es sich um Stoffgemische handelt.

Hagemeister berichtet von zwei Schülern, die mit einem Magneten Schwefelpulver und Eisen getrennt haben. Sie vermuten unter A und B Kaffeepulver, welches vom am Magneten haftenden Eisen getrennt wurde. Diese Sichtweise wird durch den Text (Bildkomposition, *den*, *unter ihm*, *eingetaucht*) latent wie manifest abgewehrt. Man könnte annehmen, dass sie durch die Gesprächssituation oder durch die beschriebenen, zu Zweifeln führenden, Irritationen herbeigeführt wurde.

4. Hagemeister vernachlässigt den Zweck einer Testaufgabe

Hagemeisters verständlicher Wunsch nach ausführlicher Debatte des Kaffeeproblems wird dem Zweck einer Testaufgabe nicht gerecht. Hier geht es ja um das Erfassen einer Fähigkeit und nicht um „Einführung in wissenschaftliche Arbeitsmethoden”, „Achtung vor der Natur”, „schön aussehen” usw. Man kann einer Testaufgabe nicht vorwerfen, dass sie eine Testaufgabe ist. Man kann aber fragen, ob sie testet, was sie testen soll. Diese Aufgabe tut das nur zum Teil.

1.1.4. Vergleich der Interpretation unter Nutzung von Objektiver Hermeneutik mit dem Ansatz des „mentalen Situationsmodells" bei Baumert u.a.

> „Welches mentale Situationsmodell muß vom Schüler konstruiert werden, um die Aufgabe richtig zu lösen? Der subjektive Problemraum wird
> (1) durch die modellhafte Skizze der Versuchsanordnung, die
> (2) im Itemstamm verbal erläutert wird, und
>
> (3) die gezielte Frage des Aufgabenstammes „Welcher Stoff könnte Kaffee sein?" (wohlgemerkt: nicht „enthalten") vorgezeichnet.
>
> Die Aufgabe verlangt die Aktivierung einer einzigen Wissenseinheit, dass Magnete Nichtmetalle nicht anziehen. Dieser Fall wird in der Lösungsalternative C durch den blanken Magneten symbolisiert - obwohl beim Eintauchen eines Magneten in Kaffeepulver Kaffeereste auch am Magneten hängen bleiben können. [Das wäre also in Wirklichkeit Lösung A. W.M.] Dies ist jedoch – ebenso wie die Abdrücke, die der Magnet nach Hagemeister im jeweiligen Stoff hinterlassen haben müsste – für das zur Lösung erforderliche Situationsmodell unerheblich. Ein angemessenes Situationsmodell ist gerade kein photographisches Abbild einer Versuchsanordnung, sondern ein konzeptueller Rahmen, in dem Situation, Aufgabenstellung und das zur Lösung der Aufgabe notwendige Wissen zusammengebunden werden.
>
> Um welche Substanzen es sich im Fall A und B handelt, lässt die Aufgabe offen, da die Beantwortung dieser Frage für die Entwicklung eines adäquaten Situationsmodells ohne Belang ist. Im Vergleich zum Fall C wird jedoch klar, dass es sich nicht (nur) um Kaffee handeln kann. Zur Lösung der Aufgabe wird also nur rudimentäres Wissen über Magnetismus verlangt; eine genaue Kenntnis ferromagnetischer Materialien ist nicht erforderlich." (Baumert u.a. 2000, S.199 f.)

Hier werden grundsätzlich unterschiedliche Fragestellungen deutlich: In der objektiv-hermeneutischen Interpretation geht es darum, die Sinnstruktur der Aufgabe zu erfassen und dadurch die Frage zu beantworten, was sie mißt. Der Text wird also daraufhin untersucht, was er sagt. Daraus zieht man dann Schlußfolgerungen darauf, was der Text beim Schüler auslösen könnte. Wenn man also herausarbeitet, welche Wissenseinheit abgefragt wird, dann schließt man daraus, dass diese Abfrage auf den Rezipienten wirken kann. Wenn mehrfach ein Bruch im Aufgabentext identifiziert wird, dann schließt man daraus, dass dieser Bruch auf den Textrezipienten wirken kann. Die Intention des Texterstellers kann ebenfalls identifiziert werden, man kann sie aber methodisch sauber von der realisierten Textstruktur trennen. So erhält man ebenso wie jede fachdidaktische Aufgabenanalyse eine Aussage darüber, dass die Aufgabe offenbar das Wissen darum testen soll, dass Kaffee nichtmagnetisch ist.[8]

[8] Die objektiv-hermeneutische Textinterpretation kann die Intention des Testerstellers allerdings nicht mit Sicherheit vollständig bestimmen: Einerseits gibt es Testaufgaben, die mehrere Lösungen denkbar erscheinen lassen, z.B. TIMSS A4. Hier zeigt erst das Lösungsschema, welche Lösung die Tester wirklich anstreben.
Andererseits ist z.B. in der hier interpretierten Aufgabe D2 aus der Interpretation heraus nicht benennbar, ob die Testersteller die Fähigkeit mittesten wollten, eine Experimentierabbildung in eine Vorstellung vom Experiment zu übersetzen. Die Interpretation zeigt, dass diese Fähigkeit mitgetestet wird, kann aber nicht identifizieren, ob die Tester das beabsichtigten.

Beim Ansatz des mentalen Situationsmodells wird ein Unterschied zwischen der Intention und der Sinnstruktur des Aufgabentexts nicht thematisiert und somit in Abrede gestellt. Da für die Objektive Hermeneutik die Unterscheidung eines gemeinten oder intendierten Sinns und einer objektiven Sinnstruktur eines Textes grundlegend ist, liegt an dieser Stelle der zentrale theoretische Unterschied.

Baumert u.a. stellen die Frage, welches Modell vom Schüler konstruiert werden muß, um die Aufgabe richtig zu lösen. Die Antwort darauf hat zwei Ebenen, die fachdidaktische und die des mentalen Situationsmodells.

Auf der **fachdidaktischen Ebene** stellen Baumert u.a. fest: „Die Aufgabe verlangt die Aktivierung einer einzigen Wissenseinheit, dass Magnete Nichtmetalle nicht anziehen.” Ich stelle im Gegensatz dazu fest, dass das Wissen, dass Kaffee nicht durch Magnete angezogen wird, aktiviert werden muß. Dieser Interpretationsunterschied ist ein rein physikdidaktischer und hat nichts mit unseren unterschiedlichen Methoden zu tun.
Baumert u.a. schließen: „Zur Lösung der Aufgabe wird also nur rudimentäres Wissen über Magnetismus verlangt; eine genaue Kenntnis ferromagnetischer Materialien ist nicht erforderlich.”
Ich komme zu einem anderen Ergebnis: Die Aufgabe verlangt zunächst eine genaue Kenntnis ferromagnetischer Materialien, sie verlangt nämlich auf einem wissenschaftlich abgesicherten Niveau das Wissen darum, dass ein spezieller (und eher selten auf Magnetismus hin betrachteter) Stoff nicht magnetisch ist. Da dieser Stoff nun aber gerade Kaffee ist, wird dieser eigentlich sehr hohe Anspruch verschleiert und damit in gewisser Weise reduziert – aber nur, solange Alltagswissen und wissenschaftliche Erkenntnis zusammenlaufen. Man sieht also, dass hier ein anderer Aufgabencharakter herausgearbeitet wird, der aber eben auch die Herkunft einiger Einwände und Irritationen von Hagemeister erklärt. Baumert u.a. erfassen hingegen mit ihrem Vorgehen nicht, was die Aufgabe mißt, sondern lediglich, was sie messen soll. Dieser Unterschied in der Interpretation rührt also nicht mehr aus unterschiedlichen physikdidaktischen Interpretationen her, sondern fußt in den unterschiedlichen Methoden der Aufgabeninterpretation.

Auf der **Ebene des mentalen Situationsmodells** werden nur triviale Aussagen generiert: Wenn der Schüler die Aufgabe lösen will, dann muß er zunächst einen subjektiven Problemraum erarbeiten. Dazu muß er die Zeichnung und die zugehörige Erläuterung sowie die Frage verstehen. Um die Frage zu beantworten, muß er wissen, dass Magnete Nichtmetalle nicht anziehen.

Auch meine Interpretation trifft ja zunächst keine nähere Aussage dazu, welchen subjektiven Problemraum der Schüler entwirft. Das ist aber auch nicht wichtig, wenn man fragt, was eine Aufgabe mißt. Ich bespreche dieses Problem ausführlich in Kapitel 2, hier nur soviel: Der objektive Aufgabentext gibt dem Schüler vor, was er zu tun hat. Dieser Text kann mit der Intention des Testerstellers zusammenlaufen oder nicht. Die objektiv-hermeneutische Interpretation arbeitet den objektiven Aufgabentext und die Intention des Testerstellers heraus und findet dabei auch heraus, ob beide zusammenlaufen. Sie beantwortet also die Frage, was die Aufgabe messen soll, aber sie antwortet auch auf die wichtigere Frage, was sie mißt. Sie kann davon ausgehend Aussagen über die Verbesserung der Aufgabe treffen: Für jede problematische Stelle ist benennbar, welche Textveränderung zu welcher Veränderung der Sinnstruktur der Aufgabe führen würde. Dabei kann sich natürlich herausstellen, dass die vorliegende Aufgabe nicht so verändert werden kann, dass sie mißt, was sie messen soll.

Nun ist klar, dass der objektive Aufgabentext bei jedem Individuum auf spezifische kognitive Gegebenheiten trifft und der subjektive Problemraum jedes Individuums ebenso spezifisch ist. Diese Spezifik – einschließlich Pathologien – kann die Objektive Hermeneutik natürlich nicht vorhersagen, aber das kann der Ansatz der mentalen Situationsmodelle auch nicht, auch hier wird eine objektiv durch die Aufgabe gestellte Anforderung vorausgesetzt.

In meinen ausführlicheren Interpretationen von Mathematikaufgaben stelle ich vor die objektiv-hermeneutische Analyse eine Darstellung der möglichen Lösungswege. Für die hier vorliegende Aufgabe würde das bedeuten, verschiedenste Argumentationswege zur Beantwortung der Frage zu generieren. Die objektiv-hermeneutische Analyse kann dann neben den anderen von ihr zu beantwortenden Fragen eruieren, welcher der Lösungswege durch den Aufgabentext latent oder manifest nahegelegt wird. Die Argumentation von Baumert u.a. leistet aber nicht mal diese Offenlegung verschiedener Lösungswege.[9] Die einzige Erklärung ist wiederum, dass die Autoren ausschließlich den Testersteller und seine Intention in den Blick nehmen.

Die Fokussierung auf den Wunsch des Testers[10] zeigt sich auch in der Argumentation zum Ausschluß von A und B: „Um welche Substanzen es sich im Fall A und B handelt, lässt die Aufgabe offen, da die Beantwortung dieser Frage für die Entwicklung des adäquaten Situationsmodells ohne Belang ist.” So kann man nicht argumentieren, wenn man den Schüler im Blick hat, der das Modell erst aufbauen muß, um damit dann die Frage zu beantworten. Für ihn stellt sich ja die Frage, ob Kaffee Eisen enthält, erst im Verlauf des Aufbaus des Situationsmodells bzw. nach dessen Aufbau.

Mit der Fokussierung auf den Wunsch des Testers wäre auch zu erklären, dass Baumert u.a. das Fehlen von Abdrücken der Magnete mit dem Argument verteidigen, dass die Abdrücke „für das zur Lösung erforderliche Situationsmodell unerheblich” seien. Man erhält den Eindruck, dass die Experimentiersituation hier nicht ernst genommen wird. Aus der Aufgabeninterpretation heraus wird aber deutlich, dass hier lediglich die latent in der Aufgabenstellung transportierte Logik des Gedankenexperiments vertreten wird. Baumert u.a. können das allerdings nicht so benennen, weil manifest von einem Realexperiment die Rede ist. Genau diese Struktur nimmt auch Hagemeister auf und verweist auf den Schüler, der sich ernsthaft auf die Experimentiersituation einläßt und für den Abdrücke in den Stoffhaufen entstehen. Im Kopf entstehen diese Abdrücke eher nicht.

[9] So gelangt man selbst beim oberflächlichen Versuch, ein mentales Situationsmodell zu entwerfen, zu der viel näher liegenden Fragestellung: Welcher Stoff könnte Eisen sein, woauf man mit A oder B antwortet. Kaffee wäre dann C – sozusagen als nichtmagnetisches Gegenstück.

[10] Diese Fokussierung ist vorrangig ein Problem der Gültigkeit der Interpretation der Testergebnisse, weniger der Beeinflussung der Testergebnisse. Mit der Notwendigkeit, die Intention des Aufgabenstellers zu erraten, liegt ein schulischer, den Schülern nicht fremder Habitus vor.

1.1.5. Vergleich der Interpretation unter Nutzung von Objektiver Hermeneutik mit dem Vorgehen von Hagemeister

Hagemeister stellt die Frage, was die TIMSS-Aufgaben messen bzw. ob sie das messen, was sie messen sollen. Er interpretiert die Aufgaben ohne kontrollierbares methodisches Vorgehen, er schaut einfach drauf. Er hat dann in einer Runde mit 3 Schülerinnen und 4 Schülern TIMSS-Aufgaben bearbeiten lassen. „Erst, wenn eine Antwort aufgeschrieben oder angekreuzt war, haben wir uns über die Aufgabe unterhalten". (Hagemeister 1999, S.161)

Es ist klar, dass bei diesem Vorgehen davon auszugehen ist, dass der Forscher den Probanden Fragen und Interpretationsweisen suggeriert, dass sie aber auch in besonderem Maße bereit sind, auftretende Irritationen offen zu legen. Beides zeigt sich in Hagemeisters Berichten.

Ein Vergleich zwischen Hagemeisters Bericht und meiner Interpretation deutet an, in welch starkem Maße die z.T. ja nur latent im Text auftretenden Irritationen wirken. Dies zeigt sich im positiven Sinne an der Bemerkung „Leider weiß ich nicht, ob Kaffee Eisen enthält". Dies zeigt sich aber besonders scharf im negativen Sinne, also darin, welche Irritationen alle **nicht** aufgetreten sind, obwohl sie sachlich durchaus nahe liegen, aber durch den Text latent verworfen werden: Hagemeister ebenso wie seine Schüler diskutieren nicht, dass jeder der drei Magnete verschieden stark sein könnte - hier wirkt das latente Signal des *Jeder der drei ...* zusammen mit dem manifesten Signal der Abbildung dahin, dass drei identische Magnete vorliegen. Es wird ebenfalls nicht diskutiert, dass die Magnete zu verschiedenen Zeitpunkten eingetaucht worden sein könnten und deshalb im Laufe der Zeit verschieden viel von dem Stoff wieder runtergefallen ist. Hier wirkt das Konstrukt der Laborsituation und das zeitliche Signal des *... eingetaucht worden*. Dieses Signal wirkt gemeinsam mit dem *Jeder der drei ...* auch gegen den denkbaren Fall, dass die Magnete vor langer Zeit eingetaucht wurden und lediglich verschieden dauerhaft magnetisch sind. Alle diese (und andere konstruierbare) Fälle werden durch den Aufgabentext verworfen bzw. nicht nahegelegt, und sie tauchen in Hagemeisters Aufgabendiskussion nicht auf.

Anders ist es mit dem Fall, dass unter A und B ein Stoffgemisch lag, z.B. Kaffee und Eisenpulver, und jetzt das Eisen an den Magneten klebt und der Kaffee unter A und B liegt. Die Interpretation zeigt, dass der Text diesen Fall nicht zuläßt, trotzdem taucht er bei Hagemeister auf. Hier zeigt sich nun die Stärke der methodischen Kontrollierbarkeit der Interpretation: Man kann sich nicht darauf zurückziehen, dass man „sich da etwas Falsches gedacht" habe. Die Interpretation ist korrekt durchgeführt und kann nicht einfach verworfen werden, nur weil sie oberflächlich nicht „paßt". Wenn der Text den Fall des Stoffgemisches nicht nahe legt, muß etwas anderes diesen Fall aufrufen: Wir wissen aus der Interpretation, dass der Text durch den doppelten Bruch eine Irritation hervorruft, und einen Zweifel an der scheinbaren Alltäglichkeit des vorliegenden Problems. Wir wissen, dass (vorrangig latent) ein wissenschaftlicher bzw. schulisch-experimenteller Kontext errichtet wird. Diesen Kontext nehmen Hagemeister und seine Schüler auf, und sie beziehen sich innerhalb dieses Kontextes auf den naheliegendsten und innerhalb ihres schulischen Horizonts liegenden Fall, nämlich den Schwefel-Eisen-Fall.

Das bedeutet aber, dass die oben genannten, eher absonderlichen Fälle (verschiedene Magnete usw.) durchaus auftauchen könnten, wenn man die Aufgabe mit Physikstudenten diskutiert.[11]

Es zeigt sich also, dass die objektiv-hermeneutische Interpretation weit mehr erklären kann als die Herangehensweise von Hagemeister. Wenn er behauptet, dass elaborierte Physikkenntnisse daran hindern, die Lösung schnell zu finden oder dass Multiple Choice die Schüler behindert, so kann ich dies genauer fassen: Ich belege textlich, wodurch ein „Kenner" und wodurch auch jeder andere irritiert wird, welche Dimensionen eines Konstrukts, das man Testfähigkeit nennen kann, diese Irritation überwinden helfen und inwieweit Multiple Choice diese Probleme abschwächt oder verstärkt. Im Gegensatz zu Hagemeister unterliegt die objektiv-hermeneutische Interpretation einer methodischen Kontrolle. Dies vereinfacht es auch, Aufgabenkritik und normative Vorstellungen von Unterricht methodisch voneinander zu trennen. Probleme der Gesprächsführung und der Suggestion werden vermieden. Man kann auf Aussagen wie „Es wird nur Textverständnis getestet." verzichten, da man genau angeben kann, an welchen Stellen Textverständnis u.a. nichtfachliche Fähigkeiten gemessen werden.

[11] Man könnte meinen, dass diese Absonderlichkeiten doch alle gar nicht angesprochen sind - und manifest spricht der Text sie wirklich nicht an. Testkritiker würden dagegen halten, dass die Aufgabe diese Absonderlichkeiten ja doch nahe lege und irgendetwas damit nicht stimmt. Erst die Entschlüsselung des latenten Aufgabentexts erklärt die Ursache für diesen Eindruck.

1.2. Zur Verwendung der Theorie der mentalen Situationsmodelle bei Baumert u.a.

Wie läßt sich erklären, dass die Baumert-Hagemeister-Debatte eine so magere Antwort auf die dort gestellte Frage, was die TIMSS-Aufgaben gemessen haben, gibt? Bezüglich Hagemeister ist die Frage schnell beantwortet: Er hat methodisch unkontrolliert in einer pragmatischen Laborkonstellation Erkenntnisse generiert. Die kann man ebenso pragmatisch diskutieren, hat aber gegen eine methodisch kontrollierte Gegenuntersuchung nur eine schwache Argumentationsbasis.

Baumert u.a. argumentieren hingegen auch selbst auf der Basis nachvollziehbarer Methoden, deren Grenzen sich aber gezeigt haben:

- Die Grenzen der Argumentation mit **Trennschärfen** hat die objektiv-hermeneutische Interpretation gezeigt: Obwohl die Aufgabe laut Baumert u.a. über gute Trennschärfewerte verfügt, trennt sie nicht sicher zwischen Wissenden und Nichtwissenden. Für eine spezielle Gruppe birgt die Aufgabe eine Benachteiligung - in diesem Fall für diejenigen, die die Aufgabe auf Alltagsniveau lösen könnten, die Frage aber auf dem durch den objektiven Aufgabentext induzierten wissenschaftlichen Niveau nicht beantworten können und gleichzeitig die entstehende Irritation nicht überwinden können.
- Die Unbrauchbarkeit des **Konstrukt**s **der „Fähigkeitsniveaus"** für die Beschreibung von Aufgabenanforderungen und „Fähigkeitsanalysen" hat sich angedeutet: Der Aufgabenpunktwert kann nicht widerspiegeln, dass die Aufgabe auf zwei verschiedenen Niveaus testet, nämlich auf Alltagsniveau und auf einem wissenschaftlichen Niveau. Die Fähigkeitsniveaus sind außerdem so beschrieben, dass für einfache Faktenerfragung, die sich aber auf laborgestützte, wissenschaftliche Absicherung des Faktenwissens stützt, gar keine Kategorie vorgesehen ist.

Im folgenden geht es um die Theorie der mentalen Situationsmodelle, die für mathematische Problemaufgaben von Kurt Reusser spezifiziert wurde. Das von Baumert u.a. für die Aufgabe D2 beschriebene Situationsmodell hilft bei der Beantwortung der Frage, was die Aufgabe mißt, wenig. Ich werde im folgenden beschreiben, was der Inhalt der Theorie der mentalen Situationsmodelle ist und in welchem Maße die argumentative Schwäche von Baumert u.a. bereits in der zugrundeliegenden Theorie ihre Ursache hat. Da Kurt Reussers Ansatz sich auf die Theorie des Textverstehens von Kintsch und van Dijk bezieht, werde ich diese kurz besprechen. Nach einer Diskussion von Reussers Ansatz wird deutlich werden, welche theoretischen und welche Umsetzungsprobleme die argumentative Schwäche von Baumert u.a. hervorrufen.

Neben dieser theoretischen Einordnung und Diskussion des interpretierenden Vorgehens von Baumert u.a. werden sich Forschungsfragen für die weiteren Interpretationen ergeben. Sie dienen dazu, die Leistungsfähigkeit der objektiv-hermeneutischen Aufgabeninterpretation gegenüber dem kognitionspsychologischen Vorgehen mit mentalen Modellen zu untersuchen, aber auch die Grenzen objektiv-hermeneutischen Vorgehens zu umreißen.

1.2.1. Die Theorie des Textverstehens von Kintsch und van Dijk und die Wirkung ihrer Rezeption bei Reusser und Baumert u.a.

Kurt Reusser schreibt zur Einführung in sein Modell:

> „Wer ein mathematisches Textproblem lösen will, steht vor einer komplexen Konstruktions- oder Strukturerzeugungsaufgabe. Er muss die sprachlich dargestellte Gegebenheit enkodieren und die darin ausgedrückte Handlungs- und Situationsstruktur vergegenwärtigen. Text- und Kognitionspsychologen sprechen in diesem Zusammenhang etwa vom Aufbau einer Textbasis (Kintsch 1974), von der Konstruktion mentaler Situationsmodelle (van Dijk & Kintsch 1983, Johnson-Laird 1983) oder von der Deutung einer Gegebenheit durch ihre Integration ins Weltwissen (Aebli 1980). Einen solchen Konstruktions- oder Interpretationsvorgang als psychologischen Prozess zu erklären, verlangt nicht nur eine Theorie über dessen Endstufe oder Ergebnis, sondern auch über die Zwischenstufen und konstitutiven Teilprozesse.” (Reusser 1989, S.15)

Mit Piaget lenkt Reusser unseren Blick auf die Einheit der strukturtheoretisch und der prozeßtheoretisch zu betrachtenden Komponenten dieses Prozesses:

> „Strukturtheoretisch lassen sich die progressiven Stufen des Verständnisses einer Textrechnung als Netze oder Systeme von Beziehungen beschreiben. Sie repräsentieren die Vergegenwärtigung der Handlungs- oder Sachsituation, wie sie der Problemlöser unter mathematischer Perspektive erzeugt. Prozesstheoretisch sind es die Prinzipien, Schritte und Merkmale eines strategisch-konstruktiven und inferenziellen Geschehens, welche einer Deutung bedürfen. [...]
> Der Aufbau einer Problemrepräsentation und deren Transformation in eine arithmetische Struktur sollte nun nicht einfach als eine „Fahrt ins Blaue” gesehen werden. Struktur und Prozess sind dort aufeinander bezogen, wo die – vorerst bloß global gegebene – Zielvorstellung der Mathematisierung (zum Beispiel in der schematischen Gestalt einer Gleichung mit Lücke) dem Verstehensprozess Richtung bzw. Perspektive gibt.” (ebenda, S.15)

Reussers Textaufgabenlösungsmodell bezieht sich u.a. auf die Theorie der Textverarbeitung von Kintsch & van Dijk (1978) und van Dijk & Kintsch (1983). Deren wichtigste Annahmen sind,

> „den Prozess der Textverarbeitung als im Prinzip zweistufigen (in dieser Zweistufigkeit aber nicht notwendigerweise sequenziellen) Vorgang aufzufassen: als Prozess des Aufbaus einer textnahen, propositionalen Mikro- und Makrostruktur (von Kintsch seit 1974 als *Textbasis* bezeichnet) und als Prozess der Konstruktion eines von der Textstruktur abgeleiteten bzw. vom Text denotierten, mentalen *Situationsmodells*. Der Prozess des Verstehens einer arithmetischen Textaufgabe besteht ... demnach darin, „von der verbalen Form der Aufgabe eine begriffliche Repräsentation zu konstruieren, auf welcher Problemlöseprozesse operieren können (...)” (Reusser 1989, S.57)

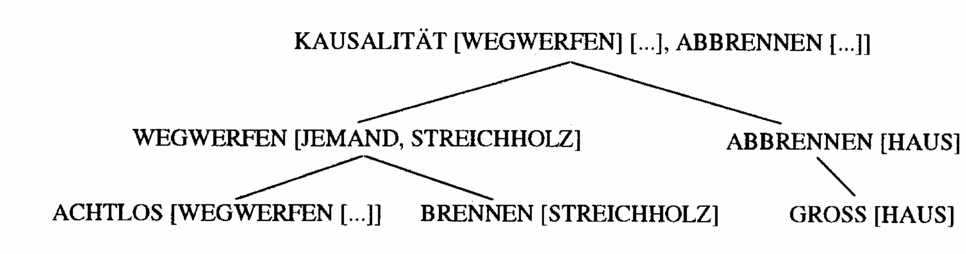
Walter Kintsch erläutert zu seiner und van Dijks Theorie der Textverarbeitung:

> „In von Konzepten der Künstlichen Intelligenz inspirierten Textverstehensmodellen stellt das Situationsmodell die entscheidende mentale Repräsentation dar. Damit ist gemeint, was der Text über

eine gewisse Situation aussagt, welche Schlüsse daraus folgen, wie die Textinformationen sich zum Vorwissen verhalten, usw. [Davon findet sich in den Situationsmodellen von Baumert u.a. nichts. W.M.] Das Situationsmodell ist die psychologisch realisierte Bedeutungsvorstellung (Repräsentation) dessen, worüber der Text eine Aussage macht. Psychologen, die sich für das Verstehen und Erinnern von Texten interessieren, müssen sich darüber hinaus mit weiteren Repräsentationsebenen beschäftigen, wie zum Beispiel der Oberflächenstruktur (den Wörtern und Sätzen im eigentlichen Sinn und ihrer linguistischen Struktur) oder der Textbasis (also dem semantischen Gehalt und der Organisation des Textes als solcher, im Unterschied zu der darin beschriebenen Situation)." (Kintsch 1994, S.41)

Diese von Kintsch so genannte Oberflächenstruktur und Textbasis arbeitet die Objektive Hermeneutik in ihren Interpretationen ebenso heraus, aber eben nicht nur sie, sondern auch die latente Sinnstruktur des Textes. Letztere spielt in der Textverarbeitungstheorie von Kintsch & van Dijk (und damit von Reusser und Baumert u.a.) keine Rolle. Damit kennt auch die von ihnen erarbeitete Textbedeutung keine latente Ebene. Die daraus erwachsenden theoretischen Probleme illustriert folgendes Beispiel. Kintsch zeigt daran die Funktionsweise des Textverstehensmodells[12]:

„Dem Text „Ein brennendes Streichholz wurde achtlos weggeworfen. Ein grosses Haus ist abgebrannt." entsprechen die folgenden Propositionen (Bedeutungseinheiten): BRENNEN (STREICHHOLZ), WEGWERFEN (JEMAND, STREICHHOLZ), ACHTLOS (WEGWERFEN (...)), ABBRENNEN (HAUS), GROSS (HAUS), KAUSALITÄT (WEGWERFEN (...), ABBRENNEN (...)). Die ersten fünf Propositionen lassen sich direkt aus dem Text ableiten. Die KAUSALITÄT (...) Proposition ist eine auf Allgemeinwissen basierende Überbrückungsinferenz, die notwendig ist, um die beiden Sätze kohärent zu machen. Die Propositionen sind in der folgenden Abbildung in einem Netz miteinander verbunden:



Dieses Netz stellt die Textbasis dar – die mentale Repräsentation des Textes. Die Textbasis ist hierarchisch geordnet, was psychologisch bedeutsame Folgen hat. So werden zum Beispiel übergeordnete Propositionen in der Regel besser behalten als untergeordnete. Nehmen wir an, der Text würde auf irgend eine Art weitergeführt. Da die Kapazität des menschlichen Arbeitsgedächtnisses strikt begrenzt ist, ist es nicht möglich, die Textbasis im Arbeitsgedächtnis beliebig zu vergrößern. Um dem neuen Text im Arbeitsgedächtnis Platz zu machen, muß die bereits konstruierte Textbasis

[12] Ein Beispiel einer solchen Netzdarstellung für einen komplexeren Text – einen Witz – findet man bei Aebli (1994, S.93). Es illustriert den Aufwand und die Aussagekraft solcher Netzdarstellungen.
Reusser (1989) erläutert ausführlich ein Beispiel für solche Netze bei den von ihm verwendeten Textaufgaben (S.95 ff.) und stellt die linguistischen Grundlagen dieser Herangehensweise dar (123 ff.).

ins Langzeitgedächtnis überführt werden. Damit der neue Text aber mit der alten Textbasis verbunden werden kann, werden ein oder zwei besonders wichtige Propositionen (die übergeordneten Propositionen in der Textbasis) im Kurzzeitgedächtnis behalten, so dass sie als Bindeglieder zum neuen Text fungieren können. Auf diese Art kann eine kohärente Repräsentation eines langen Textes konstruiert werden, obwohl immer nur ein kleiner Teil des Textes aktiv bearbeitet werden kann. Weitere Prozesse, basierend auf den sogenannten Makrostrategien, sind nötig, um auch die globale Struktur eines längeren Textes im Gedächtnis zu repräsentieren. Das Modell von van Dijk und Kintsch beschreibt diese Prozesse und erlaubt es, das Verstehen eines Textes im Detail zu simulieren. Die Ergebnisse solcher Simulationen stellen theoretische Vorhersagen dar, die auf verschiedene Arten empirisch überprüft werden können: über die Reproduktion oder das Zusammenfassen von Texten, über Lesezeiten, Inferenzen, und so fort.” (Kintsch 1994, S.42)

Diskussion:

1. Das Anliegen dieser Arbeit besteht nicht darin, die propositional orientierte Theorie des Textverstehens zu kritisieren. Ich möchte aber auf eine problematische Kernstelle der Argumentation hinweisen: Kintsch schreibt: „Die KAUSALITÄT (...) Proposition ist eine auf Allgemeinwissen basierende Überbrückungsinferenz, die notwendig ist, um die beiden Sätze kohärent zu machen.” (ebenda) Diese Aussage ist methodisch nicht zu kontrollieren, sie lebt lediglich von Schlüssigkeit. Alle anderen Propositionen könnte man als semantisch kontrollierbar bezeichnen, die Kausalität aber wird einfach auf Allgemeinwissen zurückgeführt. Gleichzeitig trägt die Kausalität als „übergeordnete Proposition” die Hauptlast der Textbedeutung. Kintsch hat recht, ohne die Kausalität sind die beiden Sätze nicht kohärent. Eine hier nur angedeutete objektiv-hermeneutische Interpretation des Textes würde aber sofort zeigen, wie das Zusammenlaufen der Sätze sprachlich zu erklären ist:

Ein brennendes Streichholz ... könnte auf zwei Wege führen: entweder auf „positive” Anwendung im Sinne des Menschen (*... sorgt für einen warmen Ofen*) oder auf Gefahr (*... gehört nicht in Kinderhände*). Das *ein* verstärkt durch seine Diffusität z.B. im Vergleich mit *das* im Gefahrfall die Gefährlichkeit. Das *achtlos* im zweiten Satzteil kennzeichnet nun eindeutig den Gefahrfall. Zugespitzt gesprochen: Ab jetzt kann es nur noch brennen. Das *weggeworfen* spezifiziert das eigentlich nur noch, die Strukturlogik der Gefahr wird aber auch nochmals verstärkt. Am Ende des ersten Satzes ist damit schon klar, dass es entweder brennen muß oder jemand die Gefahr bannt. Der zweite Satz ist also mit dem ersten schon verbunden, die Kausalität ist sprachlich vorgezeichnet. Sie muß nicht postuliert werden, sondern ist methodisch kontrolliert herleitbar.

Das Beispiel illustriert beredt die Vereinfachungen, aber auch die notwendigen Verrenkungen („Allgemeinwissen”), die die Vernachlässigung des Latenten durch die propositionale Herangehensweise mit sich bringt. In der objektiv-hermeneutischen Arbeit spielen hingegen gerade die Verwerfungen, die die Nichtkohärenz von latentem und manifestem Text mit sich bringt, eine zentrale Rolle.

2. Die vorliegende Theorie faßt den Prozeß der Textverarbeitung als im Prinzip zweistufigen (in dieser Zweistufigkeit aber nicht notwendigerweise sequenziellen) Vorgang auf: als Prozeß des Aufbaus einer textnahen, propositionalen Mikro- und Makrostruktur (*Textbasis*) und als Prozeß der Konstruktion eines von der Textstruktur abgeleiteten bzw. vom Text denotierten, mentalen *Situationsmodells*. Das Situationsmodell ist dabei die psychologisch realisierte Bedeutungsvorstellung (Repräsentation) dessen, worüber der Text eine Aussage macht.

Der Prozeß des Verstehens einer arithmetischen Textaufgabe besteht dann darin, „von der verbalen Form der Aufgabe eine begriffliche Repräsentation zu konstruieren, auf welcher Problemlöseprozesse operieren können”.

Wenn man diese Zweistufigkeit wirklich annimmt, so bleibt immer noch die **Frage, *wovon* man ein Situationsmodell aufbaut**. In den Texten von Reusser und Baumert u.a. wird immer selbstverständlich davon ausgegangen, dass man bei Textaufgaben ein Situationsmodell der Problemstellung, welche die Aufgabe stellt, aufbaut. Das vernachlässigt aber den spezifischen Zweck des Lösens einer Testaufgabe, der nicht im Lösen der Problemstellung, sondern in der Generierung der richtigen Antwort liegt. Bei Reussers Modellaufbau führt diese Sichtweise noch nicht zu Problemen, weil seine Fragestellung eindeutig auf das Verstehen der Aufgabenstellung zielt. Ihm geht es nicht um einen Kontrollprozeß, sondern um einen Lernprozeß. In diesem Sinne befaßt er sich nur mit einer begrenzten Klasse von Aufgaben. Er verweist auch darauf, dass beim empirischen Umgang mit einer Theorie stets zu fragen ist, auf welche Empirie bzw. auf welche Anwendungssituationen sich eine Theorie bezieht (Reusser 1989, S.303). Unser Problem des *Zwecks* einer Aufgabe behandelt er nur implizit:

> „Wer einen Text liest, tut dies in einem mehr oder weniger spezifischen intentionalen Kontext (...): er sucht Antwort auf eine Frage oder die Lösung eines Problems; er möchte zu einem Gegenstand X sein fragmentarisches Wissen anreichern oder besser integrieren; er liest ein Buch, um es zu besprechen, oder eine Diplomarbeit, weil er dazu ein Gutachten schreiben muß. Perspektiven oder Verarbeitungshinsichten können integrative Bestandteile eines Textes selber sein: eine explizite Wieviel-Frage in einer Textrechnung; eine Titelfrage in einer Kolumne, ... Fragen und Probleme können aber auch aus dem Kontext entstehen oder (bewußt oder unbewußt) vom Leser an einen Text herangetragen werden: ... eine mathematische Frage an einen Lehrtext zur Entstehung von Ebbe und Flut ...
>
> Der Schüler, der eine mathematische Situationsaufgabe (...) zu verstehen und zu lösen sucht, erzeugt demnach eine semantische Beschreibung oder ein Situationsmodell entsprechend der Perspektive, die er als „schematische Antizipation” (...) des Verstehensziels oder der Lösung (...) an die Aufgabe heranträgt.” (Reusser 1989, S.81)

Auch Reusser schreibt hier, dass das Lösen von Schulaufgaben das vom Aufgabensteller antizipierte Ziel ansteuert. In seiner Fragestellung geht es aber nicht darum zu untersuchen, ob Ziel und Aufgabenstellung wirklich konform laufen oder den Zusammenhang zwischen diesen Antizipationen und dem Lösungsprozeß zu erforschen.

Baumert u.a. hingegen untersuchen Testaufgaben. Damit hat die Vernachlässigung der Frage, *wovon* ein Situationsmodell entwickelt werden muß, Konsequenzen: Sie gehen mit Selbstverständlichkeit davon aus, dass die Aufgabe die Konstruktion eines Situationsmodells des gestellten Problems verlangt. Es gibt aber Fälle, in denen durch eine ernsthafte Repräsentation die Aufgabe unlösbar würde (man muß also Zusatzbedingungen kennen, z.B. über spezifisches Funktionieren von Schulaufgaben) oder in denen die Repräsentation des Problems viel aufwändiger wäre als eine sofortige mathematische Verarbeitung ohne inhaltliches Verständnis. Multiple Choice verändert den Charakter der Aufgabe zusätzlich. Wenn dann noch die Probleme des Auseinanderlaufens von latentem und manifestem Text hinzukommen, ergibt sich eine Vielzahl von möglichen Konstellationen, in denen die Intention des Aufgabenstellers und der wirkliche Aufgabentext auseinander laufen. In den Aufgabeninterpretationen werden sich viele Ausgestaltungen dieser Problemlage zeigen.
Auch Baumert u.a. verweisen immer wieder auf die Spezifität der Testsituation, ignorieren aber die genannten Gefahren. Das ist eine wesentliche Ursache dafür, dass sie an der Frage, *was* die Aufgaben messen, vorbei argumentieren.

Es hat sich gezeigt, dass diese Ignoranz nicht aus der Verwendung des Reusser-Modells herrührt, dass das Modell die spezifische Einbettung der Aufgabe aber auch nicht explizit beleuchtet, dass Reusser sie aber – über Kintsch und van Dijk hinausgehend – bereits thematisiert. Innerhalb des kognitiven Ansatzes thematisiert die Theorie der situierten Kognition den konkreten Handlungskontext explizit. Hier versteht man „unter einem mentalen Modell ... die Repräsentation einer ganz spezifischen Anforderungssituation. Mentale Modelle sind nicht als Abbilder der Realität, sondern als auf das *Handlungsziel bezogene Ausschnitte* aus der Realität zu verstehen." (Stern 1998, S.30)
„Eine Person verfügt über die Kompetenz zur Bewältigung einer Anforderung, wenn sie über Wissen verfügt, das es ihr ermöglicht, von den irrelevanten Aspekten der Situation zu abstrahieren und Wissen über die funktionalen Prinzipien der relevanten Aspekte zu aktivieren." (Stern 1998, S.35) Hier wird deutlich, dass man sich über die Anforderungssituation klar sein muß, wenn man das zu aktivierende Wissen beschreiben möchte. Wenn der Aufgabentext eine mehrschichtige Anforderungssituation herstellt, sei es (ungewollt) durch widersprüchliche latente und manifeste Signale oder (gewollt) durch die Einarbeitung verschiedener Anforderungen, dann muß auch entsprechend mehr Wissen aktiviert werden. Ein „funktionales Prinzip" kann dann z.B. auch Irritationsresistenz oder eine eng begrenzte verbale Fähigkeit sein. Mit dem Ansatz der situierten Kognition wird innerhalb des kognitiven Paradigmas deutlicher, dass man die Anforderungssituation analysieren muß. Baumert u.a. bezeichnen aber lediglich das als Anforderungssituation, was sie sich als die Anforderungssituation vorstellen. 3. Das Modell von Kintsch und van Dijk ermöglicht über die Darstellung der propositionalen Netzstruktur, die kognitive Verarbeitung längerer Texte zu untersuchen. Betrachtungen über die Repräsentation von Inhalten im Kurzzeit- oder im Arbeitsgedächtnis sind mit der Objektiven Hermeneutik nicht möglich. Hier zeigen sich Grenzen, wenn man aus der Aufgabeninterpretation heraus Aussagen über die kognitive Verarbeitung der Aufgabe treffen möchte.

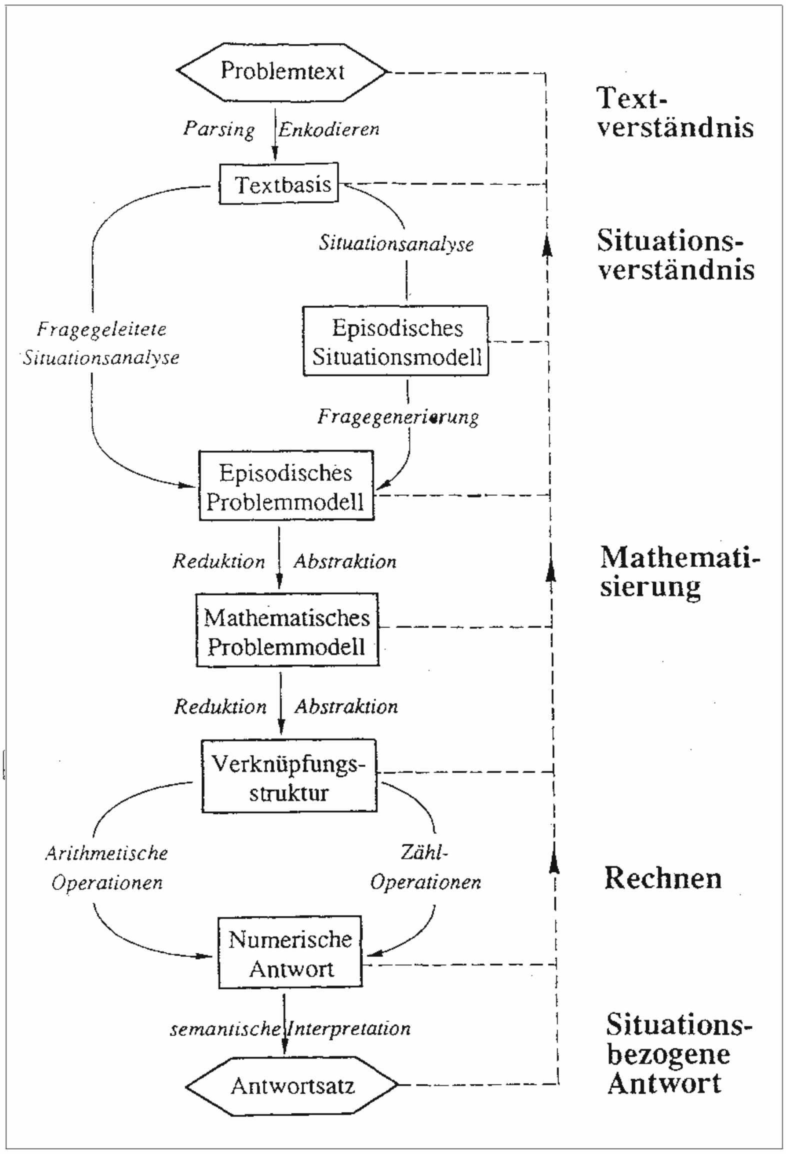
Natürlich liegt der Gedanke nahe, beide Ansätze miteinander zu verzahnen, indem man zunächst objektiv-hermeneutisch die Textbedeutung rekonstruiert und dann mit Hilfe dieses Wissens die „propositionale“ Struktur beschreibt. Über den Begriff der Proposition müsste dabei hinausgegangen werden, da latente Textbedeutungen auch ihren Platz finden müßten. So hätte man eine methodisch kontrollierte und umfassende Textbedeutungsrekonstruktion mit einer die Gesamtstruktur des Textes beschreibenden, sogar begrenzt maschinell verarbeitbaren Darstellungsweise kombiniert.

1.2.2. Vom Text zur Situation zur Gleichung: Kurt Reussers Modell des Lösens mathematischer Textaufgaben

In seiner Habilitationsschrift „Vom Text zur Situation zur Gleichung. Kognitive Simulation von Sprachverständnis und Mathematisierung beim Lösen von Textaufgaben.” entwickelt Kurt Reusser (1989) ein Modell der sprachlichen, sachlichen und mathematischen Prozesse beim Verstehen und Lösen mathematischer Textaufgaben (SPS – Situation Problem Solver). „Mit dem Prozess des Lösens einer Textaufgabe sind jene psychologischen Vorgänge gemeint, die sich zwischen der Präsentation eines Aufgabentextes und der Formulierung eines Antwortsatzes abspielen.” (Reusser 1989, S.6) „Das theoretische Kernstück bzw. die Hauptleistung von SPS bildet dabei der aufgabenspezifische *Aufbau einer episodischen Situations- oder Problemrepräsentation und deren Ueberführung in eine mathematische Problemrepräsentation*.” (ebenda, S.297)

Reusser legt als theoretisches Modell eine kognitive Simulation vor, welche Typen von Textaufgaben löst, wie sie im Mathematikunterricht der Grundschule zu finden sind. Beim methodischen Ansatz der Computersimulationen wird zunächst eine These darüber aufgestellt, wie ein kognitiver Prozeß funktioniert. Man programmiert diesen Prozeß. Wenn das Programm läuft, dann geht man davon aus, dass der angenommene Prozeß in der angenommenen Weise funktionieren *kann*. Es bedarf dann weiterer empirischer Untersuchungen, ob der Prozeß auch beim Menschen so läuft. (ebd., S.17 – 24)
Für die Frage, inwieweit Reussers Modell geeignet ist, um damit Testaufgaben zu untersuchen, ist sein Charakter als Simulationsmodell nur punktuell relevant. Deshalb diskutiere ich Computersimulationen als Methode nicht allgemein. Ich setze mich nur im Sinne meiner eigenen Fragestellung mit dem Modell und seinen sonstigen theoretischen Implikationen auseinander.

In der folgenden Abbildung ist das Modell mit Reussers Bildunterschrift dargestellt:



Text zur Abbildung:

> „Vom Text zur Situation zur Gleichung. Darstellungs- und Verarbeitungsstufen bei der Mathematisierung von Textaufgaben. Kern des sprachlich-sachlichen und mathematischen Verstehensprozesses ist der planvolle (strategische) Aufbau einer Situationsvorstellung und deren schrittweise Transformation in eine numerische Struktur. Die beiden Wege von der Textbasis zum Episodischen Problemmodell bedeuten, dass Aufgaben mit oder ohne Frage dargeboten werden können. Im letzteren Fall muss der Problemlöser oder das System eine adäquate Problemfrage selber erzeugen.“ (Reusser 1989, S.91)

Aufgabenraum und empirische Absicherung des Modells

Reusser arbeitet mit elementaren Textaufgaben: semantisch einfache Handlungsprobleme, in denen mit Objekten hantiert wird oder Objekte ihre Besitzer wechseln (eingliedrige Additions-, Subtraktions- und Ergänzungsaufgaben) (ebenda, S.11). „Der Aufgabenraum von SPS besteht aus einer Welt des (zielgerichteten) Besitztransfers. ... Eine SPS-Aufgabe lösen heisst somit, eine Besitz- und Handlungssituation, in der ein oder mehrere Koaktoren durch den Transfer von Objekten interagieren unter einer gegebenen (Frage) oder mathematischen Perspektive rekonstruieren." (ebenda, S.113)

Zwar arbeitet Reusser damit mit Aufgaben, die erheblich weniger komplex sind als die Testaufgaben, um die es in dieser Arbeit geht[13], dies stellt aber nur ein sehr schwaches Argument gegen die Verwendung des Modells dar: Die elementaren Aufgaben dienen ja lediglich der Bewährung des Modells innerhalb eines bestimmten Paradigmas. Zugespitzt formuliert: Wer dem Paradigma der Computersimulation nicht folgen kann, würde es nicht als Argument gegen das Modell betrachten, wenn es auf dem Computer nicht liefe.

Man kann aus dem schwachen Gegenargument allerdings die Frage ziehen, warum das Modell für komplexere Aufgaben nicht funktionieren sollte. Wenn man das Modell dynamisch und nichtlinear interpretiert[14], ist die Lösung komplexerer Aufgaben denkbar. Die meisten Gegenargumente werden letztendlich darauf hinaus laufen, dass hier die latente Textebene nicht berücksichtigt wird. Außerhalb dieses Arguments spricht nichts dagegen, das Modell an elementaren Aufgaben zu entwickeln und dann die Möglichkeit der Ausweitung auf komplexere Aufgaben zu untersuchen. Gleichzeitig wird daran deutlich, dass Reusser entweder nur Aufgaben verwendet hat, bei denen latenter und manifester Text zusammenlaufen - oder er hat „Alltagswissen", wie z.B. die Kausalitätsproposition bei Kintsch, als „Weltwissen" mit ins Computerprogramm geschrieben. Im Sinne des pragmatischen Umgangs mit der Theorie von Kintsch und van Dijk und der Schaffung von Vorformen „künstlicher Intelligenz" scheint mir das durchaus sinnvoll; im Sinne des Verständnisses kognitiver Prozesse müssen solche „Heilungen" klar gekennzeichnet werden, da diese Lücken ja theoretisch geschlossen werden müssen.

---

[13] Folgende Beispiele vermitteln einen Eindruck von den komplexesten Aufgaben, die Reussers System bearbeiten kann (Reusser 1989, S.116):
- Marianne wants to bring hand-written place cards to Eileen`s birthday party. So far she has written twelve cards. She (already) wrote five place cards yesterday, and she completed some more today. But she still needs six cards.
How many place cards does Marianne need altogether?
How many place cards did Marianne write today?
How many children did Eileen invite to the party?
- A few months ago, George startet breeding rabbits. Recently he got his first fourteen baby rabbits. Today, he has sold some rabbits to Mr. Maier. And now he has already nineteen baby rabbits.
How many baby rabbits did George sell to Mr. Maier?
How many fewer baby rabbits does George have now than he got first?
How many more baby rabbits does George have now than he got first?

[14] Reusser scheint das Modell durchaus dynamisch und nicht streng linear zu denken (siehe ebenda, S.79). Andererseits sieht er seinen (dann wohl doch strenger gedachten) linearen Stufenaufbau als problematisch an (ebd., S.298). Auch Baumert u.a. argumentieren linear und damit inadäquat für Problemlöseprozesse. Bei der objektiv-hermeneutischen Interpretation kann man relevante Aussagen über die Aufgabe treffen, ohne den Problemlöse"mechanismus" selbst beschreiben zu müssen. Damit hat man gewissere Aussagen über die Aufgabe, verzichtet aber auf den Anspruch, diesen „Mechanismus" näher zu beschreiben.

Reusser ordnet sich selbst eher vorsichtig ins Forschungsprogramm der kognitiven Modellierung ein: „In seiner strikten Version impliziert dieses Forschungsprogramm die Annahme, dass sich psychologische Erklärungstheorien als Computerprogramme formalisieren lassen.” (ebd., S. 295) Sein eigenes Modell SPS „stellt den Versuch dar, ... den Prozess des Verstehens und Lösens einer bestimmten Klasse von mathematischen Textaufgaben in Gestalt einer *Rahmentheorie* so detailliert und vollständig zu beschreiben, dass sich wesentliche Teile davon in der Form eines implementierten Computerprogramms formalisieren lassen.” (ebd., S.296)

In einer Fußnote schreibt Reusser dazu: „Ich bin mir nicht sicher, vermute aber, dass es sich bei dieser Rahmentheorie um eine *Perspektivtheorie* ... handelt, also um eine Theorie, welche eine *neue Sichtweise* bezüglich des in Frage stehenden Phänomenbereichs konstituiert, *ohne dass ihr empirischer Gehalt schon genau bestimmbar wäre*.” (ebd.) Bei diesem „Gedanken-Experiment” läßt sich der Prozeß

> „durch Laufenlassen auf dem Computer auf einer theoretischen Ebene in all seinen Facetten studieren, und insbesondere lassen sich mittels Simulationsexperimenten die dem Modell zugrundeliegenden Hypothesen in ihren Konsequenzen genau untersuchen. Inwiefern SPS jedoch eine valide empirisch-psychologische Erklärungstheorie für gewisse Phänomenklassen darstellt, kann – wenn überhaupt – ohne Konfrontation des Modells mit empirischen Daten nicht entschieden werden. Was im Sinne eines *sufficiency tests* als Erfolg gewertet werden darf, ist die Tatsache, dass sich die Theorie zur Umsetzung in ein lauffähiges Computerprogramm bewährt hat und damit als kohärentes theoretisches System gelten kann.” (ebd.)

Reusser ist also selbst nicht der Meinung, dass sein Modell empirisch schon abgesichert sei[15] [16]. Das läßt uns zwar den Umgang von Baumert u.a. mit dem Modell differenzierter sehen (denn sie tun so, als ob hier ein empirisch abgesichertes Modell vorliege, mit dem man mal eben über die Testaufgaben rübergehen kann), erklärt aber die Schwäche von deren Argumentation noch nicht.

Reussers Arbeit mit Propositionen und die Integration von Weltwissen

Reusser arbeitet mit Propositionen, aus denen zusammengesetzt der Text angesehen wird. Das Prinzip dieses Vorgehens ist bereits im obigen Zitat von Walter Kintsch erläutert. An dem dort dargestellten Beispiel des brennenden Hauses habe ich auch schon angedeutet, warum eine objektiv-hermeneutische Textanalyse den Text adäquater beschreibt, im dortigen Fall z.B. die sprachliche Verankerung der Kausalität besser erfaßt. Ich kann hier auch weiterhin keine umfassende Auseinandersetzung mit dem kognitiv-propositionalen Ansatz führen, möchte aber einige Anmerkungen zu Reussers diesbezügli-

[15] Empirische Arbeiten mit dem Modell (Reusser, 1989b; Staub & Reusser, 1995) stellen erste Schritte dar.

[16] Er beschreibt aber mehrere Nutzungsperspektiven für sein Modell:
„(1) als psychologisches Prozessmodell und als psychologische Erklärungstheorie;
(2) als didaktisches Modell der Anleitung zum mathematischen Problemlösen bzw. der Vermittlung von Verstehensstrategien;
(3) als pädagogisch-psychologisches Begriffsinventar der Wissens- und Verstehensdiagnose bei Lernenden;
(4) als Heuristik für psychologische und didaktische Theoretiker.” (Reusser 1989, S.303, tiefer S.303-316)
Im letzteren Sinne ist das Modell für diese Arbeit ein Werkzeug.

chen theoretischen Grundsätzen nutzen, um die prinzipiellen Unterschiede zwischen kognitiv-propositionalen und objektiv-hermeneutischen Textinterpretationen aufzuzeigen.

> „Eine Verstehensprozesstheorie, die Bedeutungsgehalte von Propositionen ausdrückt, deutet den Prozess des Verstehens als Abfolge von propositional definierten Zuständen. Für die Kodierung setzt dies allerdings voraus, dass es gelingt, idiosynkratische Bedeutungsintuitionen weitgehend auszuschalten und die dem Sprachgebrauch innewohnenden Prädikationen in reliabler Weise intersubjektiv zu identifizieren (...). Die Verwendung einer propositionalen Analyse von sprachlicher Bedeutung findet also dort ihre Grenzen, wo eine intersubjektive Kodierung empirisch kaum mehr möglich ist.” (Reusser 1989, S.77)

Hier zeigt sich die grundsätzlich unterschiedliche Herangehensweise von Objektiver Hermeneutik und propositionalem Ansatz: Die Objektive Hermeneutik sagt gerade, dass Propositionen – also die manifeste Textebene - uns die Bedeutung eines Textes noch nicht erschließen, sondern dass man zur Erschließung der Textbedeutung das Zusammenspiel von latenter und manifester Textebene betrachten muß. Damit geht einher, dass man „idiosynkratische Bedeutungsintuitionen” nicht einfach „auszuschalten” versucht, sondern beschreibt, und damit ihren Platz innerhalb der Textstruktur und ihre Bedeutung für die Textbedeutung herausarbeitet. Dabei ist auch beim objektiv-hermeneutischen Vorgehen das Ziel, „die dem Sprachgebrauch innewohnenden Prädikationen in reliabler Weise intersubjektiv zu identifizieren.”
Nicht nur eine „intersubjektive Kodierung” wird hierbei als prinzipiell möglich angenommen; das hieße lediglich, dass eine sprachliche Kodierung dem **Bewußtsein** mehrerer Personen gemeinsam ist. Der Begriff der objektiven Textbedeutung meint darüber hinausgehend, dass der Text eine unabhängig vom Subjekt und seinem Bewußtsein existierende Bedeutung hat, die im Zusammenspiel seiner latenten und seiner manifesten Elemente gründet. Diese Bedeutung ist natürlich nicht beliebig, sondern ergibt sich aus der Regelhaftigkeit von Sprache. Gäbe es sie nicht, wäre sprachliche Verständigung nicht möglich.
Damit hat man nun ein sehr viel reichhaltigeres Instrument als mit dem propositionalen Herangehen in der Hand: Man muß den Text nicht „vorzensieren”, sondern nimmt ihn einfach ernst und kann somit prinzipiell jeden Text interpretieren.

Aus der propositionalen Herangehensweise ergeben sich weiterführende Unklarheiten:

> „Eine Situationsaufgabe oder einen narrativen Text lesen und verstehen heisst, eine Handlungs- und Situationsvorstellung im Geist erzeugen, bzw. die vom Text (Autor) gemeinte Handlungssituation innerlich nachkonstruieren oder strukturell vergegenwärtigen. Ob ich mich dabei quasi als Handlungsteilnehmer in die Rolle eines Aktors versetze oder mir das Handlungsgeschehen distanzierter, quasi als Beobachter, „ansehe”, spielt keine Rolle. Wichtig ist jedoch die Tatsache, *dass* ich ... *intentional* auf die vom Text ausgedrückte Handlungssituation in ihrer *raum-zeitlichen* und *funktionalen* Struktur gerichtet bin und nicht bloss auf eine *Text-als-Text* Struktur.

> Kognitionspsychologisch gesprochen: Zwar nimmt der Verstehensprozess mit dem Aufbau einer textnahen Bedeutungsstruktur (einer Textbasis) seinen Anfang, aber er geht zugleich auch über diese hinaus, indem die Erzeugung der Textbasis eine komplexe Assimilationsleistung darstellt, die ohne die (minimale) Integration des Textes in eine subjektive Wissensstruktur nicht denkbar ist.” (ebenda, S.135 f.)

Zunächst fällt auf, dass Reusser davon spricht, die vom Text (Autor) *gemeinte* Handlungssituation nachzuvollziehen. Objektive Hermeneutik sucht das vom Text (Autor) *Gesagte*, nimmt also zunächst nur den Text ernst. Das Gemeinte erschließt sich dabei im Prinzip auch. „Im Prinzip” meint, dass natürlich auch die Objektive Hermeneutik die Unsicherheit bezüglich der Intention nicht ausschließen kann, sie kann diese Unsicherheit aber klar fassen. Wenn Reusser davon spricht, dass der Rezipient „intentional auf die vom Text ausgedrückte Handlungssituation in ihrer raum-zeitlichen und funktionalen Struktur gerichtet” sein soll, dann steht doch die Frage, woher man die Intention kennt. Kintsch hat oben auf „Allgemeinwissen” verwiesen. In Wirklichkeit ist doch auch diese Intention sprachlich vermittelt, und diese Vermittlung erfaßt Objektive Hermeneutik.[17]

Nun wissen wir – und hier fängt Reussers Problem an – dass ein Text verschieden gemeint und auch verschieden verstanden werden kann. Objektive Hermeneutik geht davon aus, dass auch dieses verschiedene Gemeintsein und Verstandenwerden nicht beliebig ist, sondern durch Analyse des Zusammenspiels von latenter und manifester Textebene beschrieben werden kann. Umgekehrt **entsteht** der Text natürlich auch durch das Zusammenwirken von Latentem und Manifestem. Die hierbei möglichen vielfältigen Kombinationen erzeugen nun gerade die Vielfalt des Gemeint-sein- und Verstanden-werden-Könnens.

Da der propositionale Ansatz die latente Textebene vernachlässigt, steht ihm die Analyse dieses Zusammenspiels nicht zur Verfügung. Er weicht auf „Intention” aus. Kintsch benutzt das Stichwort Allgemeinwissen, Reusser spricht von „Integration des Textes in eine subjektive Wissensstruktur”. Er konkretisiert an anderer Stelle:

> „Texte offenbaren uns im Verstehensprozeß keine textimmanente Bedeutung. Wir interpretieren sie und eignen sie uns an, indem wir sie – Textsegment um Textsegment enkodierend – im Lichte und mit Hilfe unseres Handlungs-, Sprach- und Weltwissens *elaborieren* (...). Textbedeutung ist somit nicht etwas transsubjektiv Fixierbares (...), sondern (erstes) Ergebnis eines genuinen, bewußten oder unbewußten Deutungsvorganges, einer Interaktion zwischen dem Text als Daten- oder Induktionsbasis und dem Vor- und Weltwissen des Lesers (...). Der Begriff der Textbedeutung ist mithin nur relativ zur Struktur eines individuellen Repertoires ... zu fassen. Oder: *vor* seiner Assimilation gibt es keinen Begriff der Textbedeutung im psychologischen Sinne.” (ebd., S.78 f.)

[17] Die Intention spiegelt sich dabei nicht ungebrochen im Text: Schon das Beispiel D2 zeigte, dass zur manifesten Intention (um das Nichtmagnetisch-Sein von Kaffeepulver wissen) latente, noch dazu widersprüchliche Intentionen hinzukamen (Umgehen mit realem oder Gedächtnisexperiment). Es zeigt aber auch, dass zwischen latenten Intentionen und nicht intendierter latenter Textbedeutung nicht mehr unterschieden werden kann, wohl aber zwischen manifester Intention und latenter Textbedeutung bzw. latenten Intentionen. Objektive Hermeneutik ist gerade an den Stellen fruchtbar, an denen die Differenz zwischen beiden Ebenen für die Forschung relevant ist.

Die objektiv-hermeneutische Textanalyse weist an dieser Stelle über Reussers Ansatz hinaus: Sie erarbeitet Textbedeutung **vor** dem „psychologischen Sinne“: Der Text hat schon eine objektive Textbedeutung, bevor er vom Individuum assimiliert wird. Aber erst mit der Assimilation beginnt Psychologie, also Individuelles. Reusser beginnt also eigentlich dort, wo Objektive Hermeneutik aufhört. Insofern muß man objektive und subjektive (die von Reusser beschriebene) Textbedeutung unterscheiden. Aber Psychologie kann auch erst dann umfassend über die Assimilation von Texten sprechen, wenn sie neben der manifesten auch die latente Textebene beachtet. Ich möchte es so ausdrücken: Der objektive Text mit latenten und manifesten Elementen trifft auf eine subjektive kognitive „Dispositionslandschaft“ (die auch „Weltwissen“ beinhaltet) und wird in diese Landschaft hinein zum subjektiven Text assimiliert. Auch dieser subjektive Text und die subjektive kognitive Landschaft haben manifeste und latente Elemente.[18]

Erst wenn man die objektive und die subjektive Textbedeutung gedanklich voneinander trennt, dann ergibt sich die Möglichkeit durchgehender methodischer Kontrolle der Interpretation beider Textbedeutungen: Man kann zunächst z.B. den objektiven Text einer Aufgabe interpretieren und hinterher mit einem Individuum ein Interview über diese Aufgabe führen. Erst vor der Folie der objektiven Textbedeutung erschließt sich die individuelle Besonderheit der subjektiven Textrezeption. Man vermeidet damit ein Problem, das Reusser beschreibt: „Die am Aufbau der Textbasis in SPS beteiligten propositionalen Bedeutungspostulate (...) sind zwar sozial vermittelte, aber in ihrer Mikrostruktur auch durch subjektive Erfahrungen mitgeprägte Assimilationsschemata. Die intersubjektive Übereinstimmung zwischen Meinen und Verstehen (...) lässt sich eben bloß pragmatisch optimieren.” (ebd., S.136)

Mit dieser Position werden auch Situationsmodelle nur pragmatisch konstruierbar. Zwar sind die Situationsmodelle von Baumert u.a. zusätzlich besonders oberflächlich konstruiert, ihre grundsätzliche Pragmatik und methodische Unkontrolliertheit hat aber an dieser Stelle ihre theoretische Ursache.

In der Sichtweise der Objektiven Hermeneutik läßt sich die Übereinstimmung zwischen subjektivem Meinen und Verstehen der Chance nach dann optimieren, wenn man einen völlig reflektierten Text benutzt, in dem latente und manifeste Textbedeutung absolut konform laufen und expliziert sind.

Es ergibt sich eine paradoxe Situation: Die Objektiven Hermeneutik argumentiert mit einer viel geringeren Bedeutung des Subjektiven als es der propositionale Ansatz tut. Dieser landet aber bei Situationsmodellen, die lediglich pragmatische Aussagen treffen, wohingegen Objektive Hermeneutik eine viel größere Palette möglicher **textinduzierter** subjektiver Interpretationen herausarbeiten kann.

Situationsmodelle

> „Unter einem Situationsmodell wird der semantische Zusammenhang verstanden, den eine Person in ihrem Geist erzeugt, wenn sie einen (narrativen) Text liest. ... Ein Situationsmodell ist das kogni-

[18] In dieser Arbeit ist es nicht möglich und für die Fragestellung auch nicht notwendig, sich näher mit der Beschaffenheit dieser subjektiven kognitiven Landschaft auseinander zu setzen. Klar ist, dass durch die Einbeziehung von latenten Elementen Emotionales nicht mehr aus dem Kognitiven ausgeschlossen werden kann.

tive Korrelat der vom Autor eines Textes gemeinten bzw. von einem Leser verstandenen Situationsstruktur (...).” (Reusser 1989, S.136)

Ich habe bereits auf die leistungsfähigere Perspektive der situierten Kognition verwiesen. Hier versteht man „unter einem mentalen Modell ... die Repräsentation einer ganz spezifischen Anforderungssituation. Mentale Modelle sind nicht als Abbilder der Realität, sondern als auf das Handlungsziel bezogene Ausschnitte aus der Realität zu verstehen.” (Stern 1998, S.30)

Die Argumentation zur objektiven Textbedeutung zeigt, dass die „vom Autor gemeinte Situationsstruktur” in einem Modell der Verarbeitung von Textaufgaben eigentlich nichts zu suchen hat. Man muß also als eigenständigen Bestandteil den objektiven Aufgabentext aufnehmen. Bei einer „idealen Aufgabenstellung“ spiegelt der objektive Aufgabentext das vom Autor Gemeinte wider; dazu müssen latenter und manifester Aufgabentext zusammenlaufen. Das Situationsmodell ergibt sich dann aus dem subjektiven Text.[19]

Ich werde - um begriffliche Verwirrungen zu vermeiden - im weiteren Verlauf der Arbeit die Begriffe in folgendem Sinne verwenden:

Ein *mentales Modell* ist die subjektive Repräsentation einer spezifischen Anforderungssituation. Ich denke dabei mit, dass dieses Modell sich aus dem subjektiven Text der Anforderung ergibt, dass dieser subjektive Text sich aus der Assimilation des objektiven Texts der Anforderung in die subjektive kognitive Landschaft ergeben hat und dass sich ein mentales Modell auf das Handlungsziel bezieht. Dieses Handlungsziel wird durch den Text und durch den Rahmen der Situation, der im „Weltwissen“ mit repräsentiert ist, vermittelt.

Es handelt sich also um den Begriff des mentalen Modells aus der situierten Kognition, aber hineinversetzt in die objektiv-hermeneutische Sichtweise. Diese Begriffsbildung bezieht sich nur auf meine Fragestellung, also Testaufgaben, und ist lediglich zur Vereinfachung der Lesbarkeit der Arbeit gedacht. Den Begriff *Situationsmodell* verwende ich im Reusserschen Sinne.

Zielgerichtetheit und strategischer Charakter des Lösens von Aufgaben

Eines der wichtigsten Probleme für die Denkpsychologie ist „die Bestimmung der richtunggebenden Faktoren, die den geordneten Ablauf des Denkens herbeiführen. ...

Der Schüler, der eine mathematische Situationsaufgabe (mit oder ohne explizite Problemfrage) zu verstehen und zu lösen sucht, erzeugt ... eine semantische Beschreibung oder ein Situationsmodell entsprechend der Perspektive, die er als „schematische Antizipation” (...) des Verstehensziels oder der Lösung (z.B. in Gestalt einer schematischen Verknüpfungsstruktur wie einer Gleichung oder Rechnung) an die Aufgabe heranträgt. Mit anderen Worten: das *Verstehen von Texten* ist kontrolliert durch die Anlässe, die Ziele und die Orientierungen eines Lesers, genauso wie das Vorliegen eines Hand-

[19] Dabei ist hier keine Aussage darüber möglich, auf welche Weise die Repräsentation von objektivem Aufgabentext, subjektivem Aufgabentext und Situationsmodell erfolgt. Ein linearer, in Sequenzen aufeinanderfolgender Aufbau ist sicherlich lediglich eine hilfreiche Veranschaulichung.

lungsziels als Vergleichsmuster die fortlaufende Kontrolle einer Handlung erlaubt (...).” (Reusser 1989, S.81)

Die objektiv-hermeneutische Interpretation nimmt diese Gedanken einerseits in ihre Fallbestimmung auf. Hier werden die Spezifitäten der Testsituation zu besprechen sein. Andererseits ergibt sich hier aber auch eine Forschungsfrage für die Interpretation: Auf welche Weise unterstützt bzw. hemmt der Aufgabentext den für die Lösung richtunggebenden Ablauf des Denkens?

Ähnliches gilt für den strategischen Aspekt des Verstehens bzw. Lösens von Aufgaben (siehe Reusser 1989, S. 82 f.) Die Fallbestimmung muß strategisches Vorgehen aufnehmen, die Interpretationen müssen untersuchen, inwiefern durch den Aufgabentext strategisches Vorgehen unterstützt bzw. gehemmt wird. Dies ist insbesondere auf die Debatte um Testfähigkeit hin zu spezifizieren, denn dabei geht es um strategisches Vorgehen und die Frage, inwieweit Aufgaben so beschaffen sind, dass strategisches Vorgehen Verstehen ersetzen kann.

1.2.3. Fazit: Baumert u.a. und die Theorie der mentalen Situationsmodelle

Baumert u.a. argumentieren gegen Hagemeister mit der Theorie der mentalen Situationsmodelle:

> „Um ein Problem oder eine Aufgabe zu lösen, muss der Bearbeiter ein mentales Situationsmodell der Aufgabenstellung entwerfen, das den Rahmen des Lösungsprozesses definiert (vgl. Reusser 1996). Mit der Entwicklung des subjektiven Situationsmodells wird entschieden, um was es überhaupt geht, welches Wissen aktiviert, welcher Lösungsweg gewählt und welche Denkoperationen durchgeführt werden. Das Situationsmodell legt auch fest, auf welchem fachlichen Anspruchsniveau eine Aufgabe behandelt wird. Die Grundmerkmale möglicher oder besser: wahrscheinlicher Situationsmodelle werden sowohl durch die Formulierung und Darbietung der Testaufgabe als auch durch die soziale Situation der Testadministration vorgezeichnet. Eine gute Testaufgabe enthält in sparsamer Form die notwendigen Hinweisinformationen, die bei Probanden, die das durch die Aufgabe indizierte Fähigkeitsniveau erreichen oder übertreffen, zur Bildung eines für die Lösung der Aufgabe adäquaten Situationsmodells führen. Die erforderlichen Hinweisinformationen werden zunächst durch die Aufgabenstellung selbst, dann aber auch durch zusätzlich mitgeteilte Lösungshinweise gegeben.” (Baumert u.a. 2000, S. 196 f.)

Ich habe am Beispiel der Aufgabe D2 gezeigt, dass die aus dieser theoretischen Grundposition erwachsenen Aufgabeninterpretationen wenig zur Klärung der Frage, was die Aufgabe mißt, beitragen. In diesem Kapitel wurden die dieser Position zugrundeliegenden Theorien des Textverstehens von Kintsch und van Dijk und der mentalen Situationsmodelle von Reusser daraufhin untersucht, inwieweit sie zur Schwäche der Aufgabeninterpretationen beitragen.

Ich fasse die Erkenntnisse unter den Stichworten der Vernachlässigung des Latenten, der methodischen Kontrolle, des objektiven Aufgabentextes und des Handlungsziels zusammen:

1. Vernachlässigung des Latenten

Die Vernachlässigung einer latenten Textebene zieht sich durch alle betrachteten Theorien. Das von Kintsch gegebene Beispiel („Ein brennendes Streichholz wurde achtlos weggeworfen. Ein großes Haus ist abgebrannt.") illustrierte die Vereinfachungen, aber auch die notwendigen Verrenkungen („Allgemeinwissen"), die die Vernachlässigung des Latenten durch die propositionale Herangehensweise mit sich bringt. In der objektiv-hermeneutischen Arbeit spielen hingegen gerade die Verwerfungen, die die Nichtkohärenz von latentem und manifestem Text mit sich bringt, eine zentrale Rolle.

Für die Interpretation von Testaufgaben ist die Beachtung der latenten Textebene von großer Bedeutung.[20] Die Interpretation von D2 zeigte bereits die durch die Vernachlässigung des Latenten entstehenden Verwerfungen und ihre Folgen in den Untersuchungen von Hagemeister. Die weiteren Aufgabeninterpretationen werden verschiedene Spielarten solcher Verwerfungen zeigen. Man kann an dieser Stelle bereits feststellen, dass das Aneinandervorbeireden in der Baumert-Hagemeister-Debatte seine Ursache darin hat, dass beide Seiten die latente Textebene der Aufgabenstellung nicht beschreiben können. Baumert u.a. können, selbst nachdem Hagemeister die entstehenden Verwerfungen andeutet, deren Ursachen nicht erkennen, weil sie die Vermischung von wissenschaftlichem und Alltagsproblem und die Vermischung von realem und Gedankenexperiment nicht herausarbeiten können. Dies gelingt erst, wenn man auch die latenten Textelemente analysiert.

2. Methodische Kontrolle

Mit der Objektiven Hermeneutik liegt eine eingeführte Methode vor, die die Analyse der latenten Textebene und damit des objektiven Texts kontrolliert ermöglicht.
Bei Baumert u.a. ist eine methodische Kontrolle nur für die statistischen Kennwerte möglich. Die Aufgabeninterpretationen entziehen sich methodischer Kontrolle, nicht nur in der Vernachlässigung des Latenten: Auch eine propositionale Analyse der Aufgabentexte findet nicht statt. Ich habe zwar am Beispiel von Kintsch angedeutet, dass das Arbeiten mit propositionalen Netzen methodische Kontrolle nicht bis in die latente Ebene hinein ermöglicht, aber doch wenigstens auf der manifesten Ebene. Das leisten Baumert u.a. nicht. Insofern ist auch die Bezugnahme auf Reusser erstaunlich, denn Reusser erstellt propositionale Netze, sobald er eine Aufgabe näher untersucht.

3. Objektiver Aufgabentext und Konstruktion mentaler Situationsmodelle

Das Thema von Kurt Reusser ist die Konstruktion mentaler Situationsmodelle. Die objektiv-hermeneutische Aufgabeninterpretation hat ihr Thema zunächst **vor** dieser Konstruktion, nämlich in

[20] Auch Baumert u.a. gehen davon aus, dass Testaufgaben „in einem standardisierten Leistungstest *Indikatorfunktion für eine latente, im Hintergrund stehende Fähigkeit* oder Kompetenz, deren individuelle Ausprägung für die jeweilige Testleistung einer Person verantwortlich ist" haben. (Baumert u.a. 2000, S.108)

der Analyse des objektiven Aufgabentexts. Weder Kintsch/van Dijk noch Reusser arbeiten damit. Ich habe in der Diskussion von Reussers Ansatz bereits angedeutet, dass mit der Annahme eines objektiven Aufgabentextes auch Textverarbeitung strukturierter und damit klarer beschrieben werden kann. Schon bei Reusser war mit Begriffen wie „Integration ins Weltwissen“ unklar, wie der Aufgabentext auf das Situationsmodell führt. Auch Baumert u.a. widmen sich dem Verhältnis von Aufgabentext und Situationsmodell:

> „Die Grundmerkmale möglicher oder besser: wahrscheinlicher Situationsmodelle werden sowohl durch die Formulierung und Darbietung der Testaufgabe als auch durch die soziale Situation der Testadministration vorgezeichnet. Eine gute Testaufgabe enthält in sparsamer Form die notwendigen Hinweisinformationen, die bei Probanden, die das durch die Aufgabe indizierte Fähigkeitsniveau erreichen oder übertreffen, zur Bildung eines für die Lösung der Aufgabe adäquaten Situationsmodells führen. Die erforderlichen Hinweisinformationen werden zunächst durch die Aufgabenstellung selbst, dann aber auch durch zusätzlich mitgeteilte Lösungshinweise gegeben. Zusätzliche Lösungshilfen sind zum Beispiel bei Mehrfachwahlaufgaben die in den Distraktoren immer implizit enthaltenen Ausschlussinformationen; bei offenen Aufgabenstellungen können dies präzisierende Hinweise sein, die einen Lösungsansatz oder ein bestimmtes Verfahren nahe legen.“ (Baumert u.a. 2000, 196 f.)

Baumert u.a. beschreiben im ersten Satz des Zitats, dass das Situationsmodell durch den Aufgabentext (und damit auch durch seine Einbettung) vorgezeichnet wird. Im zweiten Satz ist dann vom „für die Lösung der Aufgabe adäquaten Situationsmodell“ die Rede (das angedeutete Problem der Aufgabenkonstruktion über Fähigkeitsniveaus diskutiere ich unten). Es stellt sich die Frage: Von welcher Aufgabe wird hier gesprochen - von der intendierten Aufgabe oder von der „objektiven“ Aufgabe? Man stößt wieder darauf, dass die Beachtung der latenten Textebene unweigerlich zu einem anderen Bild davon führt, wie mentale Situationsmodelle aus dem Aufgabentext heraus entstehen. Baumert u.a. können also wegen ihrer theoretischen Verankerung im kognitiven Ansatz die Wirkung des Aufgabentexts auf das Situationsmodell nicht klar beschreiben.

Der dritte und vierte Satz vermischen nun aber auch noch die nächsten beiden Ebenen des Lösungsprozesses: Situationsverständnis und Mathematisierung (bzw. hier: Physikalisierung). Reussers Leistung bestand ja gerade darin, diese beiden Ebenen gedanklich voneinander zu trennen. So entsteht eine seltsame Situation: Baumert u.a. berufen sich auf Reussers Modell und nehmen die dort durch die Vernachlässigung des Latenten entstehende Unklarheit über den Zusammenhang von Aufgabentext und Situationsmodell auf. Gleichzeitig vermischen sie die nächsten beiden Ebenen des Lösungsprozesses, obwohl Reusser mit seiner Arbeit genau diese Vermischung beseitigt. Auch in den konkreten Aufgabeninterpretationen wird in keinem Fall Reussers Modellschema (wenigstens stichwortartig) verwendet. Das erweckt den Eindruck, dass hier nur so getan wird, als würde mit Reussers Ansatz gearbeitet.

4. Handlungsziel, Situationsmodell und Aufgabenlösung

In der hier diskutierten kognitiven Theorie wird von Textverstehen als zweistufigem Prozeß ausgegangen: als Prozeß des Aufbaus einer textnahen, propositionalen Mikro- und Makrostruktur (*Textbasis*) und als Prozeß der Konstruktion eines von der Textstruktur abgeleiteten bzw. vom Text denotierten, mentalen *Situationsmodells.*
„Unter einem Situationsmodell wird der semantische Zusammenhang verstanden, den eine Person in ihrem Geist erzeugt, wenn sie einen (narrativen) Text liest. ... Ein Situationsmodell ist das kognitive Korrelat der vom Autor eines Textes gemeinten bzw. von einem Leser verstandenen Situationsstruktur (...).” (Reusser 1989, S.136) Die Vermischung des Gemeinten und des Verstandenen habe ich bereits diskutiert, hier geht es mir um die Vernachlässigung des Handlungsziels in diesem Prozeß.

In Reussers Arbeit ist das Handlungsziel das Verstehen und Lösen der Textaufgabe. In der Testsituation ist das **Handlungsziel** für den Schüler hingegen zunächst, **das Kreuz richtig zu setzen bzw. die richtige Antwort hinzuschreiben**. Dieses Handlungsziel ist getrennt vom Handlungsziel der Lösung des durch die Testaufgabe gestellten Problems zu betrachten. Das vernachlässigen Baumert u.a. in ihrer Diskussion der Aufgaben. Es gibt Testaufgaben, die durch eine ernsthafte Repräsentation als Testaufgabe unlösbar würden (man muß also Zusatzbedingungen kennen, z.B. über spezifisches Funktionieren von Schulaufgaben) oder in denen die Repräsentation des Problems viel aufwändiger wäre als eine sofortige mathematische Verarbeitung ohne inhaltliches Verständnis. Multiple Choice verändert den Charakter der Aufgabe zusätzlich. Wenn dann noch die Probleme des Auseinanderlaufens von latentem und manifestem Text hinzukommen, ergibt sich eine Vielzahl von möglichen Konstellationen, in denen die Intention des Aufgabenstellers und der wirkliche Aufgabentext auseinander laufen.

Schon bei Aufgabe D2 deutet sich an, dass die Lösung des durch die Aufgabe gestellten Laborproblems gar nicht gefragt ist. Dies wird dort dadurch abgeschwächt, dass schon textlich das Laborproblem teilweise zu einem Alltagsproblem verändert wird. Es wird sich aber zeigen, dass es Aufgaben gibt, bei denen das Handlungsziel der Testbewältigung und das Handlungsziel der Problemlösung noch weiter voneinander entfernt sind. Ein denkbares, vernünftig erscheinendes Ziel der Erstellung von Testaufgaben wäre es, als Handlungsziel das Verstehen und die Lösung des Problems, welches die Aufgabe beschreibt, anzusehen. Der Schüler sollte dann das Kreuz nur richtig setzen können, wenn er dieses Handlungsziel erreicht.[21] Diese Arbeit wird erste Eindrücke zu vermitteln haben, ob solche Aufgabenkonstruktionen möglich sind.

Baumert u.a. interpretieren als Handlungsziel eher das „richtige Setzen des Kreuzes” statt der Lösung des Problems, welches die Aufgabe stellt. Das zeigt sich darin, dass ihre gesamte Diskussion der zu konstruierenden mentalen Situationsmodelle auf das Verstehen der Aufgabe im Sinne der Aufgaben-

[21] Da bei TIMSS der Testkonstruktion kein didaktisches Konzept zugrunde lag, ist dieses jeweilige Handlungsziel schwierig mit dem Aufgabentext zu vergleichen, weil es – in seiner je eigenen Widersprüchlichkeit – erst aus dem Text erschlossen werden muß. Auch daran mag es liegen, dass Baumert u.a. in ihren Aufgabeninterpretationen immer zu erraten scheinen, was die Aufgabe eigentlich testen soll.

steller zielt, unabhängig davon, ob der Aufgabentext dies überhaupt ermöglicht. Dadurch entsteht die Differenz mit Hagemeister, der das gestellte Problem (zu) ernst nimmt.

Noch schärfer als bei D2 zeigt sich die Konzentration auf die Intention der Aufgabensteller in der Diskussion um die Aufgabe M12 (Baumert u.a. 2000, S.201-206). Diese Aufgabe ist fachlich fehlerhaft, da sie vom Schüler die Anwendung des Ohmschen Gesetzes für Glühlampen verlangt, obwohl es für sie nicht gilt. Baumert u.a. räumen diesen Fehler ein. Sie behaupten aber trotzdem – selbst wenn sie später die Streichung der Aufgabe empfehlen, dass die Aufgabe geeignet sei, Wissen über das Ohmsche Gesetz zu testen. Sie erkennen auch nicht, dass die Aufgabe zu einem guten Teil lediglich proportionales Denken – ohne dass man den Begriff der Proportion kennen müßte – testet, nicht aber die Kenntnis des Ohmschen Gesetzes.

Ich habe bereits auf die bezüglich der Berücksichtigung des Handlungsziels leistungsfähigere Perspektive der situierten Kognition verwiesen. Hier versteht man „unter einem mentalen Modell ... die Repräsentation einer ganz spezifischen Anforderungssituation. Mentale Modelle sind nicht als Abbilder der Realität, sondern als auf das *Handlungsziel bezogene Ausschnitte* aus der Realität zu verstehen.” (Stern 1998, S.30)

5. Fähigkeitsniveaus als nachträglicher Ersatz für ein fachdidaktisches Konzept

TIMSS wurde in Deutschland von Testfachleuten durchgeführt, aber kaum von Fachdidaktikern begleitet. Ein fachdidaktisches Grundkonzept gibt es nicht. Umso mehr verwundert die Aussage: „Eine gute Testaufgabe enthält in sparsamer Form die notwendigen Hinweisinformationen, die bei Probanden, die das durch die Aufgabe indizierte Fähigkeitsniveau erreichen oder übertreffen, zur Bildung eines für die Lösung der Aufgabe adäquaten Situationsmodells führen.” Hier wird einerseits so getan, als ob das Konstrukt der Fähigkeitsniveaus bei der Testaufgabenkonstruktion schon fertig gewesen wäre und die Testaufgaben in dieses System hineinkonstruiert worden wären. Die Aufgaben wurden aber aus einer Aufgabendatenbank entnommen und lediglich nach statistischen und curricularen Kriterien evaluiert. Die Fähigkeitsniveaus wurden hinterher „draufkonstruiert“. Andererseits wird so getan, als ob die Aufgabentexte verändert worden wären, um die Aufgabenschwierigkeit zu variieren. Das wäre innerhalb der hier vorliegenden Art der Testkonstruktion aber sinnlos, da es lediglich darum ging, Aufgaben aller Gebiete grob auf alle Schwierigkeitsstufen zu verteilen. Gezielte Manipulationen wären hier nur möglich gewesen, wenn man von vornherein die gleiche Aufgabe mit verschiedenen „Hinweisinformationen” in den Vortest mit aufgenommen hätte. Fachdidaktisch wäre das auch interessant gewesen, aber hier wurde nicht fachdidaktisch gearbeitet. Angesichts der Hagemeisterschen Kritik scheinen Baumert u.a. lediglich den Eindruck erwecken zu wollen, als ob die fachdidaktische Durchdringung der Testaufgaben tiefer war, als sie es in Wirklichkeit war.

Kapitel 2: Zur Methodologie und Praxis der Objektiven Hermeneutik

2.1. Methodologische Bemerkungen zur Objektiven Hermeneutik[22]

In diesem Kapitel werden einige systematisierende Gedanken zur Methodologie der Objektiven Hermeneutik und zur Praxis objektiv-hermeneutischen Arbeitens dargelegt, um

- das methodische Vorgehen dieser Arbeit zu reflektieren und weitergehend als im vorigen Kapitel zu begründen
- methodologische Grundlagen der Objektiven Hermeneutik vorzustellen
- die Grenzen der Aussagekraft der Erkenntnisse dieser Arbeit auszuloten.

Mit dieser Arbeit wird die Methode der Objektiven Hermeneutik neu in die Mathematikdidaktik eingeführt. Deshalb müssen ihre Vorannahmen, ihre Erschließungskraft und ihre Grenzen besonders sorgfältig herausgearbeitet werden. Der Charakter der Objektivität der Objektiven Hermeneutik und ihres Verhältnisses zur Subjektivität von Texterstellern und Textrezipienten ist zu diskutieren. Es geht mir also nicht um eine umfassende methodologische Entfaltung der Methode, sondern um eine Einführung in ihre Grundgedanken und um Antworten auf Fragen und Zweifel aus der mathematikdidaktischen Community. Die häufig zu verzeichnende Widerständigkeit unseres Alltagsdenkens gegen das Vorgehen der Objektiven Hermeneutik fordert zusätzlich dazu auf zu kennzeichnen, welche Teile der hier generierten Erkenntnisse auch ohne bestimmte „starke“ Vorannahmen zustande kommen können: So kommt zum Beispiel die Untersuchung der *Lösungswege* der Aufgaben gänzlich ohne Objektive Hermeneutik aus.

Oevermann hat die Objektive Hermeneutik nicht nur aus ihrer Methodologie heraus, sondern auch bewußt in ihrem Charakter als Kunstlehre entwickelt. Er verweist mit diesem Begriff einerseits auf die Nichtstandardisierbarkeit des forschungspraktischen Vorgehens - mit jedem neuen Forschungsbeitrag entstehen neue methodische Varianten. *Kunstlehre* bedeutet aber auch, dass sich im Laufe der empirischen Erfahrung mit der Methode Ausschärfungen und Korrekturen ergeben haben. So wurde z.B. anfangs angenommen, dass man nur in der Gruppe interpretieren kann, um eine Vielfalt von Geschichten zu gewährleisten und ihre Wohlgeformtheit zu überprüfen. Man sprach also genügend umfassende (Sprach-)Regelkenntnis nur dem Kollektiv zu. Es stellte sich aber heraus, dass das Regelwissen des einzelnen Mitglieds der Sprachgemeinschaft genügt, um eine regelbasierte Interpretation zu gewährleisten. Dieser - im übrigen jeder Methode eigene - Charakter als Kunstlehre schlägt sich in der gelegentlichen Verschränkung von methodologischen und Kunstlehreargumenten in den folgenden Darlegungen nieder.

[22] Derzeit fehlt noch eine systematische und vollständige Einführung in die Methodologie der Objektiven Hermeneutik, Hans-Josef Wagners Buch „Objektive Hermeneutik und Bildung des Subjekts“ (2001) liefert aber eine (etwas sperrige) Darstellung methodologischer Grundlagen. Weiterhin liegt die interpretationspraktisch orientierte „Einführung in die Interpretationstechnik der Objektiven Hermeneutik“ von Andreas Wernet (2000) vor. Ich beziehe mich hier darauf und auf Aufsätze von Oevermann (1983, 1993, 1996).

Die Objektive Hermeneutik ist eine sinnverstehende Methode zur Interpretation von Texten. Sie trifft Aussagen zunächst nur über die „objektive Bedeutung“ eines Textes, eine durch die Geltung von sozialen Regeln (in dieser Arbeit: Sprachregeln) klar eingegrenzte Entität. Eine objektiv-hermeneutische Interpretation beansprucht Geltung und Objektivität also nur innerhalb des „Raumes“ dieser Regeln. Dieser Raum wird im Abschnitt „Regeln“ ausgeleuchtet.
Das Interesse empirischer Forschung erschöpft sich aber meist nicht darin, die „objektive Bedeutung“ von Texten zu erfassen. Dahinter steht oftmals das Bedürfnis, etwas über den Textersteller bzw. die Texterstellung zu erfahren. Das Herausarbeiten dieser Aussagen aus der „objektiven Bedeutung“ ist eine eigene Operation, für die auch die Objektive Hermeneutik keine methodische Kontrolle bereitstellt. Diese Operation wird im Abschnitt „Subjektivität“ diskutiert.

In dieser Arbeit geht es nicht vorrangig um die Untersuchung der Ersteller bzw. Erstellung von Testaufgaben. Die Analyse des objektiven Aufgabentextes soll vorrangig Hinweise auf die *Rezeption* der Aufgaben geben. Das ergibt sich aus dem Charakter von Testaufgaben als Meßinstrument: Man will Fähigkeiten vermessen, will also Aussagen über subjektive kognitive Dispositionen treffen. Jede Testtheorie muß dabei davon ausgehen, dass die jeweiligen Items (also eigentlich die objektive Bedeutung des Itemtexts, der „objektive Text“) mit den zu vermessenden subjektiven Dispositionen „wechselwirken“, so dass man aus dem Tun des Subjekts, also aus seinem Umgang mit dem Item, auf die Dispositionen rückschließen kann. Ich habe bereits im vorigen Kapitel bei der Kritik des kognitiv-propositionalen Ansatzes beschrieben, dass ich die *Textrezeption als Wirkung des objektiven Textes auf die individuellen psychischen Dispositionen* verstehe. Der propositionale Ansatz ignoriert die latente Textebene und kann deshalb auch das Zusammenwirken von latenter und manifester Textebene nicht analysieren. Dort, wo die latente Textebene allzu offensichtlich ist, werden durch nichtkontrollierbare Nebenannahmen wie z.B. das Vorhandensein von „Alltagswissen“ oder „Weltwissen“ beliebig Propositionen zu den Propositionen der manifesten Textebene hinzugefügt. Das Funktionieren von Sprache wird auf diese Weise imitiert, statt - wie behauptet - simuliert.
Die Analyse des objektiven Textes erzählt nie direkt etwas über psychische Wirkungen dieses Textes, sondern nur über die im Rahmen von Regelhaftigkeit im Text steckenden *Möglichkeiten*, auf die Psyche zu wirken. Erst die Untersuchung der konkreten Wirkung erzählt dann etwas über das Individuum, also über die psychischen Dispositionen, auf die der Text gestoßen ist und auf die er gewirkt hat. Diese Trennung von objektivem Text und Psyche macht einen fundamentalen Unterschied zwischen nachvollziehenden Hermeneutiken und Objektiver Hermeneutik aus. Man kann davon ausgehen, dass sowohl in den Testerstellungsgremien von TIMSS und PISA als auch bei der Deutung der Ergebnisse mit Alltagshermeneutiken oder anderen nachvollziehenden Hermeneutiken gearbeitet wurde.[23] Hier

[23] Baumert u.a. (2000) behaupten zwar, die Aufgaben nach dem propositionalen Ansatz zu interpretieren. Dies wurde aber im vorigen Kapitel widerlegt. Auch bei der Testerstellung wird bereits lediglich mit Alltagshermeneutiken gearbeitet: Experten beurteilen und verändern die aus Aufgabendatenbanken gewonnenen Aufgaben. Praktisch erhält so ein Experte einen Stapel Testaufgaben und soll entweder ihre Eignung als Testaufgabe bewerten oder ihre Paßfähigkeit zu bestimmten Curricula bestätigen oder eine Zuordnung zu Aufgabenkategorien vornehmen (siehe z.B. Bericht von Hagemeister 1999). Diese „Deutungsprozesse“ verlaufen ohne methodische Kontrolle und erhalten ihre Legitimation offenbar durch eine Mischung aus Anzahl der beteiligten Experten, starkes Sieben der Aufgaben, Internationalität der Experten und der Aufgaben, durch „empirische Bewährung“

gibt es einerseits nur eine unsystematische Wahrnehmung latenter Textelemente. Andererseits gibt es keine klare Trennung von objektivem Text als klar eingegrenzte und bestimmbare Entität und „subjektivem Text“, welcher erst in der *Wirkung des objektiven Textes in der Psyche* entsteht. Mit dieser klaren Trennung wird sofort deutlich, dass für die Beurteilung eines Meßinstruments der objektive Aufgabentext unabdingbare Grundlage ist. Erst mit seiner Kenntnis ist ein subjektiver Aufgabentext sinnvoll zu erfassen und zu deuten.

In den im Folgenden rezipierten methodologischen Texten wird die Trennung von objektiven Bedeutungsstrukturen von Texten und von subjektiven Strukturen immer wieder thematisch, aber in der Richtung der Erforschung der Fallspezifik des Texterstellers aus dem von ihm produzierten Text - denn vorrangig in dieser Blickrichtung ist Objektive Hermeneutik entstanden. Die *Rezeption* des objektiven Textes wird in den Texten zur Begründung der Objektiven Hermeneutik vor allem bei der Interpretation von Dialogen oder Trialogen thematisch. Dort zeigt sich, dass latente *und* manifeste Textebene wirken, dass also insbesondere auf den latenten Text auch geantwortet wird - ein empirisches Faktum, das wiederum in propositionalen Ansätzen keine Erklärung findet.

2.1.1. Latente und manifeste Sinnstrukturen, objektive Bedeutung

> „Zentraler Gegenstand der Methodologie der objektiven Hermeneutik sind die latenten Sinnstrukturen und objektiven Bedeutungsstrukturen von Ausdrucksgestalten, in denen sich uns die psychische, soziale und kulturelle Erfahrungswelt präsentiert. Latente Sinnstrukturen und objektive Bedeutungsstrukturen sind jene abstrakten, d.h. sinnlich nicht wahrnehmbaren Gebilde, die wir alle mehr oder weniger gut und genau „verstehen“, wenn wir uns verständigen, Texte lesen, Bilder und Bewegungsabläufe sehen, Ton- und Klangsequenzen hören, und die durch bedeutungsgenerierende Regeln erzeugt werden und unabhängig von unserer je subjektiven Interpretation objektiv gelten. Die objektive Hermeneutik ist ein Verfahren, diese objektiv geltenden Sinnstrukturen intersubjektiv überprüfbar je konkret an der les-, hör- und sichtbaren Ausdrucksgestalt zu entziffern.“
> (Oevermann 1996, S.1)

In der vorliegenden Arbeit sind die betrachteten Ausdrucksgestalten Testaufgaben. Wie in jedem Text liegen auch in ihnen latente Sinnstrukturen, also nicht unmittelbar sichtbare textliche Strukturen, die sinndarstellend und sinnbildend sind und wirken. Sie ergeben im Zusammenspiel mit den manifesten - also äußerlich sichtbaren und dem Text Inhalt und Sinn gebenden - Textstrukturen die objektive Bedeutungsstruktur.

Man kann den latenten und den manifesten Sinnstrukturen von Testaufgaben nun nicht direkt Klassen von Kompetenzen zuordnen, etwa in dem Sinne, dass mit impliziten Kompetenzen die latenten Sinn-

und dadurch, dass möglichst viele Beteiligte möglichst häufig auf die Aufgaben blicken (siehe z.B. Baumert, Lehmann u.a. 1997, S.47 ff.) .

strukturen und mit expliziten Kompetenzen die manifesten Sinnstrukturen bearbeitet werden. Erst die objektive Bedeutungsstruktur, also latenter und manifester Sinn in ihrem Zusammengehen, erzählen etwas über die mit einer Aufgabe gemessenen Kompetenzen. Man kann auch nicht sagen, dass die latenten Sinnstrukturen immer unbewußte und die manifesten Sinnstrukturen bewußte Elemente des Denkens des Texterzeugers (oder gar des Textrezipienten) sind. Das mag häufig zusammenlaufen, aber man muß zunächst klar zwischen der Sinnstruktur eines Textes und subjektiven Verfaßtheiten des Texterstellers unterscheiden. Es wird uns aber nicht überraschen, wenn sich Unbewußtes mit der latenten Sinnstruktur verbindet. So werden sich Zerstörungen des Mathematischen und Textelemente, die moralischen Grundsätzen oder didaktischen Ideologien widersprechen, meist erst mit der latenten Sinnstruktur erschließen lassen.

Das von Oevermann beschriebene „mehr oder weniger genaue Verstehen" der latenten Sinnstrukturen und objektiven Bedeutungsstrukturen bedeutet, dass jeder Textleser die objektive Bedeutungsstruktur potentiell ebenfalls eruieren kann - auch ohne eine speziell darauf zugeschnittene Methode. Auch ohne Objektive Hermeneutik erschließt sich die rechnerische Sinnhaftigkeit, aber auch die Karikatur in der folgenden Aufgabe: *Eineinhalb Hühner legen in eineinhalb Tagen eineinhalb Eier. Wieviel Eier legen zweieinhalb Hühner in zweieinhalb Tagen?* Ebenso werden die meisten Menschen das Problem des folgenden Dialogs dechiffrieren können, obwohl die manifeste Textstruktur dieses Problem nicht aufzeigt (Interpretation siehe Wernet 2000, S.47 ff.):

> „Schüler: Wann geben Sie uns die Klassenarbeiten wieder?
> Lehrer: Nächste Woche.
> Schüler: Oh, Sie haben sie doch schon 3 Wochen.
> Lehrer: Und wenn ich sie 5 Wochen hätte.
> Schüler: Meine Mutter denkt schon, ich hätte die weggeschmissen."

Der Sinn eines solchen Dialogs erschließt sich, weil man als Mitglied einer Sprachgemeinschaft über bewußtes und unbewußtes Regelwissen verfügt und weil Regelkompetenz nicht an die Fähigkeit der Regelexplikation gebunden ist. Damit ergibt sich die Möglichkeit, das inhärente (Entgrenzungs-, Beziehungs- bzw. Kommunikations-) Problem zu beschreiben. Deshalb erschließen auch nachvollziehende o.a. Hermeneutiken durchaus latente Sinnstrukturen von Texten. Objektive Hermeneutik ermöglicht aber eine **systematische** Dechiffrierung auch der latenten Textebene und damit auch des *Zusammenspiels* von manifester und latenter Textebene, also der objektiven Bedeutungsstruktur des Texts. Sie ermöglicht außerdem eine **methodische Kontrolle** und damit eine **intersubjektive Überprüfung** dieser Dechiffrierung.

2.1.2. Strukturalismus und Sequentialität in der Objektiven Hermeneutik

Schon der Sprachgebrauch „Sinnstrukturen" und „Bedeutungsstrukturen" nimmt die mit der Objektiven Hermeneutik vertretene strukturalistische Grundposition auf. Es geht in der Objektiven Hermeneutik stets darum, die Struktur von Entitäten (und zwar die Struktur von Lebenspraxis wie auch die

Struktur der von dieser Lebenspraxis erzeugten Texte) zu erfassen, die man also als strukturiert voraussetzt. Speziell auf Texte bezogen heißt das, dass ein Text eine Struktur hat, und zwar eine Bedeutungsstruktur, die auf die Strukturiertheit der diesen Text erzeugenden Lebenspraxis verweist. - Texte werden also als Bedeutungsträger und als sinnhaft seiend angenommen. Bedeutungshaftigkeit und Sinnhaftigkeit sind für sich genommen nun noch kein Grund, Texte oder Lebenspraxis als strukturiert anzusehen. In der methodologischen Grundlegung der Objektiven Hermeneutik sind aber die Begriffe Sinn, Bedeutung und Struktur immer unmittelbar aufeinander bezogen.

Diese Strukturannahme kann man sich verdeutlichen, wenn man die Entstehung (bzw. spiegelbildlich die Rezeption, siehe auch Abschnitt „Subjektivität“) von Texten betrachtet: Der Sprecher hat bei der Textproduktion an jeder Textstelle quasi unendlich viele Möglichkeiten, den Text fortzusetzen, er kann unendlich viele Wörter oder Buchstabenkombinationen an die nächste Textstelle setzen oder er kann einfach abbrechen. Nur sehr wenige dieser unendlich vielen Möglichkeiten sind bedeutungshaltig, und ihr Sinn ergibt sich nicht von allein, sondern weil sie innerhalb einer Sprache als bedeutungshaltig festgelegt sind - durch Regeln. Die *Struktur eines Textes* kann man sich nun folgendermaßen vorstellen: Der Sprecher beginnt zu sprechen, er produziert eine Bedeutungseinheit (Sequenz). Am Ende dieser Bedeutungseinheit muß er aus den unendlich vielen Möglichkeiten der Fortführung (davon endlich viele bedeutungsbehaftete Möglichkeiten) eine auswählen und die zweite Bedeutungseinheit produzieren. Dann muß er die dritte Bedeutungseinheit produzieren usw. Es zeigt sich sofort, dass mit der Textstruktur eine Bedeutungsstruktur verbunden ist, denn jede Sequenz trägt ja Bedeutung und die Entscheidung an jedem Übergang von einer Sequenz zur nächsten („Sequenzknoten“) trägt auch eine Bedeutung. Da diese Bedeutung durch Sprachregeln überhaupt erst produziert wird, kann sie mit Hilfe eben dieser Regeln auch reproduziert, also erfaßt werden. Genau das tut Objektive Hermeneutik, und in diesem Sinne steckt hinter objektiv-hermeneutischem Vorgehen immer strukturalistisches Denken.

Die beschriebene Strukturiertheit eines Textes zieht zwingend das *Prinzip der Sequentialität der Textinterpretation* nach sich: Wenn die Bedeutungsstruktur eines Textes aus den je spezifischen Entscheidungen an den Sequenzknoten abgeleitet wird, dann muß man so eine „Knotenentscheidung“ auch wirklich deuten. Nichtsequentiell arbeitende Hermeneutiken verzichten auf die konsequente Ausdeutung einer Knotenentscheidung, sie generieren aus einer Textstelle eine Hypothese und suchen dann im Text nach Beleg- oder Widerlegungsstellen für diese Hypothese. Sie benutzen den Text also als eine Art Selbstbedienungsladen oder Steinbruch für Argumente für oder gegen ihre Hypothese. Forschungspsychologisch birgt das die Gefahr eines auf die Hypothese verengten Blicks in sich: Von der Selbstkritikfähigkeit des Forschers hängt ab, ob der Text noch eine Chance hat, sich gegen die Hypothese durchzusetzen. Schwerer wiegt, dass auf diese Weise manifeste Textelemente eher wahrgenommen werden als latente, dass die Chance auf eine konsequente, detaillierte und erkenntnisreiche Feinanalyse der Knotenentscheidung verspielt wird und dass keine Strukturgesetzlichkeit erarbeitet wird. Man erkennt aber, warum auch nichtsequentiell arbeitende Hermeneutiken wichtige Resultate herausarbeiten können: Ein Forscher, der für das intuitive Erfassen latenter Bedeutungs- und Sinnstrukturen sensibilisiert ist, erfaßt die Fallstruktur, weil er die Gemeinsamkeiten der Entscheidungen an verschiedenen Sequenzknoten erfassen, strukturieren und formulieren kann; er will und kann das hinter dem Offensichtlichen Liegende erkennen (und bedient sich dabei seiner Regelkenntnis); und er arbeitet selbstkritisch mit mehreren konkurrierenden Fallstrukturhypothesen, die er in einem inneren Falsifikationsverfahren gegeneinander stellt. Objektive Hermeneutik ermöglicht nun durch ihre klare Schrittfolge bei der Interpretation den systematischen Nachvollzug und damit die methodische Kontrolle des interpretatorischen Vorgehens.

Bei der Sequenzanalyse im Sinne der Objektiven Hermeneutik wird

> „grundlegend zwischen <u>zwei ganz verschiedenen Parametern in der Determination von Sequenzen unterschieden</u>. Ein <u>erster Parameter</u> besteht aus dem Gesamt an Sequenzierungsregeln, durch die an einer je gegebenen Sequenzstelle die sinnlogisch möglichen Anschlüsse erzeugt werden und auch die je möglichen sinnlogisch kompatibel vorausgehenden Handlungen festgelegt sind und entsprechend erschlossen werden können. Diesen Parameter muß man sich vorstellen als eine <u>Menge von algorithmischen Erzeugungsregeln sehr unterschiedlichen Typs</u>. Dazu gehören z.B. ganz elementar die Regeln der sprachlichen Syntax, aber auch die pragmatischen Regeln des Sprechhandelns und die logischen Regeln für formale und für material-sachhaltige Schlüssigkeit. Dieses Gesamt an Sequenzierungsregeln erzeugt an jeder Sequenzstelle je von Neuem einen Spielraum von Optionen und Möglichkeiten, aus denen dann die in diesem Praxis-Raum anwesenden Handlungsinstanzen per Entscheidung eine Möglichkeit auswählen müssen. Welche Auswahl konkret getroffen wird, darüber entscheidet ein <u>zweiter Parameter</u>, der alle Komponenten und Elemente der Disponiertheit der verschiedenen beteiligten Lebenspraxen oder Handlungsinstanzen umfaßt. Das <u>Gesamt der Dispositionen einer je konkreten Lebenspraxis</u> macht deren Eigenart oder deren Charakter aus.“ (Oevermann 1996, S.5 f.)

Die Explizierung dieser zwei Parameter der Textproduktion sorgt für zweifache Erkenntnis durch die objektiv-hermeneutische Interpretation: Sie erzählt nicht nur etwas über das Individuum bzw. die konkrete Lebenspraxis, sondern läßt uns durch die Explizierung der möglichen Anschlüsse auch die Regeln selbst erkennen. Die Trennung der Betrachtung der Anschluß*möglichkeiten* von dem vom Textersteller dann *wirklich gewählten* Anschluß stützt den Anspruch, durch die Textinterpretation dann auch eine Aussage über den Textersteller treffen zu können.

Als einfaches Beispiel kann die Interpretation der PISA-Aufgabe „Dreieck“ (siehe Abschnitt 5.6.) dienen, weil dort nicht Sprachregeln, sondern geometrische Bezeichnungskonventionen für Punkte diskutiert werden. Verkürzt kann man den ersten Parameter so beschreiben: Man kann Punkte bezeichnen oder nicht bezeichnen. Wenn man sie bezeichnet, kann man sie innerhalb der regelhaften Bezeichnungskonventionen mit lateinischen Großbuchstaben bezeichnen oder anders. Man muß eine Entscheidung darüber treffen, mit welchem Buchstaben man anfängt, dabei gibt es verschiedene Möglichkeiten mit verschiedenen Bedeutungen (z.B. ABCD..., MNO..., PQR..., besondere Bedeutung von P (Punkt), M (Mittelpunkt), Z (Zentrum)) Außerdem ist es regelhaft, zusammengehörende Gebilde (meist in bestimmten Reihenfolgen) in geschlossener Buchstabenfolge zu bezeichnen. Der zweite Parameter ist nun die konkrete Entscheidung der testerzeugenden Lebenspraxis für eine bestimmte Bezeichnung. Im vorliegenden Fall wird im ersten Aufgabenteil eine regelhafte Bezeichnungspraxis gewählt, im zweiten Teil wird die Geschlossenheit der Buchstabenfolge verweigert. In dieser konkreten Entscheidung liegt die Eigenart der zu untersuchenden Lebenspraxis. Der aufgrund von Regelhaftigkeit methodisch kontrollierbare Teil der Interpretation endet bei der Feststellung des durch die Regelverletzung hervorgerufenen Bruchs.

> „Beide Parameter müssen analytisch klar unterschieden, aber gleichermaßen berücksichtigt werden. Mit dieser Unterscheidung hebt sich die Sequenzanalyse der objektiven Hermeneutik deutlich von anderen Analyseverfahren der Sozialwissenschaften ab, in denen, wenn die Sequentialität überhaupt berücksichtigt wird, die beiden Parameter im Begriff der Erwartung und der Erwartungs-Erwartung, also zu unserem zweiten Parameter zusammenfließen, so dass eine von der subjektiven Perspektive der beteiligten Handelnden abstrahierende, objektivierende Analyse der Sequentialität nicht mehr möglich ist.
>
> Die Leistungen der Sequenzanalyse sind vielfältig und für die objektive Hermeneutik insgesamt zentral:
>
> 1. Indem an jeder Sequenzstelle auf der Ebene des Parameters I die je eröffneten Möglichkeiten gedankenexperimentell expliziert werden müssen, bevor man sich anschaut, welche dieser Möglichkeiten faktisch eingetreten ist, gewinnt man die Folie, auf der der tatsächliche Sequenzablauf, der sich ja immer in Abhängigkeit vom Parameter II als vollzogene Auswahl bzw. Entscheidung unter den Alternanten ergibt, seine fallspezifische, präzise Kontur und Bedeutung. [...]
>
> 2. Die Sequenzanalyse schmiegt sich dem realen humansozialen Geschehen in seiner Grundstruktur an [[24]] und ist deshalb nicht, wie die sonst üblichen Meß- und Klassifikationsverfahren, eine dem Gegenstand äußerliche Methode, sondern eine der Sache selbst korrespondierende und ihr gemäße. Tatsächlich muß auch im praktischen Leben im Prinzip an jeder Sequenzstelle unter den noch offenen Optionen in eine offene Zukunft entschieden werden.
>
> 3. In die Sequenzanalyse ist gewissermaßen eine permanente Falsifikation eingebaut. Denn an jeder nächsten Sequenzstelle kann grundsätzlich der Möglichkeit nach die bis dahin kumulativ aufgebaute Fallrekonstruktion sofort scheitern. Ein strengeres Falsifikationsverfahren ist in der Methodologie der Humanwissenschaften schlechterdings nicht denkbar.“ (Oevermann 1996, S.8 f.)

Mit den Reflexionen zur Selektivität einer Lebenspraxis stellt sich die Frage, welche Lebenspraxis in Testaufgaben protokolliert ist: Es ist jene Praxis, die an jedem Selektionsknoten „eine Entscheidung trifft“. Wir können sie uns zum Teil in konkreten Handlungen des Testerstellungskollektivs vorstellen, aber auch diese soziale Struktur ist ja wiederum in andere soziale Strukturen eingebettet. Bestandteil dieser konkreten kollektiven Praxis ist nicht nur derjenige, der die Aufgabe erstmals entwirft. Es ist auch jeder, der eine Änderung anbringt. Es ist jeder, der eine Änderung vorschlägt, welche nicht ange-

[24] „Konstitutionstheoretisch ist die Sequentialitäts-Thematik das Schlüsselkonzept einer vor allem an Mead und Peirce entfalteten Theorie der Entstehung des Neuen, einer Theorie der Lebenspraxis und ihrer Zukunftsoffenheit und einer Theorie der Bildungsprozesse.“ (Wernet 2000, S. 27; zusammenfassend Wagner 2001, S. 14-32)

nommen wird. Es sind diejenigen, die eine vorgeschlagene Änderung annehmen oder verwerfen. Und es sind auch diejenigen, die keine Änderung vornehmen, obwohl sie in den Testerstellungsprozeß involviert sind. Es liegt also ein direktes Protokoll von Testerstellungspraxis vor, und die Selektionen sind kollektiv durchgeführte. Das heißt nun nicht, dass man sich hier einfach eine Art „Testerstellungsstruktur“ kollektiv verdinglicht vorstellen kann und nur diese Struktur sich in der Textstruktur der Testaufgaben wiederfindet. Auch das Testerstellungskollektiv, das uns ein Protokoll seines Tuns in Form von Testaufgaben hinterläßt, ist ja wiederum in eine soziale Struktur um Wissenschaft und Bildung eingebunden, die hier ebenfalls mitprotokolliert ist. Deshalb erzählen uns Testaufgaben ebenso etwas über das Testerstellungskollektiv wie über Wissenschafts- und Bildungssysteme.

Die spezifischen Selektionen an den Sequenzknoten zeigen auch die Textwirkung an: Die Selektionen teilen sich latent ja auch dem Rezipienten mit. Da die Selektionen auf der für den Autor wie für den Leser gleichermaßen gültigen Regelhaftigkeit von Sprache beruhen, teilt sich auch der Sinn des Textes über genau diese Selektionen mit. Und so wie sich an jedem Selektionsknoten eine Entscheidung vollzieht, und genau diese Entscheidung etwas über die texterzeugende Struktur erzählt (und genau diese Entscheidung die objektive Sinnstruktur des Textes erzeugt), so ist es auch umgekehrt beim Rezipienten: An jedem Selektionsknoten teilt sich ihm etwas mit - nämlich die objektive Sinnstruktur des Texts. Damit trifft er aber auch an jedem Selektionsknoten eine (natürlich unbewußte) Entscheidung darüber, wie er die konkrete Entscheidung der texterzeugenden Struktur deutet. Diese Deutung - also quasi die unbewußte Textinterpretation - erfolgt vor dem Hintergrund der jeweiligen subjektiven Verfaßtheit des Rezipienten. Die objektive Textstruktur trifft also auf die subjektive kognitive Struktur des Rezipienten. Deshalb ist es auch nicht möglich, dass die objektiv-hermeneutische oder irgendeine andere Interpretation *zwingende* Aussagen über die Wirkung der von ihr herausgearbeiteten Textstruktur in der Testsituation treffen kann. Nicht nur bezüglich der Erzeugungsstruktur ist lediglich die Angabe von Transformationsspielräumen (siehe Abschnitt 3.6. „Aufgabeninterpretationen und Vorhersage von Lösungspraxis“) möglich, dies gilt auch für die subjektive kognitive Struktur und damit erst recht für das Aufeinandertreffen beider Strukturen. Damit zeigt sich aber, dass ein Meßvorgang, der subjektive kognitive Dispositionen erfassen will, notwendig sehr kleine Transformationsspielräume anstreben muß - sowohl bezüglich der objektiven Textstruktur als auch bezüglich der subjektiven kognitiven Struktur. Ein erweiterter Transformationsspielraum bedeutet nämlich notwendig eine ungenauere Messung - man weiß dann nicht mehr, welche subjektive kognitive Disposition man eigentlich angesprochen und also gemessen hat.

2.1.3. Grenzen nichtstrukturalistischen Denkens bei der Deutung der Interpretationen von Aufgaben

Da diese Arbeit sich in eine Debatte auch mit Kritikern strukturalistischer Auffassungen begibt, soll hier diskutiert werden, wie weit die Aussagen der Arbeit auch ohne strukturalistisches Denken tragen. Zunächst ist dazu zu bemerken, dass die Generierung der Lösungswege ohne jedes objektiv-hermeneutische Arbeiten und auch sonst ohne strukturalistische Sichtweisen auskommt. Hier wird lediglich stoffdidaktisch dahingehend argumentiert, dass bestimmte schulmathematische Fragestellungen bestimmte Lösungszugänge zulassen. Deshalb ist es möglich, auch ohne Objektive Hermeneutik zu zeigen, dass das Kompetenzstufenmodell von PISA inhaltlich nicht begründet werden kann, weil die verschiedenen Lösungszugänge es unmöglich machen, Lösungshäufigkeiten als Lösungspraxen zu deuten (siehe Meyerhöfer 2004 und Kapitel 5).

Schwieriger wird es bei den objektiv-hermeneutischen Interpretationen selbst. Meiner Erfahrung nach besteht das Grundproblem der Akzeptanz objektiv-hermeneutischen Arbeitens in der mathematikdidaktischen Community nicht bezüglich der Rekonstruktion der Bedeutungsstruktur eines Textes, wie ich sie für die Aufgabe D2 angedeutet habe. Das liegt wahrscheinlich daran, dass man die Kette der Entscheidungen an den Sequenzknoten nicht zwingend als Struktur sehen muß, sondern einfach als Kette von Entscheidungen sehen kann, die uns etwas über den Fall erzählt. Es scheint auch aus Sicht anderer Hermeneutiken sinnvoll zu sein, so (streng sequentiell) vorzugehen - was nicht heißen muß, dass diese Hermeneutiken für sich selbst sequentielles Vorgehen für unabdingbar halten. Akzeptanzprobleme treten oftmals auf, weil aus relativ *kurzen* Textausschnitten eine Strukturaussage getroffen wird: Bei objektiv-hermeneutischem Vorgehen wird zunächst ein Textausschnitt interpretiert, also eine Bedeutungsstrukturhypothese generiert. Dann sucht man einen weiteren Textausschnitt, der möglichst diese Strukturhypothese zu widerlegen oder zu erweitern scheint, und der wird ebenfalls interpretiert usw. Das Verfahren endet bei empirischer Sättigung, also wenn klar wird, dass man keine Textstelle mehr findet, die auch nur so aussieht, als ob die Strukturhypothese erweitert oder widerlegt werden könnte. *Es ist also oftmals nicht der Akt der Strukturrekonstruktion, der auf mangelnde Akzeptanz stößt, sondern der Akt der Bildung der Fallstrukturhypothese.*

Die Interpretation einer einzelnen Aufgabe unterliegt genau diesem Problem nicht. Hier wird nämlich ein Text *in seinem gesamten Umfang* interpretiert. Bei Testaufgaben tritt dieser seltene Fall auf, weil es sehr kurze Texte sind. Die Fallstruktur muß dann nicht mit Hilfe von scheinbar nichtkontrollierbaren[25] Auswahloperationen für weitere Textstellen herausgearbeitet werden. Die Fallstruktur liegt mit der Aufgabeninterpretation bereits vollständig vor, weil mit der objektiven Bedeutungsstruktur des vollständigen Textes der Fall schon vollständig erfaßt ist - solange man als Fall die Aufgabe selbst ansieht. Die Kernfragen der Erarbeitung einer Fallstruktur aus der objektiven Bedeutungsstruktur der Aufgabe und der Falsifizierbarkeit beantworten sich hier unmittelbar: *Die Falsifizierung erfolgt immer schon in der Aufgabe selbst - der Ort, an dem falsifiziert wird, muß also gar nicht begründet werden.* Es liegen mit den Testaufgaben sogar besonders eindrucksvolle Beispiele der Falsifikationskraft der Objektiven Hermeneutik vor, weil sich in allen Interpretationen die Fallstruktur mehrfach zeigt, obwohl sehr kurze Texte vorliegen.

Damit ist nun aber auch deutlich, an welcher Stelle strukturalistisches Denken unabdingbar für weiterführende Aussagen wird: Wenn nämlich empirische Aussagen über mehr als nur die einzelne Aufgabe getroffen werden sollen. Dies betrifft einerseits den jeweiligen Test: *Wie kann man eine Aussage über*

[25] Mir scheint hier eigentlich ein Mißverständnis seitens der Kritiker der Objektiven Hermeneutik vorzuliegen: Auch die Auswahl der Textstellen und die Argumentation der empirischen Sättigung ist nachvollziehbar und kritisierbar, denn sie muß nach explizierten Kriterien erfolgen. Außerdem muß man mit einem anderen Auswahlverfahren zur selben Fallstrukturhypothese gelangen, ansonsten liegt eine fehlerhafte Interpretation vor.

einen Gesamttest treffen, ohne jede einzelne Aufgabe zu interpretieren? Ich zeige dazu bei der Betrachtung von TIMSS und PISA verschiedene Wege auf.[26] Dies betrifft aber ebenso Aussagen über die Texterstellungspraxis, sie muß für eine gültige Aussage über ihre Fallstruktur sowohl in ihrer Breite betrachtet als auch in der Generierung der Strukturaussage einer falsifikatorischen Überprüfung unterzogen werden.

In dieser Arbeit werden begrenzte Gesamtaussagen über TIMSS und PISA getroffen. *Ziel dieser Arbeit ist es aber nicht, diese konkreten Tests in der gesamten Breite des jeweiligen Konstrukts „mathematische Leistungsfähigkeit" zu kritisieren.* Es sollen vielmehr *allgemeine Probleme* aufgezeigt werden, die bisher bei der Erstellung von Testaufgaben übersehen wurden bzw. nicht oder nicht methodisch kontrolliert beseitigt werden konnten: Probleme der genauen Benennung und Operationalisierung des Meßgegenstandes, Probleme der Testfähigkeit, Probleme der latenten Zerstörung des Mathematischen etc. Die in dieser Arbeit aufgezeigten Probleme zeigen die Unbrauchbarkeit von TIMSS und PISA und die Defizite der bisher gebräuchlichen Testerstellungspraxis bereits so deutlich, dass es forschungsökonomisch unvertretbar wäre, eine Gesamtschau dieser bereits historischen Tests vorzunehmen. Deshalb ist für die Kernfragen der Arbeit strukturalistisches Denken noch gar nicht notwendig. All die auftretenden Probleme sind im Prinzip durch ein einziges Beispiel aufzuzeigen. Die Untersuchung mehrerer Aufgaben dient lediglich der empirischen Anreicherung, empirische Sättigung würde dann durch strenges strukturalistisches Vorgehen erreicht. Die Schlagkraft dieser Aussagen hat aber mit dieser empirischen Sättigung zunächst nichts zu tun. Sie ergibt sich allein aus der Möglichkeit, mit Objektiver Hermeneutik methodisch kontrolliert und systematisch auch die latenten Sinnstrukturen von Aufgaben herauszuarbeiten. Mit der Kenntnis der objektiven Bedeutungsstruktur einer einzigen Aufgabe wird bereits deutlich, dass die bisherige Erstellung von Testaufgaben keine kontrollierte Operationalisierung von mathematischer Fähigkeit leistet, dass dies die Nutzbarkeit des Tests als Meßinstrument einschränkt und dass diese Probleme heilbar sind, weil sich während einer objektiv-hermeneutischen Textanalyse bereits Alternativen ergeben oder sich die Begrenztheit der Möglichkeit des Messens zeigt.

2.1.4. Regeln

[26] Bei TIMSS interpretiere ich die „Ankeritems", das sind all jene Mathematikaufgaben, die alle Schüler lösen mußten. Bei PISA wähle ich eine Methode der kontrastierenden Fälle.

Ein Kern empirischen Forschens ist die Geltungssicherung, also das Wissen darum, warum bzw. inwiefern das Erkannte Geltung beanspruchen kann, d.h. also inwiefern und in welchem Sinne die über den realen Gegenstand getroffenen Aussagen wahr sind.[27]

> „Der Geltungsanspruch, den die objektiv-hermeneutische Bedeutungsexplikation erhebt, gründet sich auf die Inanspruchnahme geltender Regeln. *Soziales Handeln konstituiert sich entlang dieser Regeln und die Interpretation der Protokolle dieses Handelns erfolgt unter Rückgriff auf unser Regelwissen*. ... Das Konzept der Regelgeleitetheit geht davon aus, dass jede Handlung, jede soziale Praxis sich in einem Raum regelerzeugter Möglichkeiten bewegt. Die fundamentale Bedeutung der Regelgeleitetheit ist in ihrer *Nichthintergehbarkeit* zu sehen. Die Lebenspraxis kann sich ihr weder entziehen, noch kann sie die Regelgeltung außer Kraft setzen." (Wernet 2000, S.13)

Ich möchte zunächst den Regelbegriff der Objektiven Hermeneutik in seinen Herkunftsrahmen einbetten. Oevermann versteht unter Regeln

> "in scharfer Differenz zu Regularität oder Regelmäßigkeit formal ein Äquivalent zu einem Algorithmus [...], der wie ein "Naturgesetz im Kopf" des regelbefolgenden Handlungsobjekts operiert, ohne in dessen abfragbarem, bewußt verfügbarem Wissensvorrat repräsentiert zu sein und dort als Begründung von praktisch höchst wirksamen Urteilen der Angemessenheit explizit zur Verfügung zu stehen. Obwohl wie ein "Naturgesetz im Kopf" operierend, kann von diesen Regeln innerhalb gewisser Spielräume bewußt abgewichen werden, ohne dass dadurch ihre Geltung in Frage steht. Auf diese Weise ist sowohl ein Unrechts-Bewußtsein als Komplement zu einem Rechtsbewußtsein gewährleistet wie die Möglichkeit strategischer Täuschung und ästhetischer Gestaltung." (Oevermann 1993, S.115)

In dieser Arbeit kann weder das Ringen um einen fruchtbaren Regelbegriff als Grundelement der Erforschung sozialen Handelns dargestellt werden, noch soll der Oevermannsche Regelbegriff im Vergleich zu anderen Regelbegriffen diskutiert werden, da mir die Geltung der hier geleisteten Interpretationen davon unberührt erscheint. Oevermann beruft sich für seinen Regelbegriff zunächst auf Ludwig Wittgenstein. Der beschreibt verschiedene Rollen von Regeln und deutet gleichzeitig an, wie Regelverständnis angeeignet wird: Er beschreibt dabei *sprachliches Handeln* generell *als Spiel* - also als Regelgeleitetes -, weil er von der Regelgeleitetheit sozialen Handelns, insbesondere Sprechhandelns, überzeugt ist:

> „Denken wir doch daran, in was für Fällen wir sagen, ein Spiel werde nach einer bestimmten Regel gespielt! Die Regel kann ein Behelf des Unterrichts im Spiel sein. Sie wird dem Lernenden mitgeteilt und ihre Anwendung eingeübt. - Oder sie ist ein Werkzeug des Spieles selbst. - Oder: Eine Regel findet weder im Unterricht noch im Spiel selbst Verwendung; noch ist sie in einem Regelverzeichnis niedergelegt. Man lernt das Spiel, indem man zusieht, wie Andere es spielen. Aber wir sagen, es werde nach den und den Regeln gespielt, weil ein Beobachter diese Regeln aus der Praxis des Spiels ablesen kann, - ..." (Wittgenstein 1953/2001, §54)

[27] Kurt Reusser beansprucht z.B. keine Geltung seiner theoretischen Aussagen außerhalb der „kognitiven Welt" des Computers. Erst die unreflektierte Übernahme von Reussers Modell durch Baumert u.a. erzeugt ein Geltungsproblem, denn hier wird Geltung für die kognitive Welt von Schülern beansprucht.

John R. Searle, dessen Sprechakttheorie eine weitere methodologische Basis der Objektiven Hermeneutik ist, unterscheidet zwei verschiedene Arten von Regeln: konstitutive und regulative. Diese Unterscheidung klärt auch das verbreitete Mißverständnis, soziale Regeln seien mit Normen gleichzusetzen:

> „Die regulativen Regeln können wir zunächst als Regeln charakterisieren, die bereits bestehende oder unabhängig von ihnen existierende Verhaltensformen regeln - zum Beispiel regeln viele Anstandsregeln zwischenmenschliche Beziehungen, die unabhängig von jenen Regeln existieren. Konstitutive Regeln dagegen regeln nicht nur, sondern erzeugen oder prägen auch neue Formen des Verhaltens. Die Regeln für Fußball oder Schach zum Beispiel regeln nicht nur das Fußball- oder Schachspiel, sondern sie schaffen überhaupt erst die Möglichkeit, solche Spiele zu spielen. [...] Regulative Regeln regeln eine bereits existierende Tätigkeit, eine Tätigkeit, deren Vorhandensein von den Regeln logisch unabhängig ist. Konstitutive Regeln konstituieren (und regeln damit) eine Tätigkeit, deren Vorhandensein von den Regeln logisch abhängig ist.” (Searle 1969/1983, S.54 f.)

Searle beschreibt, dass eine Sprache zu sprechen bedeutet, in Übereinstimmung mit Regeln Akte zu vollziehen: „Die semantische Struktur einer Sprache lässt sich als eine auf Konventionen beruhende Realisierung einer Serie von Gruppen zugrundeliegender konstitutiver Regeln begreifen; Sprechakte sind Akte, für die charakteristisch ist, daß sie dadurch vollzogen werden, daß in Übereinstimmung mit solchen Gruppen konstitutiver Regeln Ausdrücke geäußert werden.” (ebenda, S.59) Die eine spezielle Sprache ausmachenden Konventionen sind Manifestation bzw. Realisierung von Regeln (ebenda, S.61-65).

Nun stellt sich die Frage, ob man auch Regeln folgen kann, *die man gar nicht kennt*, ob man also umgekehrt bei einem Sprecher davon ausgehen kann, dass er wirklich entlang solcher Regeln handelt. Hier muß man klar unterscheiden zwischen der *Kenntnis* einer Regel und der *Fähigkeit, sie zu explizieren*. Regelkenntnis ist Voraussetzung zur Teilnahme am Sprachspiel, ohne Kenntnis von Sprachregeln kann man nicht sprechen. Die Fähigkeit zu Regelexplikation ist hingegen keine Voraussetzung des Sprechens. Der Sprecher kann also entlang von Regeln handeln, die er kennt, ohne sie explizieren zu können. Objektive Hermeneutik zieht ihren Geltungsanspruch aus der Geltung von Regeln, muß also erklären, warum sie im Verlauf ihrer Lesartenbildung Regeln explizieren kann, die der Sprecher zwar kennt, aber eventuell selbst nicht explizieren kann und die auch der Forscher vor der Lesartenbildung oft nicht in explizierter Form kannte. Searle beschreibt eine Regelexplikation, die ähnlich abläuft wie bei objektiv-hermeneutischen Lesartenbildungen und folgert dann:

> „Um einen Ausschnitt menschlichen Verhaltens adäquat erklären zu können, müssen wir manchmal voraussetzen, dass es in Übereinstimmung mit einer Regel geschah, und zwar selbst dann, wenn der Handelnde selbst nicht in der Lage ist, die betreffende Regel anzugeben, und sich vielleicht nicht einmal der Tatsache bewußt ist, daß er in Übereinstimmung mit einer Regel gehandelt hat. Daß der Handelnde weiß, wie das und das getan werden muß, ist in bestimmten Fällen nur auf Grund der Hypothese adäquat erklärbar, daß er eine Regel des und des Inhalts kennt (erworben, internalisiert oder gelernt hat); und das gilt auch in den Fällen, in denen der Handelnde in einem we-

> sentlichen Sinne nicht weiß, daß er die Regel kennt, oder daß er das, was er tut, teilweise auf Grund der Regel tut. Zwei charakteristische Merkmale, durch die sich regelgeleitetes von bloß regelmäßigem Verhalten unterscheidet, bestehen darin, daß wir im allgemeinen Verhaltensabweichungen als irgendwie falsch oder fehlerhaft bezeichnen und daß - da Regeln, anders als auf vergangene Ereignisse bezogene Regelmäßigkeiten, automatisch auch für neue Fälle gelten - der Handelnde angesichts eines neuen Falles weiß, wie er sich zu verhalten hat." (Searle 1969/1983; S.67 f.)

Schaut man sich den Spracherwerb von Kindern an, so erkennt man auch sofort den Grund dafür, dass Sprachregeln zwar erworben und internalisiert sind, aber nicht bewußt bzw. expliziert vorliegen: Man erlernt seine Muttersprache nicht, indem man von den Bezugspersonen Sprachregeln expliziert bekommt. Die Mutter wird kaum zum Fünfjährigen sagen: "Wie oft soll ich dir eigentlich noch sagen, dass *Wohin* den Akkusativ verlangt?", sondern sie wird andeuten, dass es nicht "Ich gehe in de**m** Garten", sondern "Ich gehe in de**n** Garten" heißt. Man mag das ineffektiv finden, zumindest in der Fremdsprachendidaktik setzt man ja mehr auf das Erlernen einer Sprache über die Explikation von Sprachregeln. Deshalb fällt es manchen Menschen auch leichter, Sprachregeln für Sprachen zu explizieren, die sie kaum sprechen, als für Sprachen, die sie muttersprachlich sprechen. Ein ähnliches Phänomen findet sich bei anderen sozialen Regeln, erst ethnologische Erkenntnisse haben unseren Blick für bestimmte Regeln unseres sozialen Handelns geschärft bzw. ermöglicht. Gleichzeitig wird mit diesem genetischen Blick darauf, dass Regelwissen oftmals unexpliziert weitergegeben wird, deutlich, dass wir mit objektiv-hermeneutischem Vorgehen nicht nur latente Sinnstrukturen des vom Fall produzierten Textes freilegen, sondern auch latente Sinnstrukturen des Regelgebäudes selbst.

Die objektiv-hermeneutische Berufung auf die Geltung von Regeln wird gelegentlich mit der Unterstellung gekontert, dass Regeln dafür als unveränderlich angenommen werden müßten. Popper nennt die damit unterstellte Position "naiven Naturalismus" - sowohl natürliche als auch konventionelle Regelmäßigkeiten werden dabei als völlig unabänderlich angenommen. Natürlich sind Regeln veränderlich, solche Veränderungen kennt jeder sowohl im Sprachbereich als auch im sonstigen sozialen Leben. Einerseits muß das in der Interpretationspraxis stets berücksichtigt werden, andererseits hilft der Zwang zum Fallibilisierungsversuch, unter anderem solche Veränderungen zu entdecken. Dasselbe gilt für Texte von Sprechern mit mangelnder Regelkompetenz, z.B. sehr kleine Kinder oder Ausländer, sowie für Texte aus Subkulturen: Ein sauber durchgeführtes Falsifikationsverfahren wird hier nicht nur eine Fallstrukturhypothese hervorbringen, sondern mit dem ersten Parameter der Textproduktion auch diese Spezifika von Regelkenntnis herausarbeiten. Ich möchte den Charakter von Regeln noch etwas darstellen. Er zeigt sich am deutlichsten im Vergleich zum Charakter von Gesetzen.[28]

[28] Die Unterscheidung von regelgeleitetem und gesetzmäßigem Handeln gibt auch einen weiteren Hinweis auf einen grundsätzlichen Konstruktionsfehler in computergebundenen kognitiven Theorien wie der von Kurt Reusser. Solche Ansätze zwingen nämlich zu einer quasigesetzlichen Charakterisierung von Sprachregeln: Das Konstitutivum der Regelabweichung kann nicht programmiert werden, weil die *Bedeutung*, die eine Regelabweichung vermittelt, nicht allgemein ohne den jeweiligen sprachlichen Kontext beschrieben werden kann. Da prinzipiell unendlich viele Regelabweichungen möglich sind, kann ein Computer zwar eine Regelabweichung feststellen, nicht aber ihre Bedeutung. Man kann das verschleiern, indem man endlich viele Regelabweichungen in die Allheilmittelkiste „Weltwissen" packt, dadurch wird aber das grundsätzliche Problem nicht gelöst.

Claude Levi-Strauss spitzt den bereits von Weber ausgearbeiteten Gedanken der Unterscheidung von Regeln und Gesetzen zu, um Natur und Kultur voneinander zu unterscheiden:

> „Das Fehlen von Regeln scheint uns das zuverlässigste Kriterium zu sein, um einen natürlichen Prozeß von einem kulturellen Prozeß zu unterscheiden. ... Überall dort, wo eine Regel auftaucht, wissen wir mit Bestimmheit, daß wir uns auf der Ebene der Kultur befinden. Symmetrisch dazu bereitet es keine Schwierigkeit, in der Universalität das Kriterium der Natur zu erkennen." (Levi-Strauss 1993, S.46)

Ich halte diese Zuspitzung unter anderem deshalb für produktiv, weil sie uns bei näherer Betrachtung deutlich macht, dass die Trennung von Natur und Kultur an den Grenzflächen nicht mehr funktioniert. Für meine sich eindeutig auf dem Gebiet der Kultur bewegende Arbeit genügt es aber, die Unterscheidung von Regeln (und regelgeleitetem Handeln) und (Natur)Gesetzen (und gesetzmäßigem Geschehen) herauszuarbeiten. Ich beziehe mich dabei auf Günther Öhlschläger:

> „Wenn wir eine Beschreibung oder Teilbeschreibung etwa der Sprachkompetenz eines Sprachteilhabers als Beschreibung der sprachlichen Regeln, über die dieser Sprachteilhaber verfügt, gemacht haben, und sein sprachliches Handeln nicht mit dieser Beschreibung übereinstimmt, kann diese Diskrepanz darauf zurückgeführt werden, daß die Beschreibung falsch oder schlecht war, daß wir die Regeln nicht richtig beschrieben haben, daß wir falsche Regeln unterstellt haben, Regeln, über die der Sprachteilhaber gar nicht verfügt, aber auch darauf, daß der Handelnde einen Fehler gemacht hat. [...] Wenn aber die Bewegungen z.B. eines Planeten im Widerspruch zu den Gesetzen der Planetenbewegungen stehen, wird man auf keinen Fall annehmen, daß der Planet einen Fehler gemacht hat, abgewichen ist oder daß sich die Gesetzmäßigkeiten geändert haben, sondern vielmehr, daß diese Gesetze falsch sind und revidiert werden müssen." (Öhlschläger 1974; S.94)

Öhlschläger verweist zur Erklärung auf den unterschiedlichen Status von Naturobjekten und gesellschaftlichen Objekten in unserem Erkennen:

> „Das natürliche Geschehen wird für uns erst dadurch verständlich, daß wir die einzelnen Beobachtungen in einen Zusammenhang bringen, daß wir dem natürlichen Geschehen durch die Aufstellung von Gesetzen bestimmte Gesetzmäßigkeiten unterstellen, daß wir mit Hilfe von Gesetzen die beobachtbaren Regelmäßigkeiten erklären, wobei diese Regelmäßigkeiten erst Regelmäßigkeiten aufgrund einer Theorie sind, denn je nach der jeweiligen Theorie wird man andere Regelmäßigkeiten erkennen bzw. anderes Geschehen als regelmäßig ansehen: alle Beobachtungen sind schon theoriegeleitet.

> Bei gesetzmäßigem Geschehen kann es folglich qua Status von Gesetzen keine Abweichungen geben, da Gesetze erst und nur aufgrund der beobachtbaren Geschehen, die sich unabhängig von den Gesetzen vollziehen, gemacht werden, um diese zu beschreiben, zu erklären und für uns verstehbar zu machen. Und da die Geschehen die einzige Grundlage für die Aufstellung von Gesetzen sind, müssen alle Geschehen gleichberechtigt sein, es dürfen nicht einige schon als abweichend klassifiziert werden ... Wenn man eine Diskrepanz zwischen einem Geschehen und dem ihm unterstellten Gesetz als Abweichung und nicht als Widerspruch gegen das Gesetz auffassen würde, hätte man den einzigen Beurteilungsmaßstab für Gesetze und naturwissenschaftliche Theorien, der eben in der Übereinstimmung zwischen Geschehen und Gesetz besteht, aufgegeben und seine Theorie nicht mehr falsifizierbar und damit empirisch leer gemacht." (Öhlschläger 1974, S.96 f.)

Damit wird deutlich, dass gesetzliches Denken erst entstehen konnte, als man sich vom soziomorphen Denken verabschiedet hatte, welches natürliches Geschehen dem Willen sozialer Wesen zuschrieb. Wir unterstellen implizit, dass die Natur keiner Sozialität unterliegt, sie also weder einen Willen noch die Möglichkeit zu Individuiertheit und Fehlerhaftigkeit hat, dass sie ihre Gesetze nicht einfach verändern oder ihnen entgegenhandeln kann, dass es also keine *Möglichkeit* der Abweichung von Naturgesetzen gibt. Wir tragen also eine Struktur erst an die Natur heran, wohingegen

> „das soziale Handeln schon an sich sinnhaft strukturiert ist und nicht erst aufgrund des Herantragens einer Struktur, aufgrund des Aufstellens theoretischer Konstrukte verstehbar ist. Denn soziales Handeln ist durch diese Struktur, durch das Bestehen von Regeln, von Handlungsmustern, überhaupt erst möglich. Wenn sprachliche Zeichen beispielsweise beliebig und nicht regelhaft verwendet würden, könnte man mit jeder Äußerung Beliebiges meinen bzw. man könnte bei jeder Äußerung eines andern Beliebiges verstehen, und damit eben nichts meinen und nichts verstehen. Für jede Verständigung, für jede Interaktion ist es notwendig, daß die Handelnden über bestimmte Regeln verfügen, die intersubjektiv gültig sind und die es ihnen ermöglichen, sich verständlich zu machen und andere zu verstehen, d.h. sinnvolles, verstehbares Handeln muß immer regelgeleitetes Handeln sein. Die Intersubjektivität von Regeln wiederum ist begründet und gesichert durch die gemeinsame Praxis, man kann sich nur dann nach einer Regel richten, wenn es "einen ständigen Gebrauch, eine Gepflogenheit gibt" (Wittgenstein,§198)." (Öhlschläger 1974, S.97)

Auch Öhlschläger verweist darauf, dass die Möglichkeit des Abweichens von einer Regel diese nicht zerstört, sondern konstitutiv ist für Regeln, "denn es hat nur dann Sinn, von einer Regel zu sprechen, wenn es sowohl Handlungen gibt, die nach der Regel sind, als auch Handlungen und sonstige Aktivitäten, die nicht nach dieser Regel sind und als Abweichungen von dieser Regel gelten. Denn wenn es diesen Unterschied nicht gäbe, wären alle Handlungen beliebig und damit nicht verstehbar. Was als Fehler, als Abweichung von einer Regel gilt, ist also ebenso wie die Regel selbst und zusammen mit dieser intersubjektiv begründet." (ebenda)

Objektive Hermeneutik möchte die Bedeutung sozialer Handlungen herausarbeiten. Sie beruft sich dabei auf die Nichtbeliebigkeit, also die Bedeutsamkeit und damit Deutbarkeit sozialen Handelns. Sowohl regelkonformes als auch regelverletzendes Verhalten sind bedeutsam und damit deutbar. Da soziales Handeln entlang von Regeln stattfindet, sogar durch diese Regeln erst konstituiert wird, muß sich Objektive Hermeneutik im Interpretationsprozeß entlang dieser Regeln bewegen.

> „Über diese Regeln verfügen wir als sprach- und handlungsfähige Subjekte, ohne dass wir sie unbedingt explizieren können. Wir kennen z.B. die sprachlichen Regeln, können sie in Anwendung bringen und sind zu sprachlichen Angemessenheitsurteilen in der Lage, ohne deshalb Sprachtheoretiker sein zu müssen. Die Geltungssicherung der Interpretation hängt von der Geltung dieser Regeln ab und nicht von der Geltung der systematischen Explikation dieser Regeln." (Wernet 2000, S.13 f.)

2.1.5. Zur Objektivität der Objektiven Hermeneutik

Objektive Hermeneutik

> „ist also ein Verfahren, das sich auf die "verstehbare" Erfahrungswelt der Sozial-, Geistes- und Kulturwissenschaften nicht dadurch richtet, daß es primär deren subjektiven Niederschlag oder subjektive Repräsentanz im Bewußtsein der Handelnden nachvollzieht oder zu erschließen versucht [...]. Das wäre grundsätzlich mit Unsicherheiten behaftet und ein Verfahren, das selbst noch der zu untersuchenden Praxis angehört. Vielmehr macht die objektive Hermeneutik ernst mit den Konsequenzen aus den grundlegenden Erkenntnissen, daß
>
> 1. jede subjektive Disposition ... methodisch überprüfbar nie direkt greifbar ist, sondern immer nur vermittels einer Ausdrucksgestalt oder einer Spur, in der sie sich verkörpert oder die sie hinterlassen hat. Zutreffend entschlüsseln läßt sich daher eine solche Disposition erst, wenn man vorher die objektive Bedeutung jener Ausdrucksgestalt entziffert hat. Erst dann kann man zur begründeten Erschließung der Struktur der subjektiven Disposition selbst übergehen. Die übrigen Methoden der Forschung leiden darunter, daß sie entweder diesen Schritt der Vermittlung über die objektiven Bedeutungs- und Sinnstrukturen einer Ausdrucksgestalt auslassen oder von vornherein die Ebenen von Sinn- und Bedeutungshaftigkeit menschlichen Handelns ganz ausblenden und sich reduktionistisch auf die schiere Beobachtbarkeit äußeren Verhaltens beschränken.
>
> 2. Eine angemessene Methodologie ... muß mit der alten Anschauung brechen, derzufolge die Gegenstände der Erfahrungswissenschaften an die sinnliche Wahrnehmbarkeit gebunden sind und insofern konkret sind. Sinn- und Bedeutungsstrukturen sind grundsätzlich abstrakt. Sie lassen sich als solche sinnlich nicht wahrnehmen, aber sie sind dennoch empirisch und als empirische erfahrungswissenschaftlich analysierbar. Deshalb folgt die objektive Hermeneutik einem methodologischen Realismus, indem sie als empirisch alles das ansieht, was sich durch Methoden der Geltungsüberprüfung nachweisen läßt." (Oevermann 1996, S.1 f.)

Damit sind weitere Grenzen abgesteckt, innerhalb derer Objektive Hermeneutik arbeitet: Dies sind die Grenzen des Verstehbaren und des durch Methoden der Geltungsüberprüfung Nachweisbaren - allerdings muß sich ihrer jede Wissenschaft bewußt sein. Und es ist die Grenze der subjektiven Disposition, die methodisch kontrolliert nie direkt greifbar ist. Objektive Hermeneutik liest nur deren Spuren, behauptet aber, dass eben diesen Spuren eine objektive Bedeutung zugemessen werden kann, weil die Erzeugung einer Spur bzw. einer Ausdrucksgestalt ein sozialer Prozeß ist und somit regelhaft verläuft, genau diese Regelhaftigkeit aber objektive Bedeutungen erzeugt. Der Übergang von der Erkenntnis objektiver Bedeutungen von Ausdrucksgestalten zur Erkenntnis von subjektiven Dispositionen verläuft über Strukturbetrachtungen: Man stellt sich die subjektiven Dispositionen als eine Struktur vor und die objektiven Bedeutungsstrukturen ebenfalls als eine Struktur - und zwar als eine, welche durch die erstgenannte Struktur erzeugt wurde. Die Struktur der subjektiven Dispositionen ist grundsätzlich in Transformation begriffen, die objektive Bedeutungsstruktur ist eine statische Momentaufnahme davon.

Oevermann nimmt in der klaren und bewußten Unterscheidung der subjektiven Dispositionen und der objektiven Bedeutungen den Popperschen Gedanken der Unterscheidung von Welt 2 und Welt 3 auf: Popper (1994) argumentiert dazu in seinen Vorträgen „Erkenntnistheorie ohne erkennendes Subjekt" (erschienen 1967) und "Zur Theorie des objektiven Geistes" (erschienen 1968): „Ohne die Wörter „Welt" oder „Universum" allzu ernst zu nehmen, kann man folgende drei Welten oder Universen unterscheiden: erstens die Welt der physikalischen Gegenstände oder physikalischen Zustände; zweitens die Welt der Bewußtseinszustände oder geistigen Zustände oder vielleicht der Verhaltensdispositionen zum Handeln; und drittens die Welt der *objektiven Gedankeninhalte*, insbesondere der wissenschaftlichen und dichterischen Gedanken und der Kunstwerke." (Popper 1994, S.109) Popper begründet die unabhängige, autonome Existenz der „Welt 3" (S.111, 121f., 165-167) sowie ihre Objektivität (118-120, 163f.) und argumentiert (115-118), dass die Erkenntnis über erzeugte Strukturen grundlegender ist als die Erkenntnis über die Erzeugung dieser Strukturen. Seine Beispiele sind Spinnweben und die Erzeugung von Spinnweben, Theorien und die Erzeugung von Theorien; für diese Arbeit: Aufgabentexte und die Erzeugung von Aufgabentexten. „Im Gegensatz zu einem ersten Eindruck können wir mehr über das Herstellerverhalten lernen, wenn wir die Erzeugnisse untersuchen, als wir über die Erzeugnisse lernen, indem wir das Herstellerverhalten untersuchen. [...] Im folgenden werde ich den Ansatz von den Erzeugnissen her - den Theorien und Argumenten - den „objektiven" oder „Welt-3-Ansatz" nennen, den behavioristischen, den psychologischen und soziologischen Ansatz ... den „subjektiven" oder „Welt-2-Ansatz"." (ebenda, 117 f.) Objektive Hermeneutik folgt diesem Wissenschaftsverständnis, sie untersucht vorhandene - textliche, „objektive" - Strukturen. Ihr „objektiv"-sein bezieht sich auf die in Poppers Argumentation zentrale *Objektivität der Welt 3* (siehe 118-122). Sie untersucht Sprachstrukturen, also Objekte der Welt 3. Sie trifft Bedeutungsaussagen über diese Objekte, **die „Objektivität" dieser Aussagen bezieht sich aber nur** auf das regelerzeugte Gebilde „Sprache", welches **ein Welt-3-Objekt** ist. Schlußfolgerungen auf subjektive Dispositionen - ob auf strukturerzeugende Dispositionen (beim Sprecher) oder auf durch die Struktur erzeugte Wirkungen (beim Rezipienten) - sind davon klar zu trennen. Diese subjektiven Dispositionen gehören der Welt 2 an; vor ihrer Betrachtung müssen also zunächst (im nächsten Abschnitt) die Zusammenhänge zwischen Welt 3 und Welt 2 betrachtet werden.

Mit dieser Sichtweise wird auch klarer, warum geisteswissenschaftliches empirisches Arbeiten methodisch kontrolliert nur auf der Ebene der *Ausdrucksgestalten* erfolgen kann:

> „Bei näherer Betrachtung nimmt man nämlich die objektiven bzw. latenten Bedeutungs- und Sinnstrukturen nicht den Handlungen oder Äußerungen von Handlungssubjekten selbst ab, sondern den Spuren oder Protokollen, die sie hinterlassen haben. Diese Spuren oder Protokolle konstituieren für die Methodologie der objektiven Hermeneutik die methodisch einzig greifbare Datenebene ... Dahinter steht, dem methodisch direkten Zugriff schon entzogen, die Welt der beobachtbaren und das heißt auch: protokollierbaren Handlungen und Äußerungen, die ihrerseits wiederum auf die dahinter stehende, und dem methodisch direkten Zugriff erst recht entzogene Welt der „Subjektivität", der subjektiven Dispositionen also, ... verweisen.
> [...] Wo immer wir es mit einem geistes-, kultur- oder sozialwissenschaftlich relevanten Datum oder Ereignis zu tun haben, ist unter methodologischem Gesichtspunkt zwingend diese [eigenständige, auf anderes nicht reduzierbare, objektiv gegebene, d.h. methodisch zwingend nachweisbare] Realität der latenten Sinnstrukturen diejenige, die einzig explizit und lückenlos sowie jederzeit intersubjektiv überprüfbar und wiederholbar erschlossen werden kann und auf deren Erschließung alle anderen Schlußfolgerungen über andere Realitätsebenen zwingend sich zu beziehen haben. Die objektive Hermeneutik - oder andere denkbare und mit ihr vergleichbare Formen einer strukturalen Hermeneutik - unterscheidet sich von anderen Methodologien und Methodenansätzen ... genau dadurch, daß sie tatsächlich diese Realität der latenten Sinnstrukturen in das Zentrum der methodischen Operation stellt und explizit zum Gegenstand der Rekonstruktion erhebt, also die Operationen der Erschließung dieser Realität sowie der Ableitung von Interpretationen anderer, z.B. subjektiver oder innerpsychischer Realitäten aus dieser Erschließung nicht implizit beläßt bzw. mit Stillschweigen übergeht." (Oevermann 1993, S.113 f.)

Testaufgaben sind Protokoll einer Testerstellungspraxis. In dieser Arbeit interessiert aber nicht vorrangig diese Testerstellungspraxis, sondern der umgekehrte Prozeß der Rezeption der Testaufgaben, da die hinter einem Test stehende Behauptung ist, dass mit ihm und also mit seinen Aufgaben Aussagen über psychische Dispositionen gemacht würden. Für diese Betrachtung muß zunächst klar zwischen der Ebene der Aufgabe (Welt 3) und der Ebene ihrer subjektiven Repräsentation (Welt 2) unterschieden werden. Im Rückblick auf das vorige Kapitel zeigt sich, dass Baumert u.a. (2000) schon diese beiden Ebenen miteinander vermengen. Reusser (1989) und Kintsch (1994) hingegen trennen ebenso wie Oevermann beide Ebenen in ihrer Betrachtung deutlich und postulieren eine - weiterhin näher zu beschreibende - Wirkung der Welt 3 auf die Welt 2.

Für Multiple-Choice-Testaufgaben wird bereits an dieser Stelle deutlich, dass die Realität der latenten Sinnstrukturen bzw. der objektiven Textbedeutungen des Aufgabentextes diejenige ist, die „einzig explizit und lückenlos sowie jederzeit intersubjektiv überprüfbar und wiederholbar erschlossen werden kann". Frei beantwortbare Aufgaben liefern ein zusätzliches Protokoll einer Schülerhandlung, welches rekonstruktiv interpretierbar ist und somit *vor dem Hintergrund einer Aufgabeninterpretation* Schlüsse auf psychische Dispositionen zulässt. Eine bloße Subsumtion unter Lösungskategorien birgt hier aber - und das überrascht auf den ersten Blick - größere Gefahren der Verkürzung in sich als Multiple-Choice-Angebote, die objektiv-hermeneutisch interpretiert sind.

2.1.6. Subjektivität

Der Übergang von der Erkenntnis objektiver Bedeutungen von Ausdrucksgestalten zur Erkenntnis von subjektiven Dispositionen verläuft über Strukturbetrachtungen: Man stellt sich die subjektiven Dispositionen als eine Struktur (aus Welt 2) vor und die objektiven Bedeutungsstrukturen ebenfalls als eine Struktur (aus Welt 3) - und zwar als eine, welche durch die erstgenannte Struktur erzeugt wurde. Die Struktur der subjektiven Dispositionen ist grundsätzlich in Transformation begriffen, die objektive Bedeutungsstruktur ist eine statische Momentaufnahme davon.

Wie erfolgt nun der Rückschluß von der objektiven Struktur auf die subjektive? Das genaue Zusammengehen der Dynamik der Struktur und der statischen Momentaufnahme ist nur teilweise klar zu benennen: Objektive Hermeneutik hat den Vorteil, die Dynamik einer Struktur widerspiegeln zu können, da diese Dynamik sich an Entscheidungspunkten entäußert und die Entscheidungsknoten einer Sequenz solche Punkte sind (siehe Abschnitt 2.2., Stichwort Sequentialität). Objektive Hermeneutik arbeitet genau an diesen Punkten, und wenn der Text an verschiedenen Stellen verschiedene Entscheidungsmuster zeigt, dann kann Objektive Hermeneutik die Muster vergleichen und Widersprüchlichkeiten oder eben Transformationen rekonstruieren. Wenn solche Transformationen sichtbar sind, erscheint es überzeugend, dass das texterzeugende Individuum über psychische Dispositionen zur Erzeugung dieser Transformationsstruktur verfügen muß. Umgekehrt erscheint es allerdings problematischer: Wenn ein Text eine statische Struktur zeigt, dann liegt der Einwand nahe, dass eine in Transformation befindliche psychische Struktur hier lediglich einen Teil ihrer Gesamtheit textlich niederlegt. Diese Debatte um Statik und Dynamik ist aber lediglich Spielart eines immer akuten Erkenntnisproblems, nämlich ob wir den richtigen Moment für eine Aussage über ein Subjekt erwischen. Deshalb gibt es bei anderen Methoden Wiederholungsuntersuchungen oder große Fallzahlen, in der Objektiven Hermeneutik ist der Falsifikationsversuch unter anderem hierauf eine Antwort. Man kann nun aber nicht behaupten, dass Objektive Hermeneutik eine Garantie dafür geben kann, eine psychische Struktur in ihrer ganzen Dynamik zu erfassen. Mit dem sequenzanalytischen Vorgehen und der Betrachtung der Auswahlspezifika an den Entscheidungsknoten ist aber die größtmögliche Chance gegeben, dass eine Struktur in ihrer Dynamik erfaßt wird. Hier ist zwar eine Grenze beschrieben, an der methodisch kontrolliert keine „garantierten Aussagen" zum Verhältnis von Welt 2 und Welt 3 möglich sind, aber von den vorfindlichen Methoden ist die Objektive Hermeneutik diejenige, die die Dynamik psychischer Strukturen am ehesten widerspiegelt.

Das Problem der Erfassung der Dynamik psychischer Strukturen ist Bestandteil des Problems, dass subjektive Dispositionen niemals direkt greifbar sind, sondern lediglich vermittels Ausdrucksgestalten. Auch Objektive Hermeneutik vermag die Erkenntnislücke zwischen der objektiven Bedeutungsstruktur eines Textes und den subjektiven Dispositionen des Texterzeugers, also zwischen Welt 3 und Welt 2, nicht methodisch kontrolliert, nicht systematisch und nicht vollständig zu schließen, obwohl sie methodisch kontrolliert, systematisch und vollständig objektive Bedeutungsstrukturen zu rekonstruieren vermag. Sie stellt sich dieser Lücke aber wenigstens und ist sich der Begrenztheit ihrer Aussagen bewußt. So wird umgekehrt deutlich, dass auch die Operationalisierung eines Konstrukts in einem Test - und sei jede einzelne Aufgabe objektiv-hermeneutisch optimiert - nur zu einem Welt-3-Produkt führt. Selbst bei einem sehr kleinen Lösungsspielraum ist kein direkter Zugriff auf psychische Dispositionen möglich. Vereinfacht gesagt: Selbst wenn ein Schüler die Aufgabe 3+5 falsch löst, ist daraus keine zwingende Aussage über seine mathematischen Fähigkeiten abzuleiten. Die forsche Art, in der bei TIMSS und PISA aus grob zusammengeschusterten Aufgabensammlungen weitreichende, durch die erhobenen Daten nicht gedeckte Schlußfolgerungen gezogen werden, paßt sehr gut dazu, dass dort prinzipielle Grenzen der mit der gewählten Methode erreichbaren Erkenntnis nicht debattiert werden. Kurt Reusser entzieht sich für seine Methode noch der Diskussion der Anwendbarkeit seines Welt-3-Modells für psychische Erkenntnis, indem er die Problemlösefähigkeit seines Computerprogramms als selbständiges Ergebnis nimmt und das Axiom, dass das Denken dann ebenso wie dieses Programm funktionieren würde, in den Bereich des empirisch noch zu Bestätigenden verweist. Baumert u.a. ignorieren diese Grenze und behaupten einfach eine Paßfähigkeit von Modell und subjektiven Dispositionen bzw. Handlungen.

Das Problem der Subjektivität erschließt sich deutlicher, wenn man es im Kontext der Dialektik von Besonderem und Allgemeinem betrachtet:

> „Das vereinzelte Individuum, das reine Subjekt der Selbsterhaltung, verkörpert im absoluten Gegensatz zur Gesellschaft deren innerstes Prinzip. Woraus es sich zusammensetzt, was in ihm aufeinanderprallt, seine „Eigenschaften", sind allemal zugleich Momente der gesellschaftlichen Totalität. Monade ist es in dem strengen Sinn, daß es das Ganze mit seinen Widersprüchen vorstellt, ohne doch je dabei des Ganzen bewußt zu sein. Aber in der Gestalt seiner Widersprüche kommuniziert es nicht stets und durchgängig mit dem Ganzen, sie rührt nicht unmittelbar von dessen Erfahrung her. Die Gesellschaft hat ihm die Vereinzelung aufgeprägt, und diese hat als ein gesellschaftliches Verhältnis teil an seinem Schicksal. „Psychodynamik" ist die Reproduktion gesellschaftlicher Konflikte im Individuum, aber nicht derart, daß es die aktuellen gesellschaftlichen Spannungen bloß abbildete. Sondern es entwickelt auch, indem es als ein von der Gesellschaft Abgedichtetes, Abgespaltenes existiert, nochmals die Pathogenese einer gesellschaftlichen Totalität aus sich heraus, über der selber der Fluch der Vereinzelung waltet." (Adorno 1955/1973, S.21 f.)

Das empirische Verfahren der Objektiven Hermeneutik stellt sich dieser Dialektik von Besonderem und Allgemeinem. Im Konkreten - und das heißt immer im Besonderen, am Fall und als Welt-3-Verfahren - muß sie die Komplexität dieser „dialektisch totalisierten Monadität" aber immer abreichern: Man hat den Text und untersucht ihn sequenzweise. An jedem Sequenzknoten sind mit Hilfe von sinnlogischen Erzeugungsregeln die Möglichkeiten zu explizieren, die sich in die (textliche) Zukunft hinein eröffnen. Diese Regeln (Parameter I, siehe Abschnitt 2.1.2. „Strukturalismus und Sequentialität" und Oevermann 1996, S.7, 14) erschließen das Allgemeine. Das Besondere zeigt sich in den Auswahlmaximen (Parameter II), nach denen der Fall aus den Möglichkeiten wählt. Mit dem Parameter I erschließt sich das Allgemeine - bezogen auf den jeweiligen Entscheidungsknoten - in seiner Totalität, aber eben auch nur in der Explizierung des Be-

sonderen - nämlich der verschiedenen wohlgeformten Möglichkeiten der Entscheidung. Das Besondere in der je konkreten Entscheidung bei Parameter II wiederum zeigt sich als Besonderung des Allgemeinen.
Die Abreicherung gegenüber der allgemeinen Sichtweise auf das Subjekt als in der Dialektik von Totalität und Monadität stehend besteht nun

i) im notwendig algorithmischen Vorgehen der methodischen Praxis, welches das Allgemeine immer vor das Besondere stellt und das Besondere immer aus dem Allgemeinen zieht. Umgekehrt nimmt dieses Vorgehen ja gerade den Charakter des Besonderen als „das Besondere in dem Allgemeinen" und das Sichergeben des Allgemeinen erst aus dem Besonderen auf.

ii) Zu einem einzelnen Zeitpunkt ist der Nachvollzug der Genese sowohl der Regelentstehung als auch der Regelaneignung als auch der Fallstruktur nicht möglich. Dazu müssen mehrere Zeitpunkte empirisch untersucht werden. Umgekehrt ist der Strukturbegriff der Objektiven Hermeneutik größtmöglich dynamisch: Die Struktur wird ja nicht durch die Beschreibung statischer Elemente eruiert, sondern durch die Beschreibung der Auswahlmaximen, nach denen der Fall an den Sequenzknoten aus den Möglichkeiten auswählt. Die Tatsache, dass der Fall an jedem Sequenzknoten neu auswählen kann, wird damit berücksichtigt. Damit ist aber methodisch berücksichtigt, dass der Fall an jedem Sequenzknoten eine Strukturtransformation statt einer Strukturreproduktion vollziehen kann - deshalb bezeichnet Oevermann Transformation auch als Normalfall und Reproduktion als Grenzfall, dies generiert größtmögliche Dynamik des Strukturbegriffs.

iii) Die Fallstruktur vermag niemals das Subjekt als Ganzes zu erfassen: Objektive Hermeneutik beschränkt sich auf die Untersuchung von Ausdrucksgestalten. Nur sie geben nachvollziehbar Auskunft über das Subjekt. Die Gesamtheit (des Psychischen) des Subjekts erschöpft sich aber weder in den von ihm produzierten Ausdrucksgestalten noch in den sich daraus erschließenden Fallstrukturen. Objektive Hermeneutik stellt sich aber dieser Grenze des Erkennbaren, indem sie eben gar nicht erst behauptet, sie könne erschöpfend Auskunft über Subjekte geben, sondern sich auf die Herausarbeitung von Fallstrukturen beschränkt. Wie jede Methode unterliegt sie damit der Gefahr des Mißverstandenwerdens: Sie kann methodisch kontrolliert Auskunft nur über Ausdrucksgestalten geben. Der Rückschluß auf das Subjekt muß sowohl die Dialektik von Allgemeinem und Besonderem mitdenken als auch die Tatsache, dass Ausdrucksgestalten vom Subjekt nicht beliebig produziert werden, aber auch nicht in mechanisch beschreibbarer Weise, dass also keine mechanische Beschreibung des Zusammenhangs von psychischen Strukturen und Fallstrukturen möglich ist.

2.2. Zur Praxis objektiv-hermeneutischen Interpretierens

Bereits im Kapitel 1 wurde für die Aufgabe D2 eine objektiv-hermeneutische Textinterpretation durchgeführt. Einige methodologische Grundüberlegungen der Objektiven Hermeneutik wurden im vorausgehenden Abschnitt diskutiert. Die weitere Darlegung methodologischer Grundlagen soll mit einer kurzen interpretationspraktischen Einführung verbunden werden.

Die Kernprozedur einer Textinterpretation ist die Explikation der Textbedeutung. In seiner „Einführung in die Interpretationstechnik der Objektiven Hermeneutik" beschreibt Wernet das Vorgehen bei der Bedeutungsexplikation:

> „(1) *Geschichten erzählen*, (2) *Lesarten bilden*, (3) ... *diese Lesarten mit dem tatsächlichen Kontext konfrontieren*. Dieser Dreischritt ist von grundlegender Bedeutung für die Interpretationstechnik der Objektiven Hermeneutik. Ich werde ihn kurz erläutern ...
>
> 1. Zu einem vorliegenden Text bzw. Textausschnitt, dessen Bedeutung wir klären wollen, erzählen wir zunächst Geschichten. Das sind Geschichten, in denen der Text vorkommen könnte. Diese Geschichten unterliegen grundsätzlich nur zwei Einschränkungen: (i) Sie sollen nicht im Rahmen des tatsächlichen Äußerungskontextes liegen, sondern diesen Kontext verlassen. (ii) Es sind nur solche Geschichten erlaubt, in denen der Text uns als angemessene sprachliche Äußerung [also als „wohlgeformt" W.M.] erscheint. Im Akt des Geschichtenerzählens mobilisieren wir unser intuitives Regelwissen, lassen wir es arbeiten. Und in dem Streit um die Zulässigkeit von Geschichten problematisieren wir diese Operation.
>
> 2. Auf der Grundlage dieser erzählten Geschichten erfolgt die Lesartenbildung. Wir befragen hierbei die Geschichten auf Strukturgemeinsamkeiten hin. Die u.U. vielen Geschichten werden zu Typen gruppiert, indem Gemeinsamkeit und Differenz der Geschichten expliziert wird. Manchmal lässt sich keine Differenz feststellen und wir formulieren nur einen Typus. Häufig finden wir zwei oder drei Bedeutungstypen. Selten sind es mehr. Aus diesen Typen ergibt sich die *fallunspezifische* Textbedeutung.
>
> 3. Schließlich konfrontieren wir diese Rekonstruktion der fallunspezifischen und kontextfreien Textbedeutung, die Lesarten also, mit dem *tatsächlichen Äußerungskontext* und der darin eingelassenen *Aussageintention* des Textes. Diese Operation erschließt die Besonderheit der Fallstruktur. Auf diesem Wege gelangen wir zu *Fallstrukturhypothesen*." (Wernet 2000, S.39 f.)

Objektiv-hermeneutisches Interpretieren folgt innerhalb dieses dreischrittigen Vorgehens den **5 Prinzipien Kontextfreiheit, Wörtlichkeit, Sequentialität, Extensivität und Sparsamkeit**, die im weiteren dargelegt werden. Die Verpflichtung auf diese Prinzipien ist zwingend, gleichwohl wird in dieser Arbeit nur an ausgewählten, besonders prägnanten Stellen expliziert, dass ein solches Prinzip angewandt wurde. Dies erfolgt an Stellen, an denen einer naheliegenden Alltagsdeutung zu widersprechen ist, an denen ohne diese Prinzipien Interpretationsfehler nahe liegen oder an denen die Reichhaltigkeit der Interpretation aus der Anwendung der Prinzipien erwächst. Genauso wird der Interpretationsdreischritt nur an Stellen expliziert, an denen der Text sich einer Deutung so stark sperrte, dass eine Verschriftlichung unabdingbar wurde.

Nun ist Objektive Hermeneutik keine Ansammlung von Interpretationsregeln. Man könnte sie eher als Denkweise, als Art des Umgangs mit Texten bezeichnen, die sich aus den methodologischen Grundlagen ergeben, von denen nur ein Teil sich in Interpretationsregeln niederschlagen kann. Im Zentrum steht dabei der Anspruch, den Text als Ausdrucksgestalt einer sozialen Praxis ernst zu nehmen, die mit einer eigenen Objektivität behaftet und (sprach)regelgebunden ist. Die eher technische Aufstellung von Interpretationsregeln dient also nicht der Technisierung versus intellektueller Durchdringung des Textes, sondern der Disziplinierung des interpretatorischen Vorgehens zur Erlangung einer methodisch kontrollierten und objektiven Durchdringung.

2.2.1. Kontextfreiheit („Beachte jeden denkbaren Kontext der Äußerung!“)

> „Das Prinzip der kontextfreien Interpretation scheint auf den ersten Blick absurd. Wie kann ausgerechnet eine Methode des Verstehens auf den Kontext verzichten? Wenn wir eine Handlung verstehen wollen, sind wir dann nicht in besonderem Maße darauf angewiesen, den Kon-Text zu kennen? Offensichtlich kommt es doch darauf an, wer zu wem etwas sagt, in welcher Situation etwas gesagt bzw. getan wird, wann etwas gesagt wird, usw. Erst dann erhellt sich doch die eigentliche Bedeutung der Handlung. Demgegenüber scheint die Vorstellung einer kontextfreien Interpretation geradezu wirklichkeitsfremd.
> Ein Teil dieser Einwände ist einem naheliegenden Missverständnis geschuldet: Kontextfreiheit bedeutet *nicht*, dass die Umstände einer Handlung nicht wichtig seien zum Verständnis der Bedeutung dieser Handlung. Im Gegenteil ... spielt sogar die Operation der Einbeziehung des Kontextes eine interpretatorisch herausragende Rolle. Allerdings, so der Standpunkt der Objektiven Hermeneutik, erfolgt diese Operation erst nach der kontextfreien Interpretation. Die Kontextuierung ist der kontextfreien Bedeutungsexplikation *systematisch nachgeordnet*. Erst durch diese Nachordnung werden die beiden Dimensionen analytisch unabhängig. Die Rekonstruktion der Bedeutung eines Textes durch den Kontext läuft nämlich Gefahr, den Text ausschließlich durch den Kontext zu verstehen. Eine solche Betrachtung würde offensichtlich keine Textanalyse darstellen, sondern eine Kontextanalyse. Von vornherein würde der Text als eigenständiges Wirklichkeitsgebilde unterlaufen und missachtet werden. Denn es käme immer nur darauf an, ihn aus seinen Zusammenhängen heraus zu erläutern und ihn nicht als soziale Realität *sui generis* zu betrachten.“ (Wernet 2000, S.21 f.)

Ein Beispiel gibt die folgende Aufgabe:

> *Ein Marktplatz soll gepflastert werden. 3 Arbeiter benötigen dazu 6 Arbeitstage. Jeder Arbeiter verlegt gleich viel.*
> *Die Firma hat nun aber nur 2 Arbeiter zur Verfügung. Wie viele Arbeitstage brauchen diese beiden, um den Marktplatz zu pflastern?*

Der Satz „*Jeder Arbeiter verlegt gleich viel*.“ ist „seltsam“, man fragt sich, was er innerhalb der Aufgabenstellung zu bedeuten hat. Es liegt nahe, ihn kontextuell zu erklären, nämlich mit der innerdidaktischen Diskussion um eingekleidete Aufgaben. Offensichtlich soll dieser Satz die Anwendbarkeit umgekehrt proportionalen Arbeitens signalisieren. Diese kontextuelle Erklärung verstellt aber den Blick auf die innertextliche Realität, die diese Aufgabe und in ihr dieser Satz schafft. In der Aufgabe wird permanent eine Unklarheit über den Status des Problems als reale oder als gedankenexperimentelle Situation erzeugt. Genau diese Unklarheit produziert aber eine Aufgabenschwierigkeit, die nichts mit dem mathematischen Gegenstand oder mit einem Modellierungsprozeß zu tun hat. Wenn nun *nach* dieser Erkenntnis eine Konfrontation mit dem Kontext erfolgt, so bemerkt man sofort, dass die didaktische Position, die diese Aufgabe didaktisch „verschönt“ hat, ein Problem erst erzeugt - nämlich ein Unklarheitsproblem, welches der Schüler zusätzlich zum eigentlichen Problem zu überwinden hat, indem es nämlich nicht darum geht, wie man eine Realsituation mathematisch bearbeitet, sondern wie man solche Aufgaben zu verstehen hat. Diese Erkenntnis über den Charakter solcher Aufgaben wäre nicht möglich gewesen, wenn wir die Kontextbetrachtung und unser Wissen darüber, wie solche Aufgaben „gemeint“ sind, an den Anfang der Interpretation gestellt hätten. Erst die Konfrontation des didaktisch Gemeinten mit dem textlich Wirklichen ermöglicht hier die Erkenntnis.

> „Die Bedeutungsrekonstruktion ... basiert darauf, textkompatible *gedankenexperimentelle Kontexte* zu formulieren. In dem ersten Textzugriff werden ausschließlich diese gedankenexperimentellen Kontexte - statt des tatsächlichen Kontexts - zur Bedeutungsexplikation herangezogen. Die kontextfreie Interpretation bedient sich also einer Haltung zum Text, die in einem ersten und vorläufigen Interpretationsschritt sich dem Text in „künstlicher Naivität“ zuwendet.“ (ebd., S. 23)

Diese Operation ist vorläufig und klammert unser Vorwissen in einer spezifischen Form aus.

> „Denn ausschließlich unser Wissen um den konkret vorliegenden Fall darf bei der kontextfreien Interpretation *nicht* herangezogen werden. Die „künstliche Naivität“ bezieht sich also lediglich auf den zu interpretierenden Gegenstand einer Untersuchung. Die Kontextfreiheit der Interpretation versucht methodisch bewusst das Wissen um denjenigen Gegenstand, der im Fokus des Interesses steht, auszublenden. Forschungslogisch geht es dabei um die Vermeidung von Zirkularität. Wenn die Interpretation von dem Vorverständnis lebt und abhängig ist, dann ist sie in dessen Belieben gestellt. Diese Beliebigkeit gilt es zu vermeiden. Aus dieser Perspektive besteht der Sinn der kontextfreien Interpretation darin, gegenüber einem nicht-wissenschaftlich gewonnenen Vorverständnis größtmögliche Unabhängigkeit zu wahren.“ (ebd.)

2.2.2. Wörtlichkeit („Was da steht, steht da!“)

> „Wie schon das Prinzip der Kontextfreiheit, verpflichtet das Prinzip der Wörtlichkeit die Interpretation auf den Text. Will man den Text als Ausdruck von Wirklichkeit ernst nehmen und die Textanalyse als Wirklichkeitsanalyse begreifen, dann ist das Prinzip der wörtlichen Interpretation zwingend. Dem methodologischen Postulat der Textförmigkeit sozialer Wirklichkeit korrespondiert me-

> thodisch das Wörtlichkeitsprinzip. Und umgekehrt verunmöglicht eine Missachtung des Wörtlichkeitsprinzips die methodisch strikte Berufung auf Textverstehen als intersubjektiv überprüfbare Operation [weil die Abkehr von der Wörtlichkeit eine Auswahl bedeutet und diese Auswahl methodische Kontrolle zerstört W.M.].
>
> Das Prinzip der Wörtlichkeit besagt, dass die Bedeutungsrekonstruktion den tatsächlich artikulierten Text in seiner protokolliert vorliegenden Gestalt nicht ignorieren darf, auch und gerade dann nicht, wenn innertextliche Widersprüche auftreten.
>
> Das prominente Vorbild für die Einhaltung dieses Prinzips stellen die von Freud so bezeichneten Fehlleistungen dar. Wenn etwa „Tatsachen zum *Vorschwein* kommen“ oder die Aufforderung ergeht, „auf das Wohl unseres Chefs *aufzustoßen*“ (vgl. Freud 1916/17, 64 f.), dann stehen wir vor Texten, die zu erkennen geben, dass sie etwas sagen wollten, dies aber nicht gesagt haben. Im Alltag werden solche Fehlleistungen meist mit Lachen oder Schmunzeln bedacht. Daran sehen wir, dass beide Bedeutungsebenen erfasst werden. Wir sehen aber auch, dass in aller Regel die von der Intention des Sprechers abweichende Bedeutung nicht ernst genommen wird. Meist wird der Chef, auf dessen Wohl *aufgestoßen* wird, zusammen mit den anderen Gästen lachen und dem Versprecher keine weitere Bedeutung beimessen. Würde er beispielsweise seiner Frau zu Hause empört erzählen: *Stell dir vor, der Meier wollte auf mein Wohl aufstoßen*, müsste er mit Unverständnis rechnen: *Du meine Güte*, wird die Gattin vielleicht sagen, *leg doch nicht immer alles auf die Goldwaage*.
>
> Das Prinzip der Wörtlichkeit verpflichtet die Interpretation, den Text „auf die Goldwaage zu legen“ in einer Weise, die uns in alltäglichen Verstehenskontexten als inadäquat und kleinlich erscheinen würde.“ (Wernet 2000, S. 23 f.)

Die Interpretation von Versprechern ist nun einerseits vereinfacht, weil die mögliche Alternative - nämlich was gesagt werden *sollte* - bereits offen liegt. Sie ist andererseits erschwert, weil Alltagsinterpretationen textliche Abweichungen gedanklich heilen: Zum einen bemerkt man sie oft gar nicht, zum anderen glättet man sie im Sinne des „eigentlich Gemeinten“. So „markiert das Wörtlichkeitsprinzip die kategoriale Differenz von praktischer und wissenschaftlicher Einstellung“ besonders deutlich (ebd., S. 27). Untrennbar verbunden mit dem Nicht-Heilen ist aber der Zwang zu tiefgehender Analyse und zu falsifikatorischem Arbeiten, welche beide im Alltagshandeln kaum möglich sind - Wissenschaft arbeitet ja gerade mit dem Vorteil der Handlungsentlastung. Würde man andererseits „die wörtliche Bedeutungsschicht ignorieren, dann würde man den Text als wissenschaftliche Datenbasis missachten. Sowenig wie es einem Archäologen einfallen kann, Ausgrabungsstücke wegzuwerfen, sowenig im Rahmen quantitativer Forschung nur die Hälfte der Fragebögen berücksichtigt werden, sowenig kann ein textwissenschaftliches Verfahren Textelemente willkürlich beiseite lassen.“ (ebd., S. 24) In Testaufgaben wird man kaum Versprecher finden. Selten sieht man auch offenliegende Absurditäten wie in der PISA-Aufgabe „Bauernhöfe“, wo ein Quader als rechtwinkliges Prisma erläutert wird. Man würde dies nun in einer Alltagsinterpretation als eine Art Schrulligkeit oder fachorientierte Übertreibung beiseite kehren. Es führt aber - wie sich zeigt, wenn man den Text beim Worte nimmt - einen Umgang mit dem Aufgabeninhalt ein, der sich in seiner Grundstruktur noch mehrfach wiederholt und zu Verwerfungen führt: ein mathematisch trivialer Aufgabenanspruch wird in der Aufgabenstellung künstlich verkompliziert und fachsprachlich aufgebläht. Aber auch schon die Aufgabenüberschrift *BAUERNHÖFE* offenbart eine Grundstruktur der Aufgabe, sobald man sie wörtlich nimmt: Es geht nämlich in der Aufgabe nicht um Bauernhöfe, sondern um **einen** Bauerhof. Es geht nicht um einen Bauernhof, sondern um ein Bauern**haus**. Es geht nicht um ein Bauernhaus, sondern um das **Dach** eines Bauernhauses. Und wenn es sich überhaupt um ein Bauernhaus handelt, so hat die Bauernzuschreibung nichts mit dem Problem zu tun. Die Überschrift verspricht also größere Reichhaltigkeit, als die Aufgabe liefert. Der Schüler wird nicht nur - quasi dreieinhalbfach - in die Irre geführt, sondern auch an der Nase herumgeführt. Sowohl die Alltagswahrnehmung als auch die professionelle Wahrnehmung all derjenigen, die diese Aufgabe vor dem Test begutachtet haben, heilt diese Irreführung offenbar - sonst wäre sie verändert worden. Sobald man den Text wörtlich nimmt, offenbart er eine andere inhaltliche Dimension, in diesem Fall eine Grundstruktur, die sich noch mehrfach im Aufgabentext verifiziert. Hier bewährt sich die „methodologische Grundausrichtung der Objektiven Hermeneutik als Verfahren der Rekonstruktion *latenter Sinnstrukturen* vor allem in Abgrenzung zu einer inhaltsparaphrasierenden und auf Aussage- und Sprecherintentionen orientierten Text- und Sinninterpretation ... Wer den Text beim Wort nimmt, hat schon durch diese einfache Operation den entscheidenden Schritt getan, die Beschränkungen einer intentional-deskriptiven Interpretation zu überwinden.“ (ebd., S.26) Im Testerstellungsprozeß findet aber genau diese intentional-deskriptive Art der Aufgabenbewertung statt. Hier werden einerseits bereits vorhandene Aufgaben intentional-deskriptiv unter vorgegebene Kategorien subsumiert, statt in einem Prozeß von Konstruktion und Rekonstruktion eine Operationalisierung eines Fähigkeitskonstrukts vorzunehmen. Andererseits erfolgt auch die Interpretation der Ergebnisse nur aufgrund einer intentional-deskripitiven Einschätzung der Aufgabe, wie im Kapitel 1, unter anderem exemplarisch am Beispiel der Aufgabe D2, gezeigt wurde. Rudimentär und damit beliebig kommt Wörtlichkeit auch in der intentionalen Art von Aufgabenerstellung und -bewertung zum Tragen: Auch dort werden kontrastierende Betrachtungen vorgenommen. So, wie in der Interpretation von D2 der Vergleich von *abgebildet, gezeichnet, gezeigt, dargestellt, skizziert* oder der von *getaucht* und *eingetaucht* Erklärungskraft entfaltete, so nutzen auch Baumert u.a. dort, wo es in ihre Argumentation passt, Wörtlichkeitsbetrachtungen (z.B. Baumert u.a. 2000, S.199) Das Wörtlichkeitsprinzip wendet die hier praktizierte Beliebigkeit zu methodischer Kontrollierbarkeit hin.

2.2.3. Sequentialität („Beachte, dass der Textersteller die erste Selektionsentscheidung vor der zweiten Selektionsentscheidung getroffen hat!“)

> „Das Analyseverfahren, durch das sich die objektive Hermeneutik von allen anderen Methodenansätzen radikal unterscheidet, ist die Sequenzanalyse. Sie lehnt sich an die Sequentialität an, die für humanes Handeln konstitutiv ist. Dabei wird unter Sequentialität nicht ein triviales zeitliches oder räumliches Nacheinander bzw. Hintereinander verstanden, sondern die mit jeder Einzelhandlung als Sequenzstelle sich von neuem vollziehende <u>Schließung vorausgehend eröffneter Möglichkeiten</u> und <u>Öffnung neuer Optionen in eine offene Zukunft</u>. Insofern jedes Protokoll ausschnitthaft je konkrete Lebenspraxen authentisch ausdrückt, die sich darin verkörpern, bildet es auch den realen se-

quentiellen Prozeß ab, in dem sich diese konkreten Lebenspraxen, die an einer protokollierten Wirklichkeit handelnd beteiligt sind, ganz konkret in eine offene Zukunft entfalten.“ (Oevermann 1996, S. 6)

„Die Selektivität einer Lebenspraxis auf der Folie der durch soziale Regeln eröffneten Handlungsmöglichkeiten vollzieht sich [dabei] nicht statisch, sondern *prozessual*. Die durch Regeln eröffneten Handlungs- und Entscheidungsmöglichkeiten erscheinen ja als *Anschlussmöglichkeiten* innerhalb eines Ablaufs. Und so besteht die Rekonstruktion einer Fallstruktur nicht in der Sammlung und Systematisierung von Merkmalen einer protokollierten Lebenspraxis, sondern darin, die Selektivität dieser Lebenspraxis in der *Rekonstruktion der Ablaufstruktur der fallspezifischen Entscheidungen* zu formulieren. Die Logik der objektiv-hermeneutischen Sequenzanalyse besteht darin, „den tatsächlichen Ablauf als eine Sequenz von Selektionen zu sehen, die jeweils an jeder Sequenzstelle, d.h. einer Stelle des Anschließens weiterer Einzelakte oder -äußerungen unter nach gültigen Regeln möglichen sinnvollen Anschlüssen getroffen worden sind. Die Kette solcher Selektionsknoten ergibt die konkrete Struktur des Gebildes“ (Oevermann 1991, S. 270).“ (Wernet 2000, S. 16)

„Für die Praxis der Interpretation führt dieser Begründungszusammenhang zu einer sehr einfachen Grundregel: die Interpretation folgt streng dem Ablauf, den ein Text protokolliert. Eine entsequentialisierte Textmontage ist unzulässig.
Die interpretatorische Grundhaltung ist wiederum die, den Text als Text ernst zu nehmen, ihn also nicht als Steinbruch der Information oder als Jahrmarkt der Bedeutungsangebote auszuwerten, den Text also nicht „auszuschlachten“. Wenn der Text nämlich als Protokoll sozialer Realität angesehen wird, und wenn es um die Rekonstruktion dieser Realität gehen soll, dann gilt es, dem Text in der Interpretation gerecht zu werden.“ (ebd., S. 27)

Die methodologischen Grundüberlegungen zur Sequentialität wurden bereits im Abschnitt zu Strukturalismus und Sequentialität (2.1.2.) diskutiert.

2.2.4. Extensivität („Interpretiere alle Textelemente und bilde alle Lesarten!“)

„Das wohl auffälligste Charakteristikum der objektiv-hermeneutischen Textinterpretation besteht darin, dass die Analysen geringe Textmengen bearbeiten, dass sie sich diesen wenigen Texten aber sehr detailliert und geradezu akribisch widmen. Dieses Vorgehen provoziert zwei regelmäßig wiederkehrende Einwände: (1) Nie werde das Datenmaterial vollständig gewürdigt, sondern bloß punktuell betrachtet. [Die Methode könne damit] dem Datenmaterial nicht gerecht werden ... (2) Die akribische und detaillierte Interpretation führe zu einer artifiziellen Überprägnanz, indem sie an Nebensächlichkeiten sich in einem völlig unangebrachten Explikationsaufwand verausgabe.“ (Wernet 2000, S. 32)

Diese Argumente des Zuviel und des Zuwenig kennzeichnen einen wesentlichen Unterschied zu Alltagsinterpretationen von Texten und damit von Hermeneutiken, die ihr Vorgehen aus diesem Alltagsvorgehen ableiten. Die Möglichkeit, sich auf die „punktuelle" Auswahl von Textstellen aus dem Datenmaterial zu beschränken, ergibt sich aus der bereits beschriebenen Strukturvorstellung und einer damit verbundenen dialektischen Vorstellung von der Gesamtheit eines Phänomens. Die Extensität, also die akribische und detaillierte Feinanalyse ergibt sich einerseits als Kehrseite einer „punktuellen" Textauswahl (1), andererseits aber aus dem Ernstnehmen des Textes als Protokoll und Ausdrucksgestalt einer sozialen Praxis (2):

zu (1)

„Dem Prinzip der extensiven Feinanalyse liegt methodologisch die Annahme zu Grunde, dass sich in den protokollierten Ausschnitten sozialer Realität ein Allgemeines rekonstruieren lasse. Der Begriff der *Totalität*[29] verweist darauf, dass es keinen Sinn macht, von gesellschaftlich *unvermittelten*, zusammenhanglosen Einzelheiten auszugehen. Jedes konkrete Phänomen ist in einen allgemeinen Zusammenhang eingebettet. Die Dialektik von Allgemeinem und Besonderem stellt in Rechnung, dass das Besondere erst auf der Folie des Allgemeinen sich bildet. In der Terminologie der Objektiven Hermeneutik ...: Die Sinnstrukturiertheit sozialer Gebilde ist nicht hintergehbar. Es gibt keine Äußerungsform eines sozialen Gebildes, das die Sinnstrukturiertheit verlassen könnte.

Wenn dem so ist, dann ist es auch möglich, prinzipiell an jeder Stelle eines Protokolls der sozialen Realität deren Struktur zu rekonstruieren. Man sieht sofort, dass der Einwand, die Interpretation nur geringer Datenmengen könne das Ganze nicht in den Blick nehmen, einem anderen Begriff des „Ganzen" geschuldet ist. Wenn das Ganze als die Vielfältigkeit der Erscheinungsformen begriffen wird und das Ziel der Forschung darin besteht, diese Vielfalt irgendwie abzubilden, dann ist der Einwand gegen die Objektive Hermeneutik berechtigt. Diese aber folgt einem *strukturanalytischen* Ganzheitsbegriff, dem es nicht darum geht, den vielfältigen Erscheinungsformen *deskriptiv* Rechnung zu tragen.

Die Rekonstruktion der Strukturlogik beansprucht, das Ganze des Gebildes im Sinne der dieses Gebilde hervorbringenden Strukturprinzipien zu rekonstruieren. Diese strukturrekonstruktive Operation lässt sich an geringen Datenmengen *vollständig* durchführen. Die Triftigkeit und Aussagekraft der *extensiven Feinanalyse* bemisst sich an der *Qualität* der Interpretation, nicht an der *Quantität* des einbezogenen Datenmaterials." (Wernet 2000, S. 32 f.)

zu (2):

„Forschungspraktisch fordert das Prinzip der Extensität eine Feinanalyse des Textprotokolls, die kein Element des Protokolls unberücksichtigt lässt. Es verbietet also die willkürliche Auswahl und Auslassung von Textelementen [und fordert zusätzlich,] ausführlich zu interpretieren. ... Die Interpretation beansprucht, sinnlogisch erschöpfend zu sein. Und das heißt wiederum, dass die gedankenexperimentellen Kontexte typologisch vollständig ausgeleuchtet werden müssen." (Wernet 2000, S. 33)

[29] „Zum dialektischen Begriff der Totalität: Adorno 1981, 16 ff." (Fußnote ebenda)

Die Sequenz ist also erst dann als erschlossen zu betrachten, wenn alle denkbaren Lesarten erarbeitet und diskutiert sind. Es stellt sich in der Forschungspraxis heraus, dass man sich bei Nichtbeachtung der Extensität verirrt: Hat man sich zu schnell für eine (falsche) Lesart entschieden, weil man nicht genügend Lesarten erarbeitet hat, dann stolpert man beim nächsten Interpretationsschritt: Wenn sich die (falsche) Lesart nun zu falsifizieren scheint, muß man zurück zur vorherigen Sequenz. In mißlungenen nachvollziehenden Interpretationen kann man beobachten, wie dann der Text aufgrund einer falschen Lesart immer selektiver weiterinterpretiert wird oder wie Scheindialektiken oder Scheinwidersprüchlichkeiten konstruiert werden. In der Objektiven Hermeneutik wird das durch die Falsifizierungsforderung, durch das Prinzip der Wörtlichkeit und durch die extensive Forderung nach Vollständigkeit und Ausgiebigkeit verhindert.

> „Das Dauerproblem, das das Extensitätsprinzip aufwirft, besteht in der Frage: wann die Lesarten erschöpft sind. Die Frage lässt sich technisch nicht beantworten, sondern nur material. Es lässt sich aber sagen: je geduldiger die Suche nach Lesarten war, umso größer ist die Wahrscheinlichkeit, dass keine neuen Lesarten hinzutreten. In der praktischen Interpretationssituation erfordert die extensive Analyse also eine Haltung, die eher dazu bereit ist, noch einmal nach neuen Lesarten zu suchen, als der Ungeduld nachzugeben, endlich weiter zu kommen.“ (Wernet 2000, S. 34)

2.2.5. Sparsamkeit („Bilde nur Lesarten, die ohne Zusatzannahmen auskommen!“)

> „Das Prinzip der Sparsamkeit der Interpretation leitet sich aus den bisher erläuterten Prinzipien ab. Es schreibt vor, dass nur solche Lesarten gebildet werden dürfen, die ohne weitere Zusatzannahmen über den Fall von dem zu interpretierenden Text erzwungen sind. Der erste Schritt einer Bedeutungsexplikation besteht darin, Geschichten zu erzählen, in denen der zu analysierende Text auftauchen könnte [...]. Dabei ist es das Gebot des Prinzips der Sparsamkeit, diejenigen Geschichten auszuschließen, die darauf angewiesen sind, fallspezifische Außergewöhnlichkeiten zu unterstellen. Zwingend sind nur diejenigen Geschichten und Lesarten, die „ohne weiteres“ mit dem Text kompatibel sind.
>
> Der Terminus „Sparsamkeit“ deutet an, dass es hier zunächst darum geht, der extensiven Feinanalyse umfangslogisch Grenzen zu setzen und Extensivität nicht mit einer ziel- und endlosen Bedeutungssuche zu verwechseln. Das Prinzip hat also durchaus die forschungsökonomische Dimension, die seine Bezeichnung nahe legt. Seine eigentliche Bedeutung ist aber keine forschungsökonomische, sondern eine forschungs*logische*. Es trägt dazu bei, den Interpreten auf den Text zu verpflichten.“ (Wernet 2000, S. 35)

Ohne Sparsamkeitsforderung landet man schnell bei Unterstellungen. So kann man für jeden Text annehmen, er sei auf einer Bühne gesprochen worden. Auch die Annahme einer Pathologie gibt eine schnelle Scheinerklärung für eine schwer zu interpretierende Textsequenz. Alles das mag im Einzelfall sogar stimmen - aber die Interpretation will eine Fallstruktur ja erst entlang geltender Regeln aus dem

Text herausarbeiten und nicht bereits voraussetzen.[30] Objektive Hermeneutik gewährleistet die methodische Kontrollierbarkeit ihrer Interpretationen aber u.a. dadurch, dass sie sich ausschließlich auf den Text als Ausdrucksgestalt der zu untersuchenden sozialen Praxis beruft. Annahmen, auf die der Text nicht selbst verweist, würden genau diese Ausschließlichkeit und damit die methodische Kontrolle verletzen.

> „Vor allem soll die Einhaltung des Prinzips der Sparsamkeit verhindern, dem Fall voreilig und textlich unbegründet „Unvernünftigkeit“, „Regelverletzung“ oder „Pathologie“ zu unterstellen.
>
> Nicht etwa, weil solche Unterstellungen ethisch unstatthaft wären (das Prinzip der wörtlichen Interpretation zwingt ja geradezu zu solchen „Unterstellungen“), sondern weil und insofern der Text sie nicht erzwingt. Das Sparsamkeitsprinzip will also nicht nur die Lesartenoptionen auf die Regelgeleitetheit verpflichten, sondern will auch und vor allem die *Fallstrukturhypothesen*, also die Schlussfolgerungen über die besondere Beschaffenheit des Falles, die von einem Text protokolliert wird, an dieses Textprotokoll binden. Fallmutmaßungen, die sich vom Text in assoziativer Beliebigkeit fortbewegen, gilt es zu unterbinden. Auch hier kommt wieder das Ziel einer methodisch kontrollierten und insofern objektiven Interpretation zum Ausdruck.
>
> [...]
>
> Aus forschungspsychologischer Perspektive stehen sich die Prinzipien der Wörtlichkeit, Kontextfreiheit und Extensivität einerseits und der Sparsamkeit andererseits eigentümlich entgegen. Insbesondere das Wörtlichkeitsprinzip ist dort von besonderer Bedeutung, wo der Interpret dazu neigt, „Fünfe-gerade-sein-zu-lassen“ und fordert ihn dazu auf, die tatsächliche Gestalt des Textes reichhaltig auszudeuten. Es ermutigt zu weitreichenden Schlussfolgerungen. Das Sparsamkeitsprinzip dagegen wendet sich gegen die Tendenz, weitreichende Schlussfolgerungen unbegründet und voreilig zu ziehen. Die alltagsweltlichen Gewissheiten neigen dazu, sich dem Datenmaterial überzustülpen. Das Sparsamkeitsprinzip arbeitet dieser Tendenz entgegen.
>
> Daraus ergibt sich ein für die Objektive Hermeneutik charakteristischer Interpretationsduktus:
>
> Die objektiv-hermeneutische Textinterpretation erfordert einerseits die Bereitschaft, riskante und folgenreiche Hypothesen aus einer akribischen Textanalyse zu gewinnen, und verlangt andererseits weitestestgehende Zurückhaltung bezüglich textlich nicht zwingend indizierter Mutmaßungen.“ (ebd., S. 37 f.)

[30] Zugespitzte Beispiele für solches Vorgehen finden sich in der Geschichte der Psychatrie, wenn zu schnell Pathologien unterstellt werden. Jeder weitere vom Individuum erzeugte Text wird dann nur noch auf Kompatibilität mit der Annahme untersucht. Da fast jedes Verhalten auch mit einer Pathologiedeutung kompatibel ist, begibt man sich in einen Interpretationskreislauf, aus dem quasi kein Entkommen mehr ist, wenn nicht eine klare Falsifikationspflicht besteht.

Kapitel 3: Testen. Eine Fallbestimmung.

In diesem Kapitel wird ein Arbeitsschritt vollzogen, der in der Objektiven Hermeneutik als „Fallbestimmung“ bezeichnet wird. Hierbei geht es darum, die sozialen Strukturen auszubuchstabieren, in die der Fall - für diese Arbeit also mathematische Leistungstests bzw. die Testaufgaben - eingebettet ist. Das Ziel ist, „das Forschungsinteresse, das der Auswahl und Interpretation der Texte zu Grunde liegt, möglichst klar zu explizieren. Auf welchem theoretischen Hintergrund, mit welchen Annahmen und mit welchen Fragen, mit welchen Gewissheiten und Ungewissheiten wenden wir uns den empirischen Phänomenen zu? ... Es geht also darum, die theoriesystematische Bedeutung des Interpretationsvorhabens zu klären.“ (Wernet 2000, S.53 f.) Dieses Kapitel umfaßt die Fallbestimmung nur für mathematische Leistungstests im allgemeinen. Bei der Interpretation einer einzelnen Testaufgabe tritt als Fallbestimmung die Generierung von Lösungswegen sowie eine „stoffdidaktische Beschreibung“ der Aufgabe hinzu. Bei der Betrachtung eines Tests wiederum muß man den durch die Testersteller selbst gesetzten inhaltlichen Anspruch, also das, was sie selbst zu testen anstreben, näher betrachten.

3.1. Testaufgaben als Ausdrucksgestalten bzw. Protokolle einer sozialen Praxis

Die Kategorie der *Ausdrucksgestalt* beinhaltet alles, worin sich die erfahrbare soziale Welt präsentiert. Der Begriff *Protokoll* betont den Gesichtspunkt der ausdrucksmaterialen, überdauernden Objektivierung.

> „Dabei kann es sich um gegenständliche Objektivierungen in Produkten, um hinterlassene Spuren, um Aufzeichnungen vermittels technischer Vorrichtungen, um intendierte Beschreibungen, um institutionelle Protokolle oder um künstlerische oder sonstige bewußte Gestaltungen handeln und die Ausdrucksmaterialität kann sprachlich oder in irgendeinem anderen Medium der Spurenfixierung oder der Gestaltung vorliegen.“ Ein protokollierter Vorgang steht uns nie selbst zur Verfügung, man kann zur Kritik eines Protokolls stets nur ein anderes Protokoll heranziehen. „Auch unsere <u>Erinnerung</u> an [einen] Vorgang ist ein solches Protokoll und unser <u>Gedächtnis</u> taugt als Abgleichfolie erst dann, wenn wir es in eine prinzipiell erzählbare Erinnerung umgewandelt haben.“ (Oevermann 1996, S. 2 f.)

In dieser Arbeit werden Testaufgaben untersucht. Sie werden in einem Prozeß der Testerstellung generiert und sind damit Protokoll der sozialen Praxis der Testerstellung. Die Begrenztheit der Erschließungskraft dieses Protokolls für bestimmte Fragestellungen ist offensichtlich: So kann man zwar aus den Testaufgaben die Durchsetzung bestimmter Positionen oder Gruppierungen der beteiligten Scientific Communities herausarbeiten, für die Untersuchung des *Prozesses* der Durchsetzung wären aber z.B. Mitschnitte oder Gedächtnisprotokolle von Arbeitssitzungen, Veröffentlichungen der Testkollektive oder Arbeitsanweisungen Protokolle mit größerer Erschließungskraft.

Diese Arbeit will keine Aussage über diese Praxis der Testerstellung treffen - etwa in Form von „Soziogrammen“ oder „Psychogrammen“ der in den Testaufgaben geronnenen Dispositionen und Prozesse der Gemeinschaft der Testersteller oder einer dahinterstehenden kulturellen Gemeinschaft. Auch an jenen Stellen, an denen der Text solche Informationen geradezu aufdrängt[31], werden lediglich Folgen sich zeigender Dispositionen für den Charakter der Aufgabe als Meßinstrument thematisiert.

Anders verhält es sich mit dem von den Aufgaben transportierten Bild von Mathematik. Hier verschränkt sich der Charakter der Testaufgaben als Initiator einer Testbearbeitungspraxis unmittelbar mit einer Mitteilung über die Testersteller: Jede Testaufgabe setzt nicht nur eine Testanforderung, sondern protokolliert gleichsam,
- was die Testersteller für testenswert halten,
- welche Vorstellung von einem Test sie verfolgen
- und welches Konstrukt von dem Gegenstand des jeweiligen Meßprozesses sie haben.

Das gilt für jeden Test, wird im Fall von PISA aber dadurch zugespitzt, dass die Aufgaben explizit normativ wirken sollen. In der Diskussion über die Aufgaben führt dieser Umstand immer wieder zu Mißverständnissen, weil die Aufgaben nicht nach ihrer Eignung als Meßinstrument hinterfragt werden, sondern lediglich daraufhin betrachtet werden, ob sie „schöne Aufgaben" im Sinne einer didaktischen Norm sind. Diese beiden Ebenen sind bei der Analyse von Testaufgaben aber zu trennen: Die Untersuchung der Eignung als Meßinstrument fragt, ob gemessen wird, was gemessen werden soll. Die Untersuchung der Aufgabe als „didaktisches Instrument" fragt unabhängig davon, ob die didaktische Intention im objektiven Aufgabentext umgesetzt ist. Die Aufgabe „Bauernhöfe" (siehe Abschnitt 5.5.) ist ein beredtes Beispiel dafür, dass das didaktische Konzept zwar im manifesten Aufgabentext noch zu finden ist (wenngleich bereits dort gebrochen), dass es aber im objektiven Aufgabentext völlig zerstört ist.
Bei TIMSS behaupten die Testersteller, dass der Test ein „Kerncurriculum" (siehe Baumert, Lehmann u.a. 1997, S.179-188) wiedergeben würde. Das von den Aufgaben transportierte Bild von Mathematik protokolliert also, welches gemeinsame Bild von Mathematik die Testersteller den Curricula der beteiligten Länder zuschreiben. Mit dieser Betrachtung des Protokollstatus der Aufgaben zeigt sich bereits eine Grenze von so konstruierten internationalen Leitungstests: Das „Kerncurriculum" wird durch Analyse von Lehr**inhalten** erstellt. Ein Lerngegenstand wird dem Kerncurriculum zugeschlagen, wenn mindestens 60 Prozent der getesteten Schüler ihn laut Lehrplan behandelt haben sollten. Hier liegt die erste Störung der Gültigkeitsaussage des Tests - auf der reinen Inhaltsebene (siehe Keitel/Kilpatrick 1998). Die zweite Störung liegt nun in der Art und Weise des Umgangs mit Mathematik bzw. des vermittelten bzw. aufgenommenen Bildes von Mathematik, welches in starkem Maße latent vermittelt wird und auch in den Aufgaben implizit bleibt. Wenn die Theorie der Aufgabenersteller diese Ebene nicht kennt und sich nur an den Äußerlichkeiten des Curriculuminhalts orientiert, dann wird ihre eigene Auffassung von Mathematik in den Aufgaben unreflektiert mitprotokolliert. Das ist bedenklich, wenn diese Aufgaben dann als didaktische Norm dargestellt und genutzt werden, wenn sie zur Eichung anderer Tests oder als Instrumentarium für weitere wissenschaftliche Untersuchungen dienen. Dieser Faktor bevorzugt Schüler aus jenen Ländern oder Gruppenkontexten, welche die Aufgaben einbrin-

[31] Beispiele: TIMSS A2, im Bild eine Waage: „... *Auf der linken Seite befinden sich ein Gewicht (eine Masse) von 1 kg und ein halber Ziegelstein.* ..." PISA Bauernhöfe: „.... *Die Balken, die das Dach stützen, sind die Kanten des Quaders (rechtwinkliges Prisma) EFGHKLMN.* ..." Hier regt sich natürlich schon ein gewisses Interesse, die Genese solcher Textelemente zu verstehen.

gen.

Das von den Aufgaben transportierte Bild von Mathematik ist also immer auch ein Protokoll des von den Testerstellern in die Aufgaben eingebrachten Bildes von Mathematik. In der Interpretation werden Elemente dieses eingebrachten Bildes herausarbeitet. Nun sind verschieden weitgehende Deutungen darüber möglich, über welche soziale Praxis die Interpretation eine wie weit tragende Aussage zuläßt:

Im Beispiel der internationalen PISA-Aufgabe „Bauernhöfe“ wird z.B. herausgearbeitet, dass dort das Mathematische in seiner Autonomie gegenüber dem Realen zerstört wird und dass der mathematische Anspruch gemeinsam mit der Modellierungsanforderung der Aufgabe zerstört wird. Die inhaltliche Einfachheit der Aufgaben wird dort durch sprachliche Verschwierigungen, sprachliche Irritationen bzw. Verwerfungen und pseudomotivationale Aufblähungen zugedeckt.

- Die am wenigsten weit gehende Deutung ist nun die, dass es sich um ein Problem der speziellen Aufgabe handelt, also quasi um einen Patzer der Testerstellungsgemeinschaft. Dagegen spricht, dass von 31 internationalen Aufgaben gerade diese (und zwei weitere) ausgewählt wurde, um als „typisches Beispiel für den *Realistic Mathematics Education*-Ansatz“ (PISA 2000, S.151) zu stehen.

- Für eine weiter gehende Deutung spricht ein Blick auf die Dialektik von Allgemeinem und Besonderem. Das Besondere (die Aufgabe) steht für das Allgemeine (den Gesamttest) und ist gleichwohl ihm nicht gleichzusetzen, ebenso umgekehrt, sie sind gegensätzlich und gleichwohl untrennbar einander verbunden. Gleichzeitig ist das Allgemeine auch hier mehr als eine Summe von Besonderem, gleichwohl es sich nur im Besonderen zeigen kann. Die Rede vom Patzer ist also genauso undialektisch wie die Annahme, man hätte mit einer Aufgabe bereits vollständig beschrieben, welches Bild von Mathematik die Testersteller zeichnen bzw. vertreten

Den Prozeß der Protokollierung einer im Testerstellungskollektiv verankerten Disposition in den Aufgaben kann man sich auch sehr konkret vorstellen: Jede Aufgabe hat mitsamt dem durch sie transportierten Bild von Mathematik den gesamten Testerstellungsprozeß durchlaufen. Sie hat dabei diverse Adaptionen und Diskussionen erfahren, ist damit ein Protokoll dieses kollektiven Prozesses und erzählt damit etwas über dieses Kollektiv. Diese Vorstellung führt bereits erheblich über die „Patzerdeutung“ hinaus.

- Eine noch weitergehende Deutung bedarf einer Berufung auf strenger strukturalistisches Denken, wiederum in der Dialektik von Besonderem und Allgemeinem sich bewegend: Betrachtet man das in der Interpretation herausgearbeitete Bild von Mathematik als Teil einer Struktur „Mathematikbild“, die sich in der Aufgabe wiederfindet und die sich im Testerstellungskollektiv wiederfindet, dann wird deutlich, dass diese Struktur sich auch in den im Kollektiv vorhandenen und wirkenden individuellen und sozialen Strukturen wiederfindet. Insofern erzählt uns die herausgearbeitete Struktur „Mathematikbild“ auch etwas über Didaktikergemeinschaften und über Mathematikunterricht.

Diese sehr weit gehende Deutung bedarf strukturalistischen Denkens, wohingegen sich die Aussagen über das Testerstellungskollektiv unmittelbar auch ohne strukturalistisches Denken ergeben, und nur dialektisches Denken voraussetzen. Ohne dialektisches Denken verbleibt nur eine Möglichkeit, um über eine „Patzerdeutung“ hinaus zu kommen: Man muß alle Aufgaben interpretieren oder einen beliebigen Anteil an Aufgaben setzen, ab dem man eine Interpretation als für den Gesamttest und damit für das Testerstellungskollektiv „repräsentativ" bzw. als einleuchtend ansieht. .

3.2. Testaufgaben als Initiator einer Lösungspraxis

Tests werden nicht erstellt, um Testerstellung zu protokollieren, sondern um Eigenschaften von Meßobjekten in einem Meßprozeß zu erfassen. In dieser Arbeit werden Instrumente zur Messung zweier verschiedener Konstrukte von „Mathematischer Leistungsfähigkeit" untersucht. Die Komponenten des Instruments „Test" interessieren dabei nicht als Protokoll einer Testerstellungspraxis, sondern als Initiator einer Lösungspraxis, denn beim Testen ist Lösungspraxis Leistungspraxis, der Meßprozeß vollzieht sich also, indem der Schüler ein Resultat bzw. sein Lösen verschriftlicht. Es ist also zu fragen, inwieweit Aussagen über das Meß-instrument - also die Aufgabe - Aussagen zulassen

- über die Lösungspraxis bzw.
- über die vollbrachte und damit gemessene Leistung bzw.
- über psychische Dispositionen, also Leistungsfähigkeiten im weiten Sinne.

Im Mathematikunterricht lassen sich Aufgaben unter verschiedenen Aspekten betrachten, sie sind also Bestandteil verschiedener Praxen: Sie sind Initiator von Lernen, Ort der Selbsttätigkeit des Schülers, Instrument der Erfassung von Lernerfolgen, Zwischenerfolgen und Defiziten (also Testinstrument), Protokoll einer fachlichen, habituellen bzw. psychischen Disposition des Aufgabenerstellers (Schüler lernen ihre Lehrer u.a. über die Aufgaben kennen, die sie stellen; Schüler erhalten über Aufgaben ein Bild vom Fach und seinen Vertretern), sie sind an der Herstellung von Sozialität beteiligt. Das Lösen von Aufgaben im Unterricht ist Ort des Trainierens von Fertigkeiten, des Erlangens von Fähigkeiten, des Erwerbs von Wissen, es ist Disziplinierungsinstrument, dient der Überbrückung von Zeitlücken, ist Selbstzweck usw.

Dieser Vielfalt von Einsatzmöglichkeiten, Funktionen und Einsatzaspekten steht bei Testaufgaben eine einzige Funktion gegenüber: **Hier sind Aufgaben ausschließlich Meßinstrument.**[32] Die einzige für den Meßvorgang relevante Praxis ist die Lösungspraxis. Der erste Bruch zwischen Getestetem und zu Testendem liegt bereits innerhalb der Lösungspraxis, denn es wird nur der schriftlich fixierte Teil

---

[32] Dieser singuläre Charakter bringt den Vorteil mit sich, dass Testaufgaben als Untersuchungsobjekt nicht dem „Einwand der produktiven Wendung" unterliegen: Wenn man Aufgaben interpretiert, die im Unterricht eingesetzt werden, wird gelegentlich ein Scheinargument gegen jede Interpretation angebracht: Der Charakter einer Aufgabe sei letztlich nicht so entscheidend, denn Unklarheiten könnten geklärt, Verwerfungen im Gespräch geheilt werden, überhaupt unterliege der Umgang mit Aufgaben der Spezifik des je individuellen Umgangs mit den Aufgaben durch den Lehrer. Das stimmt, der Einwand verstellt aber den Blick auf das produktive bzw. zerstörerische Potential, das Aufgaben für den Unterrichtsprozeß in sich tragen und das man in Stundenmitschnitten auch beobachten kann. Das produktive Wenden einer problematischen Aufgabe ist eben ein schwieriger Akt, der professionelle Distanz zum eigenen Tun und zu Aufgaben verlangt, die man ja für den Unterricht eigentlich auswählt, weil sie das eigene Unterrichtsziel zu stützen scheinen. Noch dazu ist die Verworfenheit einer Aufgabe in einer handlungsbelasteten Situation höchstens ansatzweise zu dechiffrieren. Man erkennt auch sofort, welcher Fehlschluß hinter dem Einwand steckt: Der handlungsentlastet erworbenen Deutung einer Aufgabe als schulmathematisch einen objektiven Charakter tragendem Objekt wird weniger Erkenntniskraft zugestanden als einer subjektiven Deutung. Dabei wird verkannt, dass der objektive Charakter verstanden sein muß, um die subjektive Deutung verstehen zu können.
Testaufgaben unterliegen nicht der „Chaotik" von Unterrichtsgeschehen. Kein Lehrer kann sie „produktiv wenden". Deshalb muß zwischen einer für Unterricht produktiven Aufgabe und einer gut messenden Aufgabe unterschieden werden, aber deshalb unterliegt die Interpretation von Testaufgaben auch keinem „Heilungsverdacht".

des Ausflusses der Lösungspraxis erfaßt. Der Tester kann also weder gedachte noch gesprochene noch haptische Teile der Lösungspraxis in seine Auswertung einbeziehen. Erst der zweite Bruch ist dann die Verkürzung des Verschriftlichten in einer standardisierten Kodierungs- und Auswertungsprozedur, in der nicht nur das Latente, sondern auch alles andere nicht in die Standardisierung Passende weggeschnitten wird.

Umgekehrt erfolgt die Initiierung der Lösungspraxis durch die Aufgabe und durch die Testrahmung. Die Aufgabe tritt dem Schüler als sprachliches Objekt und damit als Objektivierung des Testers entgegen. Hier erfährt er, was er tun soll. Und ausschließlich aus der Erfüllung dieses Sollens erwächst das Urteil über den Schüler bzw. der Meßvorgang. Deshalb muß die Analyse des Meßvorgangs und der Folgen des Meßvorgangs für den Schüler bzw. für die Gemeinschaften, die durch Tests den Schüler beurteilen und ihn gleichzeitig prägen, an der Aufgabe ansetzen. Dabei muß unterschieden werden zwischen der Analyse des eigentlichen Meßvorganges und den gesellschaftlichen Prämissen und Folgen des Messens.

3.3. Testen und Gesellschaft

Schule hat nicht nur eine Erziehungs- und Bildungsaufgabe, sie dient auch der Vergabe von Zukunftschancen mittels Bildungszertifikaten. Es besteht das Ideal, dass Bildungszertifikate widerspiegeln, inwieweit der Schüler die gesellschaftlich vorgegebenen Bildungs- und Erziehungsziele erreicht hat. Die Abweichungen von diesem Ideal sind aus der breiten Debatte um Schulzensuren bekannt, und das Hauptziel von Tests ist es, eine Zertifizierung vornehmen zu können, die das Erreichen der Bildungs- und Erziehungsziele valider und genauer widerspiegelt als die herkömmlichen Zertifizierungsverfahren.

Wenn Aufgaben im Testvorgang nun ausschließlich als Meßinstrument relevant sind, dann führt dies nicht nur zu der Frage, ob das Instrument mißt, was es messen soll - diese Frage wird in dieser Arbeit für jede betrachtete Aufgabe beantwortet. Es führt auch zu Fragen nach den gesellschaftlichen und fachinhaltlichen (und damit nach pädagogischen) Prämissen und Folgen des Testens.

Offensichtlich sind die Folgen, wenn der Test unmittelbar zu einer folgenreichen Zertifizierung führt: Der Getestete muß die singuläre Funktion der Testaufgabe erkennen und bedienen. Ein Schüler, der z.B. eine darüber hinausgehende Auseinandersetzung mit dem Aufgabeninhalt führt, weil er Aufgaben als Ort des Lernens bzw. der Debatte bzw. der Herstellung von Sozialität mit dem Auswerter ansieht, verliert Zeit bzw. arbeitet im schlechtesten Fall am Kern vorbei: nämlich sein Kreuz an der richtigen Stelle zu setzen - der Scanner schaut sich eben keine intellektuell interessanten Randbemerkungen an, sondern erfaßt, ob und wo das Kreuz gesetzt wurde; und der Auswerter einer offenen Antwort kann diese Antwort lediglich kategorisieren und einen Punkt vergeben oder verweigern.

Zunächst ist es also sinnvoll, den personal folgenreichen Test vom personal folgenlosen Test zu unterscheiden und verschiedene Fragen an diese Tests zu formulieren. Ist ein Test personal folgenlos, also der Test einer Population, so steht nur die Frage: Wird das gemessen, was gemessen werden soll? Ist ein Test personal folgenreich, also der Test einer Person, so muß man zusätzlich fragen: Wird das prämiert, was prämiert werden soll?

In meinen Aufgabenanalysen laufen beide Fragen zusammen. Dies liegt nicht vorrangig daran, dass TIMSS und PISA, die eigentlich personal folgenlos sein sollen (und auch von ihrer statistischen Konstruktion her keine individualen Testinstrumente sind), in der Gefahr stehen, als personal folgenreiche Testinstrumente fehlgebraucht zu werden. Es liegt vorrangig daran, dass diese Tests als politische Instrumente bereits einen Folgenreichtum in sich bergen. Dass dieser Folgenreichtum kein personaler ist, sondern sich auf Gruppen bezieht, erzeugt für diese Tests lediglich eine spezifische Folgenhaftigkeit:

Von Seiten desjenigen, dessen Zukunftschancen mit dem Testresultat beeinflusst werden, geht es ausschließlich darum, die Anforderung des Testers zu erfüllen. Dies gilt für Schüler, deren persönliche Bildungszertifikate vom Test abhängen. Es gilt für Lehrer bzw. Lehrerkollektive, deren Testergebnisse materielle oder nichtmaterielle Zuwendung auslösen. Und es gilt für Länder, Populationen bzw. Teilpopulationen, für die die Testergebnisse folgenreich sind, z.B. für kollektive Selbst- und Fremdbewußtseinsausprägungen und politische Entscheidungen. An dieser Stelle ordnen sich die Diskussionen um die Verschränkungen der Ebenen der Bedeutsamkeit der Tests nach TIMSS (insbesondere im Vergleich mit Japan) ein: Für deutsche Schüler haben nichtzensierte Leistungstests innerhalb des eigenen Bildungszertifikatserwerbs und damit für die Vergabe von Zukunftschancen bisher keine Bedeutung. Dies ändert sich bereits auf der Ebene der Schulen, wenn Bildungsadministrationen die Schulresultate zur Kenntnis erhalten. Auf der nationalen Ebene wurden TIMSS wie PISA in Deutschland gar als nationale Katastrophen wahrgenommen - und die Diskrepanz dieser nationalen Bedeutsamkeit zur mangelnden Bedeutsamkeit für den einzelnen Getesteten als ein Grund des nationalen Scheiterns beklagt.
Man kann sich nun beliebige (schwer kontrollierbare) Szenarien des Umgehens mit dieser Diskrepanz denken, z.B. Schüler, die sich als Mitglieder der deutschen Testnationalmannschaft besonders anstrengen oder Schüler, die die Testergebnisse negativ manipulieren, weil sie kollektive negative Resultate für wünschenswert halten, oder auch Auswerter, die in die eine oder die andere Richtung tendenziös auswerten. Auch auf der Ebene der Schule, deren Mittelzuweisungen von Testresultaten abhängen, kann der Test für den einzelnen Schüler zum Mittel der Loyalitätsbekundung wie auch zum Mittel der Kritik an den Verhältnissen werden.
Hinzu kommt, dass auch die sich selbst wissenschaftlich rahmenden Leistungstests auf eine *Kultur* der persönlichen Bedeutsamkeit von Tests Bezug nehmen und in ihr Rückhalt finden: Wer gewohnt ist, sich für Zensuren einer Leistungs„messung“ zu unterwerfen, unterwirft sich auch einem Test, bei dem er eigentlich jederzeit gehen kann. Die Tester versuchen auch, eine an zensurrelevante Leistungsermittlung angelehnte Rahmung zu schaffen. Der Testleiter tritt den Schülern nicht als Vertreter einer wissenschaftlichen Praxis, sondern als eine Art zentraler Prüfer entgegen. Dadurch und durch die strenge Durchführungsnormierung wird der Rahmen „verschärft“, das wirkt der „Entschärfung“ durch Nichtzensierung, also der potentiell größeren Autonomie eines Getesteten gegenüber einem Zensier-

ten, entgegen.

Von Seiten des Testers geht es darum, im Meßvorgang wirklich das zu messen, was gemessen werden soll. Dies ist ein nicht lösbarer Widerspruch zum Anliegen des Getesteten, denn zugespitzt interessiert den Tester das Bildungszertifikat des Getesteten ebenso wenig wie den Getesteten das Konstrukt des Testers. Je entfremdeter Tester und Getesteter sowie Lernprozeß und Leistungsprozeß sind, desto stärker wirkt und zeigt sich dieser Widerspruch. Wie viel weniger er z.B. in einer Klassenarbeit wirkt, zeigt sich in der Begründungslast, die ein Lehrer trägt, welcher die gleiche Klassenarbeit mehrfach verwendet, seinen Tests also nicht auf die jeweilige Lerngruppe und Lernprozesse hin spezifiziert.

Umgekehrt fußt die Teilnahme an einem Leistungstest immer auf einem in je verschiedener Weise bildungszertifikatorientierten Denken, und die Durchführung eines Leistungstests ist ebenfalls ausschließlich innerhalb bildungszertifikatorientierten Denkens möglich - außerhalb bildungszertifikatorientierten Denkens findet Auseinandersetzung mit intellektuellen Gegenständen immer in einer Weise statt, die sich von Leistungstests entfernt.[33] Selbst die Durchführung von Leistungstests in wissenschaftlichen Untersuchungen verweist in letzter Konsequenz auf einen dort wirksamen bildungszertifikatsorientierten Leistungs- und Lernbegriff.[34]

[33] Dies merkt man selbst in Situationen der spielerischen Durchführung von Leistungstests, z.B. in Zeitschriften. Die narzistische Befriedigung ergibt sich hier nicht aus der Leistungserbringung selbst, sondern daraus, dass man einen äußerlich gesetzten Anspruch erfüllt und dafür eine Belohnung erhält. Es geht also nicht um die Sache selbst, sondern um das Erteilen einer äußerlichen Anerkennung. Die narzistische Befriedigung ist deshalb gering, wenn der Test so leicht ist, dass man nicht mal nachzusehen braucht, ob die eigene Lösung richtig ist. Die (hier eventuell nur scheinbare) Asymmetrie zwischen Tester und Getestetem ist dann nicht groß genug, der Tester verringert damit sein Recht auf Anerkennen und Zertifikatserteilung.

[34] Es scheint weder historisch noch strukturell zufällig zu sein, dass subsumtives Denken in den Wissenschaften und ein subsumtiv argumentierender Leistungsbegriff zusammenlaufen. Eben dieser „subsumtive Leistungsbegriff" in seiner in Leistungstests zugespitzten Variante scheint wesentlicher Teil dessen zu sein, was ein gewisses Unbehagen an diesen Tests hervorruft. Andererseits entspricht er in seiner Subsumtivität sowohl einem wissenschaftlichen Paradigma als auch dem herkömmlichen schulischen Leistungsbegriff: Dieser ist ebenso subsumtiv, denn im Kern wird der Schüler auch hier daraufhin „abgehakt", ob er bestimmte Leistungskategorien abdeckt. Dies zeigt sich besonders klar in diversen Leistungskatalogen, die dem Lehrer ermöglichen sollen, die Erfüllung von Leistungskategorien abzuhaken und die Haken zu einer Zensur zusammenzuziehen. Diese Subsumtion wird durch pädagogische Relativierungen immer wieder aufgebrochen und verschleiert - die damit verbundene Erklärungsnot macht die Subsumtivität umso deutlicher. Wegen der Subsumtivität des schulischen Leistungsbegriffs ist auch das Unverständnis der Tester für das Unbehagen an Tests erklärbar, denn die Tests bilden ja nur die konsequente Umsetzung dieses subsumtiven schulischen Leistungsgedankens.

3.4. Die Beschränkung der Autonomie[35] des Getesteten und des Testers durch den Meßprozeß

Strukturlogisch beruht Testen auf Asymmetrie zwischen Tester und Getestetem, die hier näher untersucht werden soll.

Die Widersprüchlichkeit des Wunsches des Testers, das Vorhandensein von Komponenten eines Konstrukts „Leistung“ zu vermessen, und des Wunsches des Getesteten, die Anforderung des Testers zu erfüllen, ermöglicht (i) erst einen Testvorgang und erschwert (ii) ihn gleichzeitig:
(i) Wären diese Wünsche nicht widersprüchlich, so käme kein Testvorgang zustande, sondern ein Aushandlungsprozeß zwischen Gleichberechtigten über einen Inhalt - entweder im Sinne eines Arbeitsbündnisses zum Lernen oder im Sinne einer wissenschaftlichen Debatte. Hier zeigt sich auch, dass sowohl beim Lernen in einem Arbeitsbündnis (z.B. in Volkshochschulen oder anderen Lernkontexten, die nicht auf einen Zertifikatserwerb hinauslaufen) als auch in einer wissenschaftlichen Debatte weder eine Zensierung noch irgendeine Form von Testen denkbar ist. Jede Form von Test, die für einen der Beteiligten nicht mehr wenigstens dem Prinzip nach debattierbar ist, zerstört das Arbeitsbündnis bzw. die wissenschaftliche Debatte hin zu einer asymmetrischen Beziehung. Jede Form von Test hingegen, die debattierbar ist, ist kein Test, sondern Gegenstand eines Aushandlungsprozesses.[36]
Mit einem Aushandlungsprozeß wäre aber kein Meßprozeß mehr möglich, es sei denn man setzte voraus, dass mindestens einer der Beteiligten nach dem Aushandlungsprozeß den Inhalt des Aushandlungsprozesses „vergessen“ und sich dann dem Meßprozeß aussetzen würde. Erst eine Asymmetrie zwischen den Beteiligten ermöglicht, dass der Eine am Maßstab des Anderen gemessen wird. Diese Asymmetrie und damit das Verlassen des Aushandlungsprozesses stellt somit bereits eine Verminderung der Autonomie des Getesteten her. Sobald die Asymmetrie zusätzlich mit einer Abhängigkeit verbunden ist, muß der Eine den Anderen bedienen und ist damit zusätzlich in seiner Autonomie beschränkt. Mit je größerer personaler Bedeutsamkeit die Abhängigkeit verbunden ist, desto größer ist der Bedienungsdruck - sei es, weil man sich als Mitglied der Nationalmannschaft fühlt oder weil persönliche Zukunftschancen vom Resultat abhängen.

[35] Ich benutze hier einen Autonomiebegriff, der die Dialektik von Individuellem und Sozialem aufnimmt. Autonom sein heißt dabei immer nur, autonomer zu werden. Es heißt, sich seiner selbst und der Welt gewisser zu werden und gleichzeitig sich der Ungewißheit seiner selbst und der Welt deutlicher zu stellen. Autonom handeln heißt sich selbst stärken im sozialen Handeln und das soziale Ganze und seine Teile stärken im individuellen Handeln.
Das Verhältnis von Schule und Autonomie ist ein Gespaltenes. Schule will und soll die Autonomie des Schülers entwickeln. Das spiegelt sich bereits im Bildungsbegriff, aber auch in Begriffen wie Kompetenzentwicklung (personal, sozial, fachlich, methodisch usw.), Stärkung der Persönlichkeit, staatsbürgerliche Erziehung, kulturelle Köharenz usf. All diese und ähnliche Begriffe verweisen nicht nur auf die inhaltliche Abreicherung, die der Versuch einer quasi-endgültigen und doch allgmeinen Konkretion des Autonomiebegriffs mit sich bringen würde. Sie verweisen auch darauf, dass Schule Autonomie nicht nur entwickelt, sondern ebenso zerstört. Das liegt nicht nur am schulischen Zertifizierungs- und Selektionsauftrag. Das liegt auch an der schulischen Aufgabe der Integration von Kindern und Jugendlichen in eine Gesellschaft, die mindere oder vielleicht sogar verminderte Autonomie nicht nur produziert, sondern zur Aufrechterhaltung ihrer selbst auch benötigt.
[36] Das gilt auch für schulische Leistungskontrollen in ihrer herkömmlichen Form, wenn sie prinzipiell debattierbar sind. Nicht ohne Grund nehmen autonomieorientierte Pädagogiken Aushandlungsprozesse auch für Leistungserfassungen auf ihre Agenda. Dahinter steht nicht nur ein autonomieorientiertes Menschenbild, sondern auch ein dynamischer Leistungsbegriff.

(ii) Der Widerspruch der Anliegen der in den Testprozeß Involvierten erschwert den Testvorgang, denn die beschriebene Asymmetrie verlangt auf der Seite des Getesteten nach Ausgleich durch Vergrößerung seiner Autonomie gegenüber dem Testvorgang. Der Testvorgang wir dadurch gestört: Innerhalb der Strukturlogik der Asymmetrie beim Testen gibt es unterschiedliche Möglichkeiten der Handlungsanpassung. Die erste Möglichkeit ist Konformität, d.h. man bearbeitet die Testaufgaben „nach bestem Wissen und Gewissen". Die zweite Möglichkeit sind unterschiedliche Grade von Verweigerung. Die klarste Form von Verweigerung ist die Nichtteilnahme, also das Verlassen des Handlungsrahmens. Die dritte Möglichkeit sind verschiedene Formen der strategischen Bearbeitung des Handlungsrahmens: Der Druck, den Tester bedienen zu müssen, verlangt nach Wegen, ihn möglichst effektiv und auch unabhängig vom wirklichen Vorhandensein der zu testenden Fähigkeit zu bedienen. Er verlangt sogar, den Tester selbst dann in seinem Meßkonstrukt zu bedienen, wenn die eigentlich zu vermessende Fähigkeit im Meßkonstrukt fehlerhaft operationalisiert ist. Eine Bearbeitung des Handlungsrahmens, sogar eine gewisse Immunisierung, beginnt bereits im bewußten Umgang mit der Strukturlogik der Testsituation. Der Schüler würde sich dann einerseits seine Stellung im Testprozeß bewußt machen und die Folgen seines Testergebnisses bzw. seiner Testverweigerung in der Beurteilung seiner Person, seiner Schule, seines Bundeslandes usw. übersehen können. Er würde sich andererseits dem Problem der Autonomiezerstörung in jeder einzelnen Aufgabe stellen können, indem er einerseits Testfähigkeit entwickelt, indem er andererseits die Tester material zwingend auf Defizite der Aufgabenstellung (z.B. Zweideutigkeiten) hinweist.[37] Es ist nun nicht so, dass die Entwicklung von Testfähigkeit (siehe Schlußkapitel) einen autonomieverstärkenden Charakter hat, wie es auf den ersten Blick scheint. Sie stellt nur die Bearbeitung des asymmetrischen Handlungsrahmens dar. Wer effektive Wege findet, die Tester zu bedienen, entkommt der Autonomiezerstörung ebensowenig wie der konform handelnde. Selbst Verweigerung stärkt Autonomie noch nicht, sondern entzieht sich nur dem Handlungsrahmen ihrer Zerstörung. Durch denjenigen, der Autonomie einschränkt, kann das strukturelle Autonomieproblem des Testens (wie wahrscheinlich jedes Autonomieproblem) auf drei grundsätzliche Weisen bearbeitet werden: durch Autonomiegewährung und -stärkung, durch Verschleierung von Autonomiebeschränkung und durch offene Repression. In jeder Testsituation wird das Problem der Autonomie des Getesteten gegenüber dem Testvorgang auf jeweils spezifische Weise bearbeitet.

Die konkrete Gestaltung der Testsituation ist eine solche Form der Bearbeitung: Die hohe Standardisierung der Testsituation dient u.a. der Herstellung des äußerlich gleichen Grades dieser Ausgestaltung des Handlungsrahmens für alle Getesteten.[38] Der Umgang mit dem Autonomieproblem erfolgt andererseits in jeder einzelnen Aufgabe. Eine Aufgabeninterpretation wird also auch Aussagen über den Umgang mit der Autonomie des Getesteten treffen, muß dabei aber berücksichtigen, dass bereits die Testsituation selbst eine Einschränkung von Autonomie notwendig mit sich bringt.

[37] Gerade in Schulen mit hoher Autonomieorientierung erlebt man während Testsitzungen Schüler, die sagen: „Man kann diese Aufgabe so und so lesen, was ist gemeint?" Die Testleiter dürfen und können diese Frage nicht beantworten, da im standardisierten Prozedere keine Unterscheidung zwischen Hilfe und Suche nach („eigentlich undenkbaren") Fehlern der Aufgabenstellungen möglich ist.

[38] Eine Interpretation dieser Ausgestaltung müßte auf die Testleiterskripte, auf die dem Test vorausgehenden Briefe an die Eltern und an die Schule und auf Protokolle der schulischen Vorbereitung des Tests zurückgreifen.

Die aufgezeigte Autonomiezerstörung, die sich auch in den Interpretationen immer wieder zeigen wird, ist unabhängig vom Bewußtsein der Betroffenen. Außerdem wird sie durch die Institution Schule von Anfang an als Normalität eingeführt und von allen Beobachtern gedanklich geheilt. Ein frühes schulisches Beispiel ist die Lösung der Subtraktionsaufgabe Drei minus Vier in Klasse 1: Wenn ein Schüler hier sachlich richtig minus Eins angibt, so muß der Lehrer einen Umgang mit dem Problem finden, dass dieser Schüler mehr weiß, als er wissen darf. Oftmals erlebt man eine Rüge oder zumindest sichtbares Unbehagen, weil der Schüler nicht die „richtige" Antwort „nicht lösbar" gegeben hat. In der Diskussion mit Lehrern um diese Autonomie- (und Sach-) -zerstörung bekommt man pädagogische, heilende Argumente zu hören, z.B. dass die „schwachen Schüler" nicht „verwirrt" werden sollen. Solche Heilungen durchziehen nicht nur pädagogisches Handeln, sondern auch das Handeln der Tester: Auch von ihnen bekommt man – wenn man Autonomiezerstörungen oder andere Verwerfungen bei Testaufgaben anspricht – immer wieder zu hören, dass die Schüler schon wüßten, wie sie damit umzugehen haben. Vielleicht stimmt das sogar, wenn sich die Testaufgaben dem den Schülern vertrauten pädagogischen Habitus anschmiegen. Aber einerseits unterliegen internationale Tests verschiedenen pädagogischen Habitus. Andererseits steht eine reine Reproduktion des überkommenen pädagogischen Habitus im Widerspruch zum Anspruch der Tester, den Weg in verbesserte pädagogische Praxis weisen zu können.

3.5. Grenzen komparatistischen Vorgehens für die Erklärung von Wirkzusammenhängen

Wollte man internationale Vergleichstests wie TIMSS und PISA ausschließlich als wissenschaftliche Instrumente verstehen, so müßte man scheitern. Bereits die Frage nach der Erkenntniskraft eines Mehrländerleistungsvergleichs ist schwierig zu beantworten, denn das Wissen um Testpunktunterschiede befriedigt ja zunächst lediglich ein voyeristisches Interesse. Die Erfahrungen aus der Komparatistik (siehe Schriewer 1999) zeigen die mangelnde Erkenntniskraft von Ansätzen, die „gleichsinnig oder geradezu deterministisch wirksame makro-soziale Bedingungs- bzw. Funktionszusammenhänge" (ebenda S.63) unterstellen. Auch TIMSS und PISA suchen nach solchen multikausalen Erklärungsmustern. Selbst wenn dort im Sinne wissenschaftlicher Rhetorik die Begrenztheit solchen Denkens betont wird, bleiben im Kern doch nur Korrelationsmuster übrig, deren Interpretation zwar Erkenntnis suggeriert, aber weder Erkenntnis generiert noch methodischer Kontrolle unterliegt (siehe z.B. PISA 2000, S.183-185).

> „Entgegen solchen [Zusammenhangs-] Annahmen aber hat vergleichende Forschung eine beeindruckende internationale Variationsbreite von historisch-kulturell realisierten Problemlösungsmustern oder -strategien zutage gefördert." (Schriewer 1999, S.64) Beispielsweise sind die „Zusammenhänge zwischen Erziehung und Wirtschaftswachstum, Erziehung und Mobilisierung oder Erziehung und politischer System-Integration [...] weder direkt, noch einlinig, noch in unterschiedlichen Gesellschaften gleichförmig wirksam. Sie sind vielmehr in der Regel nur schwach ausge-

> prägt, nur partiell wirksam, dysfunktional oder schlichtweg kontraproduktiv. In jedem Fall sind sie „highly problematic“ und nur im Sinne von Interrelationen begreifbar. Solche Interrelationen sind ihrerseits eingelagert in und überformt durch weitere soziale Beziehungsmuster. [...] Den skizzierten Strängen vergleichender Forschung ist gemeinsam, daß ihre Befunde einmünden in den Aufweis einer Vielzahl variierender Interrelationsgefüge und Entwicklungspfade. Deren Komplexität läßt sich mit Hilfe von System-Modellen beziehungsweise systemisch inspirierten Typologien zwar konzeptionell fassen, sie im einzelnen aufzuhellen bleibt jedoch vergleichend-historischen Prozeß-Analysen überlassen. [...] Die im Vergleich zutage tretenden empirischen Evidenzen führen in aller Regel weniger zur Bestätigung, sondern nötigen zur Revision der als regelmäßig unterstellten makrosozialen Zusammenhänge.“ (ebd., S. 65 f.)

Auch Vergleiche wie TIMSS und PISA zeigen immer nur, dass keine allgemeingültigen Aussagen zu Wirkzusammenhängen möglich sind, obwohl die Forscher das unterstellen. So wird in der Deutung der TIMSS-Videostudie unterstellt, die hohen erreichten Testleistungen hingen mit dem in den Videos gezeigten „guten“, konstruktivistischen und kognitiv anspruchsvollen Unterricht zusammen. Bereits innerhalb der Videostudie wird ignoriert, dass nach diesem Erklärungsmuster die deutschen Schüler besser abschneiden müßten als die Schüler in den USA, weil im Vergleich der deutsche Unterricht mit dem Lehrer-Schüler-Gespräch konstruktivistischer vorgeht und insgesamt kognitiv anspruchsvoller ist. Innerhalb der Deutung der guten japanischen Testleistung als Ergebnis guten Unterrichts werden alle anderen Einflüsse von vornherein aus der Betrachtung ausgeschlossen. Wenn man umgekehrt versucht hätte, zunächst das Zusammenspiel denkbarer bzw. empirisch sich zeigender Einflußfaktoren zu analysieren, dann wäre man gar nicht auf die Idee gekommen, sich in dieser Weise auf den Unterricht selbst zu konzentrieren. Auch die ethnographischen Länderstudien (Case Studies zu Deutschland, Japan, USA) leisten wenig Erklärung, weil sie sich auf Deskriptionen beschränken. Diese Deskriptionen stützen sich zwar auf vorgegebene Beobachtungsschemata, nach denen die Verhältnisse in den drei Ländern jeweils beobachtet werden sollten. Aber auch hier zeigt sich, dass aus den konkreten Gegebenheiten der Länder heraus so verschiedene Aspekte in den Vordergrund der Betrachtung rücken, dass Vergleiche unmöglich werden, geschweige denn dass Wirkzusammenhänge deutlich würden. Entsprechend beliebig werden dann in der Interpretation der Videos und der Testergebnisse einzelne Aspekte der Ergebnisse der Case Studies herausgepickt und zu Interpretationen zusammengezogen, die sich auf keine Systematik mehr stützen können. Das wesentliche Argument für die so vollzogene „Analyse“ bleibt dann ein Testwert oder eine Korrelation, die die Argumentation angeblich stützen. Die regelmäßige Aussage „Es besteht weiterer Forschungsbedarf.“ wird dann zur Floskel, die sich äußerlich auf die Endlichkeit von Erkenntnis bezieht, im Kern aber Erkenntnis desavouiert, weil sie Erkenntnis vorgaukelt, wo keine ist und Erkenntnis auf Datenauswertung reduziert.

3.6. Operationalisierung

Das Problem der konstitutiven Asymmetrie und des damit verbundenen Widerspruchs zwischen den Anliegen von Tester und Getestetem bildet den unhintergehbaren Rahmen jedes Leistungstests und damit des Meßprozesses selbst. Die Nichtbeachtung dieses Rahmens führte zu der in Kapitel 1 beschriebenen Fehlinterpretation von Testaufgaben durch die TIMSS-Gruppe. Sie hatte nicht gefragt, auf welche Weise der Schüler zum richtigen Kreuz an der richtigen Stelle gelangt. Sie hatte sich lediglich darauf konzentriert, wie sich die Tester den Lösungsprozeß wünschen. Die Tester hatten sich also auf die von ihnen gewünschte Lösungspraxis konzentriert. Einen Meßprozeß zu verstehen setzt aber voraus, möglichst jede denkbare, durch die Aufgabe initiierte Lösungspraxis zu berücksichtigen. Die in dieser Arbeit vorgestellten objektiv-hermeneutischen Aufgabeninterpretationen stellen dafür ein Analyseinstrument dar.

Andersherum existiert für die Entwicklung eines Testinstruments eine Idealvorstellung: Das zu untersuchende Konstrukt - hier mathematische Leistungsfähigkeit - muß in ein Testinstrumentarium hinein operationalisiert werden. Da jeder Meßvorgang eine Subsumtion einer Eigenschaft des Meßobjekts unter eine konstruierte Meßgröße ist, muß sowohl die Meßgröße in ihren Facetten als auch der Mechanismus der Zuschreibung der Ausprägung einer Meßgröße zur Eigenschaft des Meßobjekts beschrieben werden. In unserem Fall müßte also zunächst dargelegt werden, was mathematische Leistungsfähigkeit ausmacht, in welche Facetten man sie sich also „zerlegt" denkt. Diese Facetten müssen zunächst noch nicht meßbar sein, denn erst nachdem man das Konstrukt konstruiert hat, erfolgt die Operationalisierung, d.h. die Festlegung des Mechanismus der Zuschreibung der Ausprägung einer Meßgröße zur Eigenschaft des Meßobjekts. In Kapitel 5 wird die Operationalisierung bei der Erstellung von PISA untersucht[39]. Obwohl die Testerstellung hier mathematikdidaktisch orientiert erfolgt und ein umfängliches Testinstrumentarium verwendet wird, gibt es kein nachvollziehbares Operationalisierungsinstrumentarium, sondern nur mehr oder weniger grobe Zuordnungen von Aufgaben zu mathematischen Fähigkeiten. Es existiert hier zwar eine differenziert ausgeprägte Vorstellung vom Konstrukt der mathematischen Leistungsfähigkeit, aber der eigentliche Akt der Operationalisierung hat nicht stattgefunden.

Die Konstruktion eines Meßkonstrukts und die Operationalisierung erfolgen in Wirklichkeit nicht nach der oben beschriebenen Idealvorstellung, sondern ambivalent, komplementär und dialektisch. Je bewußter man sich diesen Eigenschaften stellt, desto differenzierter und treffsicherer wird der Meßprozeß ausfallen, desto deutlicher wird aber auch das Bewußtsein der Begrenztheit des Meßbaren werden. Objektive Hermeneutik kann in diesem Prozeß einerseits zur Analyse des jeweiligen sprachlichen Zwischenprodukts dienen, andererseits liefert der Prozeß der Geschichten- und Lesartenbildung gerade die für eine adäquate Operationalisierung notwendigen sprachlichen Alternativen und damit neue, in ihrer sprachlichen Bedeutung klar beschriebene Formulierungen. Oevermann spricht auch

[39] Bei TIMSS gab es weder ein didaktisches Konzept noch eine Operationalisierung. Hier wurde aus gemeinsamen Unterrichtsinhalten der beteiligten Länder ein „Kerncurriculum" konstruiert und diesen Inhalten wurden Aufgaben zugeordnet.

vom „Selbstheilungspotential der Textförmigkeit sozialer Wirklichkeit“. Die Interpretationen zeigen aber auch, dass mit dem Bewußtwerden der Alternativen auch die Unschärfe und die Begrenztheit des Meßbaren bewußt werden.[40]

3.7. Aufgabeninterpretationen und Vorhersage von Lösungspraxis

Testaufgaben sind Initiator, nicht Determinator einer Lösungspraxis. Welche Schlüsse für die Lösungspraxis lassen sich nun aber aus der in der objektiv-hermeneutischen Aufgabeninterpretation herausgearbeiteten Sinnstruktur einer Aufgabe ableiten? Oevermann schreibt zur Prognosekraft von herausgearbeiteten Fallstrukturen:

> „Die fallspezifische Strukturierungsgesetzlichkeit zeigt sich in dem Maße, in dem wiederkehrend dieselben Möglichkeiten systematisch ausgeschlossen werden, die nach allgemein geltenden Regeln sinnlogisch ebenso richtig gewesen wären wie die tatsächlich selegierten. [...] Sie selbst konstituiert durch ihre als Allgemeines und Besonderes zugleich bestimmbare Selektivität die Autonomie des Einzelfalls als Strukturpotential. [...] Die Strukturierungsgesetzlichkeit erlaubt nicht deterministische Prognosen, sondern immer nur die Angabe von offenen Transformationsspielräumen.“ (Oevermann 1983, S. 274 f.)

Dies gilt auch für den umgekehrten Fall der Wirkung einer Fallstruktur. So wie in einem Dialog jeder Dialogpartner - entsprechend der bei ihm wirkenden Fallstruktur - die Möglichkeit vielfältiger, in jedem Fall bedeutsamer Reaktionen hat, so eröffnet auch die Fallspezifik jeder Aufgabe einen Raum von Handlungsmöglichkeiten. So strebt die Erarbeitung möglicher Lösungswege für eine Aufgabe zwar Vollständigkeit an, und jeder Lösungsweg verweist auf bestimmte Leistungsdispositionen. Dies bedeutet nun aber nicht, dass ein Schüler, der über die entsprechende Leistungsdisposition verfügt, diesen Weg auch wirklich einschlägt und bewältigt. Es bedeutet auch nicht, dass pures Raten (welches von intuitivem Vorgehen zu unterscheiden ist) ausgeschlossen werden kann - insbesondere bei Multiple-Choice-Aufgaben. Auch Mischwege mit vielfältig verschlungenen Fehl- und Halblösungen sind nicht auszuschließen. Wegen all dieser Möglichkeiten ist es auch wenig aufschlußreich, wenn die Schüler im Test aufgefordert werden anzukreuzen, welchen Lösungsweg sie gewählt haben. Hier zwingen die Tester den Schüler lediglich, sein eigenes Denken ihrer mechanistischen Lösungsvorstellung zu subsumieren.

[40] Mit einer präzise durchgeführten Operationalisierung wird sich auch zeigen können, welche Skalenart bei einem Test überhaupt sinnvoll nutzbar ist. Momentan scheint der Wunsch nach Intervall- und Verhältnisskalen stärker als die Reflektion über deren inhaltliche Adäquatheit zu sein. Für diese Arbeit ist diese Frage aber irrelevant, da sich bei den Interpretationen herausstellt, dass die vorliegenden Aufgaben inhaltlich nicht mal ordinal anzuordnen sind.

Auch aus den anderen in den Interpretationen herausgearbeiteten Aufgabeneigenschaften sind nur Transformationsspielräume ableitbar. So wird z.B. selbst die unlösbare PISA-Aufgabe „Kontinent“ (siehe 5.9.), die aus der Wertung genommen werden mußte, von Schülern „richtig“ gelöst, also mit dem Kreuz an der richtigen Stelle bedacht. Dies liegt nicht nur an einer angereicherten Ratewahrscheinlichkeit und daran, dass Testfähigkeit ermöglicht zu ahnen, was die Tester wollen. Es liegt auch daran, dass Nichtentscheidbarkeit zwischen zwei Lösungen auch immer die Option in sich trägt, dass der Schüler nur eine der zwei möglichen Lösungen sieht bzw. „sehen will“. Auch die TIMSS-Aufgabe A4 ist eigentlich unlösbar - es sei denn man nimmt sein Wissen über das Funktionieren von eingekleideten Aufgaben hinzu oder man erfaßt gar nicht erst die Vieldeutigkeit des Problems. In abgeschwächter Form tritt das gleiche bei Irritationen auf: Auf Irritationen kann man „reinfallen“ oder nicht. Das Hinweggehen über Irritationen kann bewußt oder unbewußt erfolgen.

Klarer wird der Begriff vom Transformationsspielraum, wenn man sich den Oevermannschen Strukturbegriff vor Augen führt:

> „Ich ziehe den Begriff „Strukturiertheit“ dem der „Struktur“ vor. Darin soll zum Ausdruck kommen, dass im Gegenstandsbereich der Soziologie mit „Struktur“ eines Gebildes immer nur dessen „Strukturiertheit“ der Sache nach gemeint sein kann, wenn nicht „Struktur“ von vornherein die Oberfläche der erstarrten Erscheinung meinen soll oder gar den trivialen bzw. leeren Formalismus einer Menge von Elementen, die in einer zu spezifizierenden Relation zueinander stehen. Eine Struktur hat ein soziales Gebilde als Ergebnis eines historischen Bildungsprozesses, der strukturtheoretisch als Prozeß der Transformation einer Struktur zu bestimmen ist, und aufgrund von Prozessen der beständigen Reproduktion dieses Ergebnisses, wobei aufgrund geringfügiger Abweichungen die Reproduktionen einer Struktur über Zeit zu mehr oder weniger deutlichen Transformationen sich kumulieren. Die Reproduktion einer Struktur ist schon immer eine Abstraktion vom allgemeineren Prozeß der Transformation. Beide, Reproduktion wie Transformation, werden durch das gekennzeichnet, was man die fall- oder typenspezifische Strukturierungsgesetzlichkeit nennen kann.“ (Oevermann 1983, S. 270)

Hier ist nun nicht nur auf die Dynamik der sozialen Praxis von Testerstellung verwiesen, die sich in Reproduktions- und Transformationsmustern in den Aufgaben zeigen. Es ist auch auf die Dynamik verwiesen, die Testaufgaben in ihrer Wirkung in subjektiven „Landschaften“ auslösen. Und auf die Dynamik, die der Fallstuktur(iertheit) des getesteten Subjekts sowieso schon innewohnt.

Es ist also nicht möglich, deterministisch von der Textstruktur auf die Lösungspraxis zu schließen. Aus der Verschränkung von Textstruktur und Lösungspraxis läßt sich auf subjektive Dispositionen schließen, erst die Mit-Analyse von Lösungspraxis läßt Schlüsse auf diese Dispositionen zu. Man kann allerdings versuchen, den Transformationsspielraum einer Aufgabe einzuschränken, um eine enger begrenzte Disposition zu erfassen. Das heißt man kann den Text so verändern, dass der Lösungsspielraum verkleinert wird. Damit ist verbunden, dass eine enger begrenzte Disposition ermöglicht, die Aufgabe zu lösen, dass also nur diese enger begrenzte Disposition gemessen wird.

Kapitel 4: Der TIMSS-Test

Die TIMS-Studie (Third International Mathematics and Science Study) wurde im Jahr 1997 veröffentlicht. TIMSS hatte fünf Komponenten (vgl. Baumert, Lehmann 1997, S. 35 f.):

1. „International vergleichende Curriculum-Studie zu Mathematik und Naturwissenschaften, zu der Lehrplan- und Schulbuchanalysen ebenso wie die Erfassung des tatsächlich unterrichteten Stoffes gehören"
2. Mehrdimensionale Leistungsstudie zu Mathematik und Naturwissenschaften - 7./8.Klasse und Abschlußjahrgang der Sekundarstufe II (an einer dritten Studie zum Ende der Grundschulzeit hatte Deutschland nicht teilgenommen); in dieser Arbeit wird ausschließlich der Test 7./8.Klasse betrachtet
3. „Befragungen der Schulleiter zur Schulorganisation und Schulkultur sowie der Fachlehrer zu ihrem Unterricht und der allgemeinen Berufstätigkeit"
4. Serie von Fallstudien (case studies), die mit ethnographischen Methoden den mathematisch-naturwissenschaftlichen Unterricht in Deutschland, Japan und den USA „ideographisch im Kontext von Schule, Schulverwaltung, Elternhaus und Jugendkultur untersuchen"
5. TIMSS Videotape Classroom Study in Deutschland, Japan und den USA, „die anhand einer großen Stichprobe von Videoaufnahmen ganzer Unterrichtsstunden unmittelbaren Einblick in den Mathematikunterricht in diesen drei Ländern erlaubt."

Die Veröffentlichung der ersten TIMSS-Ergebnisse war in Deutschland von einer für damalige Verhältnisse ungekannten medialen Inszenierung begleitet. Das mediale Interesse galt fast ausschließlich den Länderrangfolgen der Leistungsstudien. Die Ergebnisse wurden als eine Art nationale Katastrophe interpretiert. Es folgten intensive Debatten in der bildungspolitischen Öffentlichkeit wie auch in den beteiligten Fachdidaktiken. Im politischen Raum und in Teilen der Fachöffentlichkeit wurde verstärktes Testen beschlossen bzw. gefordert, eine Tendenz, die auf verschiedenen Ebenen Umsetzung erfährt und im Begriff „Testwelle" ihren Niederschlag fand. Beispielhaft seien hier nur der KMK-Beschluß zur regelmäßigen Durchführung von Vergleichsuntersuchungen, die Teilnahme aller Bundesländer an der PISA-Studie und die in Kaiser u.a. (2001) dargestellten Untersuchungen genannt. Hinzu kam eine vorher nicht gekannte wissenschaftspolitische Aufmerksamkeit für die Disziplinen rund um den mathematisch-naturwissenschaftlichen Unterricht. Im Anschluß an den „TIMSS-Schock" wurde eine Reihe von Forschungs- und Implementationsprojekten entwickelt und bewilligt, die auch der Mathematikdidaktik einen personellen Entwicklungsschub brachten.

In der mathematikdidaktischen Diskussion hat TIMSS eine so reichhaltige Aufnahme gefunden, dass die Angabe einzelner Arbeiten wenig sinnvoll ist. Das Verdienst der Studie liegt hier nicht so sehr in reichhaltigen Ergebnissen, sondern im Vorhandensein eines als relativ objektiv wahrgenommenen Diskussionsbezugspunktes. Viele Autoren ziehen TIMSS (das gleiche gilt später für PISA) als Beleg für ihre früheren bzw. jetzigen Positionen an, ohne dass diese Bezugnahmen belegt würden oder belegbar wären. Diese gewisse Beliebigkeit der Deutung der Resultate überrascht wenig: Bereits im vorigen Kapitel wurde unter Bezug auf Schriewer (1999) diskutiert, dass internationale Vergleichsstudien vorrangig die Grenzen der Vergleichbarkeit und die Unmöglichkeit kausaler Aussagen zeigen. Aus der Vielzahl der erhobenen Einzelheiten lassen sich aber Ausschnitte als „Beleg“ vielfältiger Positionen herausschneiden. Offensichtlich gibt es auch noch keine Verfahren zur sinnvollen und umfassenden Verknüpfung und Fokussierung der erhobenen Informationen und Daten. So wurden die in den Case Studies (NISA 1998, 1999a, 1999b, komprimierte Vorfassung: Stevenson und Nerison-Low 1997) ethnografisch erworbenen Erkenntnisse nie mit den anderen TIMSS-Bestandteilen zu einem stimmigen Ganzen verzahnt. Das Geflecht von Schul- und sonstiger Gesellschaftswirklichkeit ist in diesen Studien relativ dicht und vielschichtig beschrieben. Die bei Baumert und Lehmann (1997, S.22-27, 217-219) erbrachten Deutungen und schulpolitischen Schlußfolgerungen erscheinen dagegen als unzulässige Vergröberungen. In den Case Studies - und erst recht bei Hinzunahme der anderen Studien - ergibt sich ein so differenziertes Bild, dass man zwar hinterher viel über die verschiedenen Möglichkeiten gesellschaftlicher Rahmenbedingungen und schulischer Praxis weiß, aber eben auch erkennt, dass sich jedenfalls keine zwingenden Schlußfolgerungen für die Gestaltung der Rahmenbedingungen bzw. der schulischen Praxis ergeben. Genau mit solchen Schlußfolgerungen treten aber die TIMSS- und später auch die PISA-Macher auf den Markt der bildungspolitischen Deutungshoheiten.

Für die Fragestellung nach dem Charakter von TIMSS als Meßinstrument ist die Marktmacht von TIMSS wenig relevant, weil dieser Markt von der jeweils neuesten Großstudie bestimmt wird. Interessant ist TIMSS als historisches Exempel, und als geeichtes Meßinstrument, welches weiterhin für forschungsindustrielle Zwecke genutzt wird (siehe z.B. Köller, Trautwein 2001; Klieme, Bos 2000).

Das mittlerweile zu registrierende Versanden der inhaltlichen und methodischen Debatte um TIMSS (und auch um die nachfolgenden Großstudien) ist verständlich. Es erscheint wenig attraktiv, ständig abgeschlossenen Studien hinterherzuarbeiten bzw. sich dem Mainstream in seinen Lauf zu stellen. Da TIMSS aber immer wieder argumentativ herangezogen wird und der TIMSS-Test als Vorbild für eine Reihe in die Schulen wirkender Tests dient, scheint mir eine nähere Betrachtung noch notwendig. Die hier vorliegenden objektiv-hermeneutischen Interpretationen von TIMSS-Aufgaben bestätigen diese Notwendigkeit, denn es zeigen sich erhebliche Zweifel der Eignung der Aufgaben als Meßinstrument.

4.1. Zur methodischen Auseinandersetzung mit TIMSS in der deutschen Mathematikdidaktik

Sehr viel kürzer als die Liste angeblicher Schlußfolgerungen aus TIMSS ist die Liste methodischer Auseinandersetzungen mit der Studie. Dies hat sicherlich mit den nicht zugänglichen Daten, aber auch mit der geringen Halbwertzeit des Interesses für TIMSS zu tun. Die Erkenntnis- und Erschließungskraft von TIMSS für die Mathematikdidaktik scheint nicht so hoch zu sein, dass wenige Jahre nach Veröffentlichung erster Resultate noch weiterführende Auseinandersetzungen damit stattfinden würden.

Eine erste umfassende Auseinandersetzung mit dem mathematischen Teil von TIMSS leistete Weigand (1997). Er verweist zunächst auf die *mangelnde Nachvollziehbarkeit* der Studie: „Viele der im Test verwendeten Aufgaben sowie die Unterrichtsmitschnitte auf Videobändern sind noch nicht allgemein zugänglich. Es ist unklar, inwieweit manche Länder die Auswahl der Testklassen bzw. der „Videoklassen“ beeinflussen konnten. Schließlich trägt auch die Geheimhaltung der Ländervergleichstests nicht dazu bei, konstruktive Schlußfolgerungen ziehen zu können.“ (Weigand 1997, S. 518)
Zur *Behauptung mangelnder Problemlösefähigkeiten* bei deutschen Schülern bemerkt er: „Wenn von einem problemlösenden Unterricht gesprochen wird, dann denkt man an Aufgaben, die den Lernenden in der Weise fordern, daß er nicht einfach bekannte Algorithmen anwenden, sondern daß er selbst Lösungsstrategien entwickeln, anwenden und überprüfen muß. ... [Er beschreibt zwei Aufgaben, die bei TIMSS als Problemlöseaufgaben bezeichnet werden.] Mit diesen Aufgaben kann sicherlich nicht die „Problemlösefähigkeit“ von Lernenden überprüft werden. So hat auch in Amerika das Wort *problem solving* eine viel weitergehende Bedeutung als in Deutschland. Von daher sind Ergebnisse der TIMSS-Studie mit Vorsicht zu beurteilen, die deutschen Schülern eine geringe Problemlösefähigkeit zubilligen.“ (ebenda, S. 521)
Zum verwendeten *Konstrukt von „mathematischer Leistungsfähigkeit“* bemerkt er: „Geht man von dem HEYMANNschen Katalog der Aufgaben der allgemeinbildenden Schule [...] oder auch von den weitgehend akzeptierten Lernzielen des Mathematikunterrichts (Mathematisieren lernen, Begriffe bilden lernen, heuristische Strategien kennenlernen, Beweise lernen ...) aus, dann beziehen sich die zum Lösen der TIMSS-Aufgaben benötigten Fähigkeiten nur auf einen kleinen Teil dieser Ziele. Wenn zukünftig verstärkt Schule und Unterricht getestet werden sollen, dann müssen auch Methoden entwickelt werden, die es ermöglichen, das Erreichen dieser übergeordneten formalen Lernziele des Mathematikunterrichts zu überprüfen.“ (ebenda, S. 521 f.)
Er verweist weiter darauf, dass der *Einsatz neuer Technologien* bei TIMSS keine Rolle spielt: „Die Ergebnisse der TIMSS-Studie dürfen aber nicht dazu führen, den Unterricht ausschließlich auf das traditionelle Lösen von Aufgaben auszurichten und darüber die Chancen des Einsatzes neuer Technologien zu vergessen.“[41] (ebd., S. 521) Schließlich stellt er die damals weit verbreitete Forderung nach Übernahme der Unterrichtspraxis der bei TIMSS „siegreichen“ Japaner in Frage: Man müsse sich genauer ansehen, was die aus der TIMSS-Videostudie heraus interpretierten Beschreibungen japanischen Mathematikunterrichts wirklich bedeuten, was z.B. „anspruchsvolle Probleme und Beweise“ dort sind. „Es sollte auch mitbedacht werden, daß sich - wie Kenner des japanischen Schulwesens immer wieder herausstellen - die autoritäte Struktur des japanischen Schulsystems mit dem Anspruch an die Schüler, „Schulfragen zu Lebensfragen“ zu machen, auf heutige europäische Verhältnisse nicht übertragen läßt.“ (ebenda)

[41] Hier liegt ein Beispiel für einen Trend bei der Rezeption von TIMSS durch die Mathematikdidaktik vor: Viele Wissenschaftler verweisen darauf, dass gerade jenes Gebiet nicht genügend im Tests auftaucht, dem sie selbst ihre Aufmerksamkeit in der Forschung bzw. der schulischen Implementation gewidmet haben.

Exkurs: Japanischer Mathematikunterricht und seine Übertragbarkeit auf Deutschland.

Die Problematik der Nichtübertragbarkeit von Unterrichtspraxis in ein anderes Land ist am Beispiel von Japan auch von anderen Autoren diskutiert worden. Die TIMSS-Videostudie (siehe Stigler u.a. 1999, Stiglers Ausgangsposition siehe Stevenson/Stigler 1992 bzw. Stigler/Hiebert 1999, zum Problem der selektiven Wahrnehmung bei TIMSS-Video Kaiser 2000, S. 184 f.) stellte in ihrer Deutung der Videos japanischen Mathematikunterricht als vorbildhaft und nachahmenswert dar. Die deutsche TIMSS-Gruppe faßt das in folgender Weise zusammen: „Japanischer Mathematikunterricht ist Problemlöseunterricht. Er schult mathematisches Verständnis und mathematisches Denken. Mathematikunterricht in Deutschland und den USA ist eher Wissenserwerbsunterricht, der auf Beherrschung von Verfahren zielt. In Deutschland werden mathematische Konzepte im Unterrichtsgespräch, das auf eine einzige Lösung hinführt, entwickelt, in den USA vom Lehrer vorgestellt und von den Schülern angewandt.“ (Baumert, Lehmann 1997, S. 215)

Die Behauptungen der TIMSS-Video-Deutung über den vorbildhaften konstruktivistischen Mathematikunterricht ignorieren viele Widersprüche: Zunächst stellen japanische Forscher ähnliche Notwendigkeiten der Veränderung von Unterrichtspraxis dar, wie man sie auch aus Deutschland kennt (z.B. Nôda 1998, Harada 1998). Dieses Darstellen von Defiziten auf internationalem Parkett ist umso bemerkenswerter, als es in Japan eine auf allen Ebenen ausgeprägte Kultur der positiven Außendarstellung im Vergleich zu kritischer Innendarstellung gibt (siehe Rohlen 1980).

Außerdem wird das japanische Schulwesen oft als starr und unterdrückend wahrgenommen, es gibt eine auffallend hohe Selbstmordrate japanischer Schüler. Die Rolle des wachsenden Leistungsdrucks hin zur high-school-Aufnahmeprüfung nach der 8.Klasse und zur Aufnahmeprüfung an den Universitäten wird in der TIMSS-Video-Deutung und der Deutung der Testresultate ignoriert. An diesen Stellen der Schullaufbahn werden Zukunftschancen verteilt, so dass die Schulen von der direkten Verteilung von Zukunftschancen durch Zensuren entlastet sind. Der äußere Leistungsdruck und das Nachmittagsschulwesen entlasten den Unterricht von der Verantwortung für Übungs- und Trainingsprozesse und verschleiern teilweise Motivierungsdefizite des Unterrichts. (Schümer 1998, 1999, Darstellung der Rolle und Arbeitsweise der Juku-Nachmittags-Schulen und Wechselwirkung mit dem öffentlichen Schulwesen bei Rohlen 1980) Der Leistungsdruck scheint sich altersmäßig immer weiter nach unten verlagert zu haben, je stärker aufgrund der wirtschaftlichen Krise der Konkurrenzdruck geworden ist. Dennoch konstatiert Keitel nach einer Japan-Visite - im Widerspruch zu TIMSS-Video: „Während die Schüler der Sekundarstufen häufig eher desinteressiert-abwesend den pausenlosen Lehrervorträgen zuhören oder lustlos-beflissen das sorgfältig geplante Tafelbild des Lehrers abschreiben, zeigen sich in den Grundschulen die Schüler lebendiger und *aufgeweckt*-interessiert, sie erhalten deutlich mehr Gelegenheit zu Eigenaktivitäten.“ (Keitel 1998) Dieser Eindruck von Unterrichtswirklichkeit steht in Einklang mit dem von Schümer (1997) herausgearbeiteten sehr hohen Redeanteil der Lehrer im japanischen Mathematikunterricht - ein Resultat, das innerhalb der euphorischen Darstellung bei TIMSS erklärungsbedürftig bleibt.

Diese Aussagen stehen offenbar in Widerspruch mit den bei TIMSS behaupteten Resultaten. Kenner japanischen Mathematikunterrichts äußern den Eindruck, dass die veröffentlichten Stunden sogenannten Meisterstunden ähneln, die vom Erziehungsministerium veröffentlicht werden, um den Lehrern „guten“ Unterricht zu zeigen. (Keitel 1998, detailliert Lewis 1998) Man hätte dann kein Abbild japanischen Unterrichts, sondern lediglich dessen, was man in Japan für guten Unterricht hält. Man kann dann natürlich fragen, inwieweit diese Wunschvorstellung den alltäglichen Unterricht durchwirkt.

Eine umfangreiche Auseinandersetzung mit methodischen Problemen rund um TIMSS leistet Kaiser (1998) Ich diskutiere hier sechs Problemkreise:

1. Sie stellt die Frage nach der Angemessenheit der zugrunde liegenden Methodologie der probabilistischen Testtheorie (Item-Response-Theory). Eine dort geleistete Kernoperation ist es, aus der Lösungshäufigkeit einer Aufgabe und der Wahrscheinlichkeit, die Aufgabe zu lösen, die Fähigkeit einer Testperson zu schätzen. Kaiser kritisiert die mit Hilfe dieses Vorgehens von der TIMSS-Gruppe getroffenen Aussagen, z.B.: „Die Abbildung illustriert, daß die Grundrechenarten in der 8.Klasse auf dem durchschnittlichen Fähigkeitsniveau von Hauptschülern mit „einiger Sicherheit“ beherrscht werden ..., Verständnis von Brüchen kann erst auf Realschulniveau erwartet werden ...“ Sie verweist darauf, dass solche Aussagen nur auf der Basis gesicherter Kenntnisse möglich sind, was unter mathematischen Fähigkeiten zu verstehen ist und dass diese Kenntnisse nicht existieren.

In meinem Artikel „Das Kompetenzstufenmodell von PISA – Eine empirische Dekonstruktion“ (Meyerhöfer 2004) erweitere ich die Kritik an Aussagen wie denen der TIMSS-Gruppe. Ich zeige dort, warum aus Lösungshäufigkeiten nicht auf Fähigkeiten geschlossen werden kann und warum Aussagen wie die zitierten inhaltlich leer sind.

2. „Darüber hinaus vertreten Baumert/Lehmann u.a. ein nicht komplex genug angelegtes Verständnis mathematischer Fähigkeiten; so verstehen sie darunter mathematisches Problemlösen, den verständigen Umgang mit mathematischen Symbolen, Begriffen und Modellen sowie mathematisches Denken in inner- und außermathematischen Kontexten. Es fehlen Fähigkeiten zur kritischen Einschätzung und Bewertung mathematischer Methoden, die angesichts einer zunehmenden Mathematisierung unserer Welt immer zentraler werden.“ (ebenda, S. 7 f.)

In dieser Arbeit wird nicht untersucht, ob das TIMSS-Leistungskonstrukt alle Dimensionen abdeckt, die aus mathematikdidaktischer Perspektive wünschenswert wären. – Die Betrachtungen von Kaiser und Weigand zeigen bereits, dass das nicht der Fall ist. In dieser Arbeit wird nur gefragt, ob mit den Aufgaben wirklich jene Dimensionen mathematischer Leistung gemessen werden, die gemessen werden sollen.

3. „Das nächste Problem ist die Kontrollierbarkeit der Ergebnisse: TIMSS hat eine Fülle hochkomplexer Ergebnisse produziert. Die öffentliche Diskussion bezieht sich in der Regel nur auf die Ranglisten. Die als nicht zufriedenstellend angesehenen Ranglistenplätze dienen in vielen Ländern – so auch Deutschland – dazu, großangelegte schulische Reformen zu begründen. ... Differenzierte Argumentationen sind in der öffentlichen Diskussion eher selten und es ist fraglich, ob sich die zweckgebundene Verwendung der TIMSS-Ergebnisse durch die Politik überhaupt verhindern läßt.“ (ebenda, S. 8) Kai-

ser übersieht, dass das Problem durch die Forscher selbst verursacht ist. Es handelt sich bei diesen internationalen Leistungstests eben nicht um wissenschaftliche Projekte zur Generierung von Erkenntnis. Es handelt sich um kulturindustrielle Produkte, die sich in einem wissenschaftlichen Raum verorten. Das wird nicht nur deutlich, wenn man die Vermarktungsstrategien betrachtet; in der Vermarktung dieser Studien wird weder das Erkenntnisinteresse der Öffentlichkeit befriedigt noch wird die Marktmacht in einem positiven Sinne genutzt, um vorhandene bzw. neugewonnene Erkenntnis zu popularisieren. Die Forscher selbst vermarkten in erster Linie das Sensationsversprechende – und eben sich selbst. Nach TIMSS lag inhaltlich ja nicht die Schlußfolgerung nahe, nun verstärkt zu testen. Diese Schlußfolgerung haben die an TIMSS beteiligten Forscher suggeriert, um ihre Marktposition zu stärken. Es lag auch nicht nahe, die Korrelationsforschung zu stärken. Es lag viel näher, die Implementation von vorhandenem Wissen zu stärken und die quantitativen Resultate qualitativ auszuloten und zu hinterfragen, TIMSS also als Kristallisationspunkt für weitere Forschung zu nutzen. Dies wurde aber durch die Veröffentlichungspolitik der TIMSS-Gruppe und durch die von der TIMSS-Gruppe suggerierte Forschungsausrichtung auf weiten Strecken verhindert. Die Ernsthaftigkeit des öffentlichen Interesses an wirklich tiefgehender Erkenntnis und an der Implementation von Alternativen wurde durch die Sensations- und Machtorientierung der Forscher in grober Weise herabgewürdigt. Das Problem des mangelnden Erkenntnisinteresses reproduziert sich im wissenschaftlichen Konstrukt selbst. Dies wurde für TIMSS im ersten Kapitel exemplarisch herausgearbeitet und wiederholt sich bei PISA.

4. Kaiser stellt im weiteren „die Frage nach dem Anregungspotential von TIMSS: Was können wir von dem Mathematikunterricht aus völlig anderen Kulturen – wie z.B. der japanischen – wirklich lernen? Sind nicht Vergleiche zum Unterricht in Ländern mit ähnlichem ökonomischen und gesellschaftlichen Hintergrund und vergleichbaren Bildungssystemen deutlich anregender?“ (ebenda) Zweifellos sind solche Vergleiche suggestiver bezüglich des Vergleichs von Bildungspraxis. Die Mehrländerstudien führen uns aber deutlicher die prinzipielle Begrenztheit der Folgerungen aus solchen Studien vor Augen (siehe Schriewer 1999). Das Potential von internationalen Vergleichen ist eben gerade ein Anregungspotential, nicht ein Folgerungspotential. Wichtiger für das Anregungspotential ist nicht die Fremdheit der Kultur, sondern die Forschungsfrage und –methode: Das Anregungspotential von TIMSS erwächst nicht aus dem Leistungstest, sondern aus den ethnografischen Case Studies und (außerhalb der methodischen Probleme) aus der Videostudie.

Dimensionalität von Fähigkeiten

5. „Des weiteren verlangt das probabilistische Testmodell die Annahme einer einheitlichen Fähigkeitsdimension, und es ist durchaus fragwürdig, inwieweit die Ergebnisse so unterschiedlicher Tests wie die zur Algebra oder zur Geometrie wirklich Rückschlüsse auf eine Leistungsdimension erlauben.“ (ebenda, S. 8)

Der Umgang mit der Frage nach der Mehrdimensionalität eines solchen Tests endet zwangsläufig in der Frage, was die Realität ist, die hier in Rede steht: Es ist unmittelbar einsichtig, dass ein mathematischer Leistungstest Fähigkeiten in mehreren Dimensionen messen muß, wenn er einigermaßen komplex ist. Es ist ebenso einsichtig, dass wir über reine intellektuelle Konstrukte sprechen, wenn wir über „Fähigkeiten“ bzw. über Fähigkeitsdimensionen sprechen. Mit diesen Konstrukten versuchen wir, die empirische Erfahrung eines „Könnens“ intellektuell zu verarbeiten. Will man so ein Konstrukt auf eine greifbare Realitätsebene herunterbrechen, so landet man bei der Feststellung, dass es sich bei einer Fähigkeit um eine bestimmte Gehirn-Körper-Konstellation handeln muß.

Es ist nun ebenfalls einsichtig, dass wir keine Aussage darüber treffen können, wie verschiedene Fähigkeiten oder Fähigkeitsdimensionen im Gehirn zusammenhängen (vom Zusammenspiel mit dem Rest des Körpers ganz zu schweigen). Wenn man neuronal denkt, scheint es nicht mal sinnvoll, Fähigkeiten und Fähigkeitsdimensionen zu unterscheiden, weil die **Neuronennetze, deren Vorhandensein und Gestalt eine Fähigkeit konstituieren**, nicht scharf voneinander getrennt sind. – Man würde sich ja eine Fähigkeit mit mehreren Dimensionen als einen Cluster von miteinander verbundenen Neuronennetzwerken vorstellen, wohingegen eine eindimensionale Fähigkeit nur *ein* solches Neuronennetzwerk wäre. Wir wissen aus der Hirnforschung, dass dieses *eine* Neuronennetzwerk nicht „frei im Raum schwebt“, sondern wiederum mit anderen Neuronen und Neuronennetzwerken verbunden ist. Das bedeutet nun aber, dass man nicht wirklich abgrenzen kann, ob eine Fähigkeit als alleinstehend und eindimensional betrachtet werden kann oder als Bestandteil einer mehrdimensionalen Fähigkeit betrachtet werden muß, die sich in einem Clusterelement konstituiert.

Die Betrachtung zeigt bereits, dass man nun nicht dem Impetus erliegen darf, das Problem in Starre oder Beliebigkeit aufzulösen: Ein starres Konzept wäre es, die verschiedenen Fähigkeitsdimensionen als unverbunden zu betrachten. Ein Konzept der Beliebigkeit wäre es, die Unterschiede zu verwischen und so zu tun, als ob nur eine einzige Dimension vorliegt. Man sieht sofort, dass auch ein starres Konzept beliebig und ein Beliebigkeitskonzept starr ist – jedenfalls im Vergleich zur Wirklichkeit. Das bleibt vom Prinzip her auch so, wenn man zu mehreren Dimensionen übergeht, denn dann reproduziert sich das Problem ja nur auf einer anderen Anzahl.

Das Problem läßt sich nur auflösen, wenn man sich der Realität in ihrer dialektischen Verbundenheit stellt. Dies ist aber nur möglich, indem man versucht, Realität zu rekonstruieren. Hier würde das bedeuten, aus Ausdrucksgestalten einer - mathematische Probleme lösenden - Praxis das **Zusammenspiel von beteiligten kognitiven Elementen zu rekonstruieren**. Natürlich bietet uns auch dieses Vorgehen weder direkten noch vollständigen Durchgriff auf die Realität, aber es läßt der Realität wenigstens die Chance, sich in unserem wissenschaftlichen Erkennen adäquat widerzuspiegeln.

Das Vorgehen beim Testen gestattet solch rekonstruktives Vorgehen nicht. Es ist uns aber immerhin möglich, uns wenigstens eine ungefähre Vorstellung davon zu erarbeiten, mit welchen Fähigkeiten man eine Aufgabe lösen kann. Dies wird der jeweils erste Bestandteil der Aufgabeninterpretationen sein. Dabei zeigt sich, dass bereits auf der Ebene der einzelnen Aufgabe verschiedene Fähigkeiten ein Lösen ermöglichen. Diese mögen zum Teil auf verschiedenen Dimensionen liegen.

Bezüglich der Dimensionalität von TIMSS haben Knoche und Lind (2000) versucht, die Frage über Faktorenanalysen zu beantworten. Sie stellen fest, dass in ihrer ersten Analyse sowohl für die einzelnen Stoffgebiete als auch im Gesamttest „stets ein starker Hauptfaktor auftritt. Trotzdem sind im Sinne der Korrelationsanalyse stets weitere Faktoren z.T. nicht unerheblich an der Erklärung der Daten be-

teiligt. In einer lernpsychologisch oder didaktisch orientierten Studie würde man diese Faktoren auf jeden Fall berücksichtigen." (Knoche, Lind 2000, S. 17) In einer zweiten Analyse, die sich innerhalb des Raschmodells bewegt, untersuchen sie fünf „leichte" und fünf „schwere" Aufgaben. Sie errechnen einen Hauptfaktor, mit dem alle Aufgaben etwa gleichstark positiv korreliert sind. Ein zweiter Faktor korreliert mit den „leichten" Aufgaben gleichstark negativ, mit den „schweren" Aufgaben gleich stark positiv. „Obwohl der Test im Sinne der Raschskalierung eindimensional ist, kann man das damit geprüfte Merkmal als *zweidimensional* im Sinne der klassischen Faktorenanalyse ansehen. Der Hauptfaktor wäre dabei als allgemeine Leistungsfähigkeit zu interpretieren. Beim zweiten Faktor würde man so etwas wie eine Schwierigkeitsdimension vermuten, da die leichten Aufgaben negativ und die schweren Aufgaben positiv mit ihm korreliert sind." (ebenda, S. 19 f.)
Diese Argumentation antwortet nur scheinbar auf die Frage von Kaiser nach der Adäquatheit der Eindimensionalität des Tests. Die Faktorenanalyse reproduziert die Defizite der Dimensionalitätsbetrachtungen der probabilistischen Testtheorie: Bei beiden Betrachtungen geht es nur darum, ob eine Aufgabe bezüglich ihrer Lösungshäufigkeit auf der gleichen „rechnerischen Dimension" wie die anderen Aufgaben liegt. Die inhaltlichen Füllungen dieser Dimensionen interessieren nicht. – Sie sind auch nur oberflächlich, holzschnittartig sowie unter Abschneiden erheblicher Teile des untersuchten Wirklichkeitsbereiches zu „beschreiben" (siehe z.B. ebenda, S. 15 ff.). Es handelt sich bei Dimensionen wiederum um die Mischung von inhaltlich beschreibbaren Dimensionen. Damit wird egal, ob man eine oder drei oder auch siebenundzwanzig solcher in sich verwaschenen Dimensionen mißt. Solange man grundsätzlich glaubt, diese Verwaschungen bei der Bearbeitung seiner Forschungsfrage in Kauf nehmen zu können oder zu müssen, ist es von untergeordneter Bedeutung, in wie viele und in welche der Dimensionen die Unschärfe hineingewaschen wird. Es ist davon auszugehen, dass nicht nur die sogenannten Metafähigkeiten, sondern auch Testfähigkeit in ihren verschiedenen Aspekten in alle Dimensionen eingewaschen werden. Ebenso ist aber davon auszugehen, dass auch die verschiedenen mathematischen Fähigkeiten sich in verschiedenen Dimensionen wiederfinden.
Eine Faktorenanalyse bewegt sich bezüglich der Frage nach der Dimensionalität der untersuchten Fähigkeit auf einer doppelt inadäquaten Ebene. Die Betrachtung auf der unmittelbar physischen Ebene zeigt, dass eine Faktorenanalyse das Erkennen der Dimensionalität von Fähigkeiten nicht unterstützen kann: Der Einfluß bestimmter „Clusterelemente" auf das Gesamt der Fähigkeit und die physische Konstitution dieses Gesamt sind nur in der unmittelbaren Untersuchung sowohl der Aufgabe als auch des Physischen zu erschließen. Man kann sie nicht „ausrechnen", indem man ein paar Lösungshäufigkeiten miteinander verrechnet. Ein Zusammenhang zwischen diesen beiden Ebenen von „Dimensionalität" ist nicht zu erkennen.
Aber auch die Ebene jener theoretischen Konstrukte, die wir uns vorstellen, wenn wir von einer Fähigkeit sprechen, wird durch eine Faktorenanalyse verfehlt: Unsere theoretischen Konstrukte entstehen daraus, dass wir erkannte Gemeinsamkeiten in unseren mathematischen und sonstigen weltlichen Konstrukten zusammendenken. Die Elemente dieses Zusammendenkens (die nur vielleicht mit den oben beschriebenen physischen Elementen zusammenlaufen) können aber nicht gegriffen werden, indem man die Lösungshäufigkeiten von Aufgaben verrechnet. Sie können gegriffen werden, indem

man zunächst die Aufgaben analysiert. Dabei zeigen sich die zur Lösung der Aufgabe erforderlichen Fähigkeitsdimensionen in ihrer dialektischen Vielheit und Verwobenheit. Die in dieser Arbeit vorgestellten objektiv-hermeneutischen Interpretationen leisten eine solche Analyse in einer ersten Annäherung. Wollte man die Dimensionalität von Fähigkeiten untersuchen, so müßte der Zusammenhang sowohl verschiedener Fähigkeitselemente als auch verschiedener Fähigkeiten näher betrachtet werden. Bereits in den Interpretationen dieser Arbeit ist aber erkennbar, dass wir es nicht mit starren Dimensionen von Fähigkeiten zu tun haben, sondern dass Spielräume vorliegen: Einerseits kann man Aufgaben auf verschiedene Weisen lösen, die mehr oder weniger miteinander verwoben sind. Aber auch innerhalb jeder dieser Vorgehensweisen bestehen vielfältige Variabilitäten. Die Vorstellung hinter der Faktorenanalyse muß all diese Spielräume verleugnen, da sie ja nur eine einzige Lösungshäufigkeit für eine Aufgabe als Datum zur Verfügung hat. Es gibt keinen schlüssigen Hinweis darauf, dass diese Defizite „sich rausrechen“. Die mit der Faktorenanalyse errechnete Dimensionalität agiert auf einer Realitätsebene, die eigener Art ist. Ihr Zusammenhang mit unseren intellektuellen Konstrukten von Leistung ist bestenfalls spekulativ, schlimmstenfalls nicht vorhanden. Selbstverständlich wird man die Ergebnisse jeder Faktorenanalyse inhaltlich deuten können: Man muß nur die Betrachtungsebene vergröbern und gleichzeitig annehmen, dass das Abschneiden inhaltlicher Aspekte durch eine große Anzahl an Items geheilt wird. Wenn man den Vergröberungs- und Abschneidungsprozeß ehrlich beurteilt, erhält man einen Eindruck davon, wieviel Chance die Realität noch hat, sich in den Ergebnissen der Analyse widerzuspiegeln.

Curriculare Validität

6. Kaiser diskutiert im weiteren das Problem der curricularen Angemessenheit des Tests angesichts der Vielzahl beteiligter Länder. In Deutschland wurden nur der Lehrplan von Nordrhein-Westfalen und die Lehrbücher zur Geometrie und Algebra des Bayrischen Schulbuchverlages für die Curriculumanalyse verwendet. Sie verweist auf Analysen, die zeigen, dass Testergebnisse umso besser sind, je intensiver die getesteten Lerninhalte unterrichtet wurden. „Diese Ergebnisse legen die Befürchtung nahe, dass das Problem der curricularen Validität für so viele unterschiedliche Länder, wie sie TIMSS berücksichtigt, gar nicht lösbar ist.“ (Kaiser 1998; S. 8)

Die curriculare Validität ist auch ein Kernthema von Keitel und Kilpatrick (1998). Sie leisten aus diesem Blickwinkel eine umfassende Kritik an internationalen Studien und konstatieren zu TIMSS:

> „Eine Fiktion war, daß die Rahmenrichtlinien und Lehrbücher (als die Repräsentanten der Curricula) „in gleicher Weise“ universell die Unterrichtspraxis beeinflussen, daß sie die wichtigste Quelle des Lehrerwissens darstellen und daß sie nicht nur das Berufswissen der Lehrer, sondern auch ihre Praxis wesentlich bestimmen.
>
> Eine zweite Fiktion war es, daß die Konzentration auf zählbare und meßbare Aspekte von Schülerleistungen die wichtigsten, hinreichend relevanten Informationen für die Datenverarbeitung sicherstellt. Diese Konzentration führte auch zu Konstrukten wie „opportunities of teaching and learning“, „performance expectations“ und „career perspectives“. Solche eher artifiziellen Konstruktionen

können Forscher mehr zu Spekulationen als zu ernsthaften Interpretationen einladen. ... Die kodierten Daten des Curriculumanalyseprojekts, die als Basis für alle weitere Datenverarbeitung und für die graphischen und numerischen Darstellungen dienten, sind nicht ernsthaft für alle Länder kontrolliert worden, noch konnte das Kodierungsverfahren von der Forschergruppe in allen Fällen selbst nachgeprüft werden, wenn die Daten schließlich vorgelegt wurden. ... Die Studie hängt ganz überwiegend von den nationalen Kodierungen ab, von den nationalen Experten und deren ... Verständnis von den Zielsetzungen der Studie, den Instrumenten und den Verfahrensweisen. Da die Kodierung eng auf die Interpretation der Konstrukte bezogen war, die Konstrukte selbst häufig mit Vorurteilen ihrer Entwickler behaftet waren, erschien in der Folge die Curriculum-Analyse im Vergleich zu den allgemeinen Ergebnissen der Haupt-TIMS-Studie nicht aussagekräftig. Die begleitenden Kontextstudien, die die empirische Leistungsstudie komplementieren und deren Ergebnisse stützen sollten, wurden irrelevant und halfen nicht, die Testdaten zu interpretieren, in einigen Fällen führte es sogar dazu, daß die Ergebnisse der Kontextstudien den Testdaten widersprachen.
Als eine dritte Fiktion der Curriculum-Analysis-Studie erwies sich, daß in Vergleichsstudien Zusammenarbeit vorausgesetzt werden kann, und notwendig entsteht damit Gleichberechtigung der Partner und nicht Wettbewerb. Aber von Anfang an war eine strenge Hierarchie zwischen den großen und den kleinen, den weit entfernten oder armen Ländern vorherrschend, und „natürlich" war meistens die (Englisch sprechende) Forschergruppe durch die großen Länder festgelegt, und diese Gruppen bestimmten im wesentlichen, wie das allgemeine Curriculum definiert wird, und die Konnotationen und Kodierungen." (Keitel/Kilpatrick 1998, S. 524 f.)

Großstudien als politische und kulturindustrielle Phänomene

Keitel und Kilpatrick setzen sich ausführlich damit auseinander, wie politische Rahmungen und die finanziellen Quellen von Großstudien die wissenschaftlichen Resultate unterminieren. Ich halte es für fruchtbarer, diese Betrachtungen unter der Begrifflichkeit von Kulturindustrie zu führen.

Ich beziehe mich dabei auf den Adorno`schen Begriff von Kulturindustrie und auf seine Erweiterung des Begriffs über den Bereich der Kunst hinaus, insbesondere in den Bereich der Wissenschaft hinein. Diese Erweiterung ist fruchtbar, weil sie uns das Phänomen der Großstudien und ihrer begrenzten Erkenntniskraft besser verstehen läßt. Der Begriff der Kulturindustrie lenkt unseren Blick auf die Entfremdung des Forschers vom Forschungsprozeß, aber auch auf das so entstehende Produkt.

Der Rückgriff auf den Begriff der Kulturindustrie anstelle der Fokussierung auf die politische Einbettung von Großstudien erweitert die zunächst eher holzschnittartige Gegenüberstellung von Geldgeberländern und „Ranking-Staffage", von armen und reichen Forschern sowie von politischem Geldgeber und wissenschaftlichem Geldnehmer (Keitel/Kilpatrick 1998). Das Problem von kulturindustriellen Großprojekten ist ja gerade, dass sie alle Beteiligten in entfremdete Produktion zwingen. Dies gilt für die beteiligten Vertreter des politischen Raums ebenso wie für die Forscher, die beteiligten Lehrer und Schüler. Bezüglich der Entfremdung unterscheiden sich die Forscher aus armen und solche aus reichen Ländern nur graduell.

Ein weiterer Vorteil der kulturindustriellen Begrifflichkeit ist, dass sie einen schärferen Blick auf die

Zerstörung von Wissenschaftlichkeit durch kulturindustrielle Projekte gestattet: Ein wesentliches Unterscheidungskriterium zwischen Wissenschaft und Kulturindustrie ist die Positionierung des Projekts im wirtschaftlichen Raum: Jeder Forscher ist darauf angewiesen, selbst Geld zu verdienen und auch den für sein Projekt notwendigen Mitarbeitern den Lebensunterhalt zu sichern. Wirtschaftliche Absicherung ist also Voraussetzung für jegliches Betreiben von Wissenschaft als Institutionalisierung von Erkennen. Zu Kulturindustrie wird Wissenschaft, wenn das Wirtschaftliche nicht mehr nur Voraussetzung von Erkennen ist, sondern dem Erkennen inhaltlich vorgelagert ist. Zugespitzt ausgedrückt: Der Wissenschaftler benötigt Gelder, weil er ein Erkenntnisinteresse hat und die Befriedigung dieses Interesses eine wirtschaftliche Basis benötigt. Der „Kulturindustrielle" benötigt ein Erkenntnisinteresse, weil er ein wirtschaftliches Interesse hat und die Befriedigung dieses Interesses ein Erkenntnisinteresse zur gesellschaftlichen Legitimation benötigt.

Die konkrete Frage an ein Projekt lautet in dieser Sichtweise immer: Was ist hier das Erkenntnisinteresse? Die **Ernsthaftigkeit dieses Interesses** offenbart sich unter anderem in der Forschungspraxis und in der öffentlichen Darstellung der Ergebnisse durch die Forscher. Wenn die Forscher selbst Rangreihenuntersuchungen an die erste Stelle ihrer Resultatspräsentationen stellen, dann offenbaren sie, dass sie an erster Stelle keinerlei Erkenntnisinteresse haben. Eine Rangreihe stellt eben keine Erkenntnis in einem wissenschaftlichen Sinne dar, sondern befriedigt für sich genommen lediglich ein voyeristisches Bedürfnis. Diesem Bedürfnis soll keine Berechtigung abgesprochen werden, aber es ist eben eine fundamentale Feststellung, dass Wissenschaftler an die erste Stelle der Präsentation einer wissenschaftlichen Untersuchung ein Resultat stellen, welches keinen wissenschaftlichen Wert hat. Man erkennt auch, dass der politische Charakter der Studie ebenfalls nicht im Vordergrund des Interesses der Forscher steht. Schließlich stellt eine Rangreihe auch keinerlei Unterstützung für politisches Handeln dar. Sie erzeugt im Gegenteil sogar Handlungsdruck, welcher wiederum entfremdetes Handeln im politischen Raum erzwingt: Ein kompetenter Bildungspolitiker erkennt ja sofort, dass diese Studien ihm keinerlei authentische Entscheidungshilfe durch neue Erkenntnis liefern kann. Er wird aber gezwungenermaßen den Eindruck erwecken, als ob er eben solche Erkenntnisse aus dem wissenschaftlichen Raum in politisches Handeln umsetzt. Der Entfremdungsdruck im politischen Raum wird verschleiert, wenn man lediglich Geldgeber und Geldnehmer in Beziehung setzt, nicht aber die kulturindustrielle Rahmung dieser Studien in Betracht zieht.

4.2. Die Konstruktion des Testinstrumentariums bei TIMSS (Population 7./8. Klasse)

Ich beschränke meine Darstellung auf jene Teile des theoretischen und technischen Konzepts von TIMSS, die sich auf die Konstruktion der Aufgaben und die Bestimmung ihres Meßinhalts beziehen. Das Gesamtkonstrukt wird bei Baumert, Lehmann u.a. (1997) zusammengefaßt, dort finden sich auch Literaturangaben zum Nachvollzug des technischen Apparats.

Zur Testentwicklung vermerkt die deutsche TIMSS-Gruppe:

> „Als erster Schritt zur Konstruktion der Leistungstests für Mathematik und die naturwissenschaftlichen Fächer wurde an der University of British Columbia in Vancouver, Kanada, eine internationale Datenbank angelegt, in der potentiell geeignete Testaufgaben systematisch gesammelt wurden.

> In die Datenbank wurden sowohl Aufgaben, die sich in anderen Untersuchungen bewährt hatten, als auch neuentwickelte Aufgaben der teilnehmenden Forschungsgruppen aufgenommen. Nach einer vorläufigen Überprüfung der curricularen Validität für die teilnehmenden Länder und einer nochmaligen Begutachtung durch Fachwissenschaftler aus 10 Ländern (Subject Matter Advisory Commitee) wurden 335 ausgewählte Aufgaben in der Regel doppelt durch zwei unabhängige Übersetzer in 30 Sprachen übertragen und im Frühjahr 1993 in einer Pilotuntersuchung eingesetzt. Um entdeckte Mängel und Lücken zu beseitigen, wurde[n durch ein Assessment-Center in den USA] ... zusätzliche, in der Regel bereits empirisch bewährte Testaufgaben mit bekannten Eigenschaften zur Verfügung [gestellt] oder neue Aufgaben entwickelt. Nach nochmaliger Überprüfung der Itemsammlung durch das Subject Matter Advisory Commitee wurde eine Vorversion der Leistungstests im Frühjahr 1994 in 43 Ländern erprobt. Schließlich wurden insgesamt 286 Aufgaben für die Hauptuntersuchung ausgewählt, von denen jeder Schüler etwa 70 Aufgaben zu bearbeiten hatte.“ (Baumert, Lehmann u.a. 1997, S. 47 f.)

Es liegt bei TIMSS also kein Konstrukt von mathematischer Leistungsfähigkeit vor, welches beschreibt, was mit diesem Test gemessen werden *soll*. Es hat demzufolge auch keine Operationalisierung eines solchen Konstrukts stattgefunden. Der Meßgegenstand wurde in einer Art Freihandverfahren festgelegt. Es existiert aber ein Katalog, der bestimmte Sachgebiete und bestimmte Anforderungsarten abdecken soll. Die Sachgebiete sind: Zahlen und Zahlenverständnis, Messen und Maßeinheiten, Algebra, Geometrie, Proportionalität, Darstellung und Analyse von Daten, Wahrscheinlichkeitsrechnung. Die Anforderungsarten sind: Wissen, Beherrschen von Routineverfahren, Beherrschung von komplexen Verfahren, Anwendungsbezogene und mathematische Probleme. (Anzahl der zugeordneten Aufgaben ebenda, S. 48) Die Kategorien werden disjunkt gedacht. Das bringt mit sich, dass beim Versuch der Einordnung einer Aufgabe in die Kategorien erhebliche Anteile des Aufgabencharakters vernachlässigt werden müssen, dass die Kategorienzuordnungen also nicht richtig sein *können*. Deshalb wird die Zuordnung in den Interpretationen nicht überprüft. Das Problem wird im PISA-Kapitel vertieft diskutiert.

Die Beschreibung der Aufgabenentwicklung durch die TIMSS-Gruppe illustriert die Probleme, die oben und in Kapitel 3 bereits dargestellt wurden: 45 Länder waren beteiligt, aber Experten aus nur 10 Ländern wählen die Aufgaben aus. Es gibt keine methodische Kontrolle darüber, was die Aufgaben messen, wie also die Geltung der Testaussage erzeugt wird. An die Stelle eines wissenschaftlichen Verfahrens der Operationalisierung und der kontrollierbaren Geltungserzeugung tritt eine Berufung auf nichtkontrollierbares Expertenwissen und auf „empirische Bewährung“, wobei unklar gelassen wird, was diese Bewährung „in der Regel“ ausmacht.

> „Um die curricularen Anforderungen der 7. und 8. Jahrgangsstufe in den Leistungstests möglichst breit abzudecken und die Untersuchungsteilnehmer nicht ungebührlich zu belasten, erhielt jeder Schüler jeweils nur eine Untermenge der insgesamt verwendeten Testaufgaben. Die Testaufgaben wurden so zusammengestellt, daß hinreichend präzise Populationsschätzungen erreicht werden konnten, auch wenn nicht jedem Schüler jede Testaufgabe vorgelegt wurde.“ (ebenda)

Es gab bei TIMSS 8 Testhefte mit jeweils unterschiedlichen Aufgaben. Um ihre statistische Vergleichbarkeit zu gewährleisten, sind sie durch sogenannte Ankeritems miteinander verzahnt. Das sind Items, die in drei oder vier verschiedenen oder in allen acht Testheften vertreten sind. Nur 6 Mathematik- und 6 Naturwissenschaftsaufgaben dienten als gemeinsamer Pool für alle acht Testhefte.

> „Die Aufgaben haben unterschiedliche Formate. Neben Aufgaben mit Mehrfachwahlantworten (multiple choice items) wurden offene Aufgabenstellungen verwendet, die den in Deutschland in der Schule üblichen Aufgaben entsprechen. Für die Bearbeitung der offenen Aufgaben stand ein Drittel der Testzeit zur Verfügung." (ebenda, S. 49 f.)

4.3. Auswahl der Aufgaben zur Interpretation, Generalisierbarkeit

Jede empirische Untersuchung steht vor dem Problem der Generalisierbarkeit ihrer Aussagen. Dies gilt auch, wenn das empirische Datum ein Test ist. In den folgenden Interpretationen werden sich verschiedene Probleme zeigen. Es wird sich zeigen, dass nicht präzise anzugeben ist, welche mathematische Fähigkeit die Aufgaben messen. Verschiedene Dimensionen von Testfähigkeit werden sich zeigen. Probleme der Zerstörung des Mathematischen werden sich zeigen. Und immer wieder kann man entgegnen, dass die jeweilige Aufgabe problematisch sein mag, dass dies aber keinen Rückschluß auf den Gesamttest zuläßt.

Gegen diese Entgegnung steht zunächst ein Strukturargument: In den Interpretationen wird eine Struktur rekonstruiert. Es ist eine Struktur, die im Testerstellungsprozeß gewirkt hat und deren Konkretion in Kapitel 3 angedeutet wurde. Es gibt keinen Grund anzunehmen, dass diese Struktur nur bei einer einzigen oder bei genau zwei oder genau drei Aufgaben gewirkt hat, bei den anderen Aufgaben hingegen nicht.

Ein weiteres Argument liefert die oben beschriebene Praxis der Testerstellung. Es liegt kein Fähigkeitskonstrukt vor und es hat keine Operationalisierung stattgefunden. Es überrascht also nicht, wenn sich auch in den Aufgaben herausstellt, dass das Gemessene beliebig bzw. nicht präzise zu benennen ist. Der Prozeß der Aufgabenerstellung unterlag keiner methodischen Kontrolle. Es überrascht also nicht, dass Verwerfungen zwischen dem manifesten und dem latenten Text auftreten. Probleme der Testfähigkeit und der Zerstörung des Mathematischen zeigen sich aber oftmals erst, wenn man auch den latenten Text analysiert. Diese Fähigkeit ist unsystematisch zwar in Expertenwissen enthalten, aber dieses Expertenwissen hat offenbar trotz des erheblichen personellen Aufwandes nicht ausgereicht, um die vielen Verwerfungen zu verhindern. Es bleibt unklar, woran das liegt. Starke Kandidaten der Erklärung sind die kulturindustrielle Einbindung, das Übersetzungsproblem, mangelnde Sensibilität für Probleme des latenten Textes in der beteiligten Forscherpopulation und prinzipielle Probleme der methodenlosen Erfassung des Latenten.

Ich habe versucht, das Problem der Generalisierbarkeit durch eine Aufgabenauswahl zu bearbeiten, welche auch Anhänger von Repräsentationsdenken überzeugen kann: Ich habe jene sechs Aufgaben ausgewählt, die jeder Schüler lösen mußte, die also den Kern der statistischen Verzahnung der Testhefte bilden. Damit sind einerseits all jene Aufgaben erfaßt, die wirklich in jedem Testheft standen. Dies sind aber auch jene Aufgaben, welche durch die TIMSS-Gruppe als besonders geeignet für diese zentrale Funktion exponiert wurden. Von diesen Aufgaben ist zu erwarten, dass sie mit besonderer

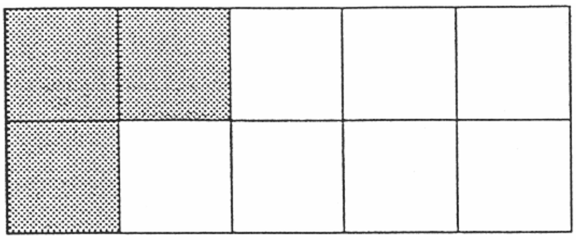
Sorgfalt ausgewählt und auf ihre Brauchbarkeit als Meßinstrument kritisch hinterfragt wurde. Diese Aufgaben dürften ein umfängliches Prozedere der Überprüfung durchlaufen haben. Es stellte sich heraus, dass von diesen sechs Aufgaben zwei Aufgaben als Testaufgaben unbrauchbar sind. In einem Fall wird vorrangig nicht der mathematische Inhalt, sondern die Fähigkeit zum Erkennen der Struktur von Schulaufgaben getestet. Im zweiten Fall testet die Aufgabe eher verbale und Kombinationsfähigkeiten und eine gewisse Unverfrorenheit im Umgang mit der mathematischen Anforderung als eine mathematische Fähigkeit. Drei weitere Aufgaben sind zumindest problematisch. Die Mechanismen der Aufgabenauswahl, -kontrolle und -bewertung bei TIMSS scheinen also grundsätzlich zur Erstellung eines Meßinstruments ungeeignet zu sein.

4.4. Aufgabeninterpretationen

4.4.1. Die TIMSS-Aufgabe A1

Betrachte die Figur. Wie viele von den kleinen Quadraten muß man ZUSÄTZLICH schattieren, damit $\frac{4}{5}$ der kleinen Quadrate schattiert sind?

A. 5
B. 4
C. 3
D. 2
E. 1



Lösungswege und Struktur des Problems

- Vorgehen vom Gegebenen aus: Das Rechteck ist in 10 Quadrate unterteilt bzw. aus zehn Quadraten zusammengesetzt. Drei der Quadrate, also ein Anteil von $\frac{3}{10}$ der Quadrate, sind bereits schattiert. Es sollen aber $\frac{4}{5}$, also **Fehler! Textmarke nicht definiert.**$\frac{8}{10}$ der Quadrate schattiert werden. Es müssen also noch 5 Quadrate bzw. der Quadrate schattiert werden.
- Vorgehen vom Gesuchten aus: $\frac{4}{5}$ der Quadrate sind acht Quadrate. Um dies zu erkennen, arbeitet man entweder mit den zehn Quadraten oder mit fünf „Doppelquadraten“. Da bereits drei Quadrate schattiert sind, muß man noch 5 weitere schattieren.
- Rückwärtsarbeiten mit Multiple Choice: Man wählt ein Lösungsangebot, z.B. 3, und schattiert dann drei Quadrate. Nun kann man sich überlegen, ob $\frac{4}{5}$ der Quadrate schattiert sind. Dann kann man entweder andere Angebote ausprobieren oder mit der Überlegung des zweiten Punktes weiterarbeiten.
- rein rechnerische Bearbeitung: wenn man erkannt hat, dass $\frac{3}{10}$ zu $\frac{4}{5}$ ergänzt werden sollen $\frac{3}{10}+\frac{x}{10}=$ **Fehler! Textmarke nicht definiert.**$\frac{4}{5}$ bzw. die Differenz von $\frac{3}{10}$ und $\frac{4}{5}$ bestimmt werden soll.

In dieser Aufgabe besteht eine geometrische Figur aus zehn gleich großen Teilfiguren. Ein Teil der

Teilfiguren ist gekennzeichnet, wobei der Blick bei Betrachtung der Figur noch nicht auf *Bruch*teile fokussiert ist. Diese eindeutige Fokussierung auf die Teile *als Bruchteile* erfolgt erst, wenn die Aufgabe gestellt wird, dass „$\frac{4}{5}$ der Quadrate“ gekennzeichnet sein sollen.

Das Ganze (das Rechteck) ist in zehn Zehntel aufgeteilt. Diese zehn Zehntel sind aber so angeordnet, dass man es auch als in fünf Fünftel oder in zwei Halbe aufgeteilt sehen kann. Die Fünftel und die Halben sind dann allerdings selbst wieder geteilt. Die naheliegendste Sichtweise ist also die des Aufgeteiltseins in Zehntel. Eine Anforderung aller Lösungswege besteht im Dechiffrieren der Graphik, also im Erkennen der Tatsache, dass hier ein Ganzes aufgeteilt ist und dass man die Aufteilung als Aufteilung in Zehntel bzw. als Aufteilung in Fünftel lesen kann. Wer graphisch ausschließlich die Aufteilung in Zehntel erkennt, muß $\frac{4}{5}$ in $\frac{8}{10}$ umwandeln können, um zu erkennen, dass die Ergänzung von $\frac{3}{10}$ zu $\frac{8}{10}$ (bzw. ihre Differenz) gesucht ist. Wer graphisch ausschließlich die Aufteilung in Fünftel erkennt, muß mit einem geteilten Bruch umgehen können, da er anderthalb Fünftel zu vier Fünftel ergänzen muß.

Als dritte Möglichkeit verbleibt ein Wechsel zwischen beiden Sichtweisen, wobei man die Figur als anschaulichen Anker hat.

Sollte man in der Sichtweise der Fünftel arbeiten, so muß zur Beantwortung der Frage noch erkannt werden, dass die Quadrate Zehntel des Ganzen darstellen und deshalb in Zehntel umgerechnet werden muß.
Die Aufgabe fordert also bei allen Lösungswegen das Erkennen von Bruchteilen aus einer Zeichnung heraus, das Umgehen mit einem Bruch als Darstellung für einen Bruchteil und die Ergänzung des einen Bruchteils zum anderen bzw. die Bildung einer Differenz zwischen beiden.

Interpretation des Aufgabentextes

Betrachte die Figur.
Die Formulierung schafft eine Fokussierung auf eine Figur und damit eine kurze, klare, zeitsparende Hinführung auf den Gegenstand der Aufgabe. Die Aufmerksamkeitsbindung verbindet sich mit der Vorbereitung bzw. Begleitung einer sich auf den Fokussierungsgegenstand beziehenden Frage bzw. Mitteilung. Immer stellt diese Aufforderung eine Instruktions- bzw. Prüfungssituation und eine Asymmetrie zwischen Sprechendem und Angesprochenem her. (Das verdeutlicht sich bei der kontrastiven Konstruktion von symmetrischen oder umgekehrt asymmetrischen Situationen - ein Untergebener könnte nicht wohlgeformt *„Betrachten Sie die Figur.“* zu seinem Vorgesetzten sagen.) Im Zusammengehen mit dem Bild betont „die Figur“ - im Vergleich etwa zu „das Bild“ oder „die Zeichnung“ - den mathematischen Kontext, ohne eine nähere Fokussierung auf einen Aspekt der Figur vorzunehmen wie z.B. durch „das Rechteck“.

Wie viele von den kleinen Quadraten [...]
Die bildungssprachliche Verwendung der Genitivkonstruktion *Wie viele der kleinen Quadrate* wird vermieden. Dies kann man äußerlich auf einen Adressaten- bzw. Autorenkreis außerhalb von Bildungsschichten beziehen, für den der Genitiv etwas Fremdes, Irritierendes und Aufgesetztes hat. La-

tent ist ***von den** Quadraten* dynamischer als *der Quadrate*, legt also nahe, mit einem Teil der Quadrate etwas zu machen. *Von den* kleinen Quadraten werden einige sprachlich separiert, also herausgehoben. Die latente sprachliche Unterstützung des Lösungsvorgehens wird also bezahlt mit sprachlicher Vergröberung des Textes. (Man könnte sich an dieser Stelle mit der Übersetzung des amerikanischen *How many **of the*** ... ins Deutsche auseinandersetzen. Es geht aber in der Interpretation der Aufgabe nicht um die Entstehung von Konstrukten, sondern um ihren Charakter als Meßinstrument, also um das Resultat. In diesem Gedanken steckt auch die Annahme sprachlicher Autonomie der deutschen Arbeitsgruppe.)
Es wird besonders auf die *kleinen* Quadrate verwiesen. Dieser Verweis hat einen vereindeutigenden Aspekt, denn es wird ausgeschlossen, dass sich der Schüler mit den aus den kleinen Quadraten zusammengesetzten „großen" Quadraten auseinandersetzt. Es würde Zeit kosten, bis dem Schüler völlig klar wäre, dass es um die „großen" Quadrate nicht gehen kann, weil die Aufgabenstellung für sie sinnlos wäre. Der Verweis hat aber auch einen verwirrenden Aspekt, denn die Wahrscheinlichkeit, daß sich Schüler von sich aus mit den „großen" Quadraten auseinandersetzen, ist ausgesprochen gering. Man wird also hier darauf gestoßen, besondere kleine und eventuell sogar große Quadrate zu suchen. Lediglich als Hilfe für den Schüler, der nicht weiß, was Quadrate sind, könnte man sich „kleine Quadrate" vorstellen. Dieses Argument zerbricht aber daran, daß außer den Quadraten gar nichts da ist, womit man arbeiten kann. Die Vereindeutigung zerstört sich also selbst. An dieser Stelle erhalten testerfahrene Schüler einen Vorteil, weil sie mit solchen testvereindeutigenden und beschleunigenden Konstruktionen vertraut sind. Sie werden einerseits nicht irritiert, andererseits wird für sie die Bearbeitung der Aufgabe wirklich beschleunigt.

Um den Charakter der nun folgenden Sinneinheit herauszuarbeiten betrachte ich Möglichkeiten, die Sinneinheit *Wie viele von den kleinen Quadraten [...]* wohlgeformt fortzuführen[42]. Ich ordne die Möglichkeiten nach ihrer Komplexität und probiere verschiedene Bezeichnungen für das Kennzeichnen der Teile aus:

... sind eingefärbt? (Mengenerfassung)
... sind nicht gefärbt? (Mengenerfassung)
... müssen noch schraffiert werden, damit 5 Quadrate schraffiert sind? (Summen- bzw. Differenzbildung)
... müssen geschwärzt werden, damit 3 Quadrate weiß sind/bleiben? (Doppelte Differenz- bzw. Summenbildung)
... müssen noch ausgemalt werden, damit $\frac{4}{5}$ der Quadrate ausgemalt sind? (Differenz- bzw. Summenbildung mit zusätzlicher Schwierigkeitskonstruktion)
... müssen noch eingeschwärzt werden, damit eine quadratische Anzahl an Quadraten eingeschwärzt ist? (Differenz- bzw. Summenbildung mit komplexerer zusätzlicher Schwierigkeitskonstruktion, 2 Lösungen)

[42] Sinn einer solchen Generierung der Möglichkeiten der wohlgeformten Fortführung des Textes ist in der objektiven Hermeneutik das Auffinden der Besonderheit der dann wirklich gegebenen Aufgabenstellung, indem man die bestehenden Möglichkeiten dem vorgefundenen Text gegenüberstellt.

[...] muß man ZUSÄTZLICH schattieren, [...]

Die Tester haben eine komplexere Aufgabenversion gewählt. In dieser Sinneinheit wird zunächst ein „mittleres Maß“ an persönlicher Ansprache des Schülers gewählt: Es heißt nicht „sind zusätzlich zu schattieren“ oder „müssen zusätzlich schattiert werden“, was eine völlige Adressatenunbezogenheit darstellen würde. Es heißt auch nicht „mußt du zusätzlich schattieren“, dies wäre ein besonderer Bezug auf den Schüler persönlich. Die gewählte „mittlere“ Variante des „man“ schließt den Schüler in eine unbestimmte Gruppe mit ein, nimmt somit die Testsituation direkt auf. Es ist aber ein persönliches Handeln, nämlich durch *man*, notwendig. Im Vergleich zu *sollte man* wird eine größere Verbindlichkeit der Frage geschaffen. Auch die konjunktiven Varianten *müßte man*, *müßtest du* oder *wären zu* würden ein direktes Durchführen des Schattierens weiter wegschieben. Im Vergleich dazu wird der Schüler hier geradezu aufgefordert, wirklich zu schattieren.

ZUSÄTZLICH ist nicht nur stärker als das von mir vorgeschlagene „noch“, es ist auch noch groß geschrieben. Dies könnte man zunächst als Hilfe interpretieren, da es vor der Angabe der insgesamt zu schattierenden Quadrate warnt. Ein Blick auf die durch Multiple Choice angegebenen Lösungsvarianten zeigt aber, dass diese Hilfe nicht notwendig wäre; der Schüler würde spätestens hier seinen Irrtum bemerken. Es handelt sich bei der Großschreibung also um einen weiteren Beschleuniger für das Lesen und Bearbeiten der Aufgabe. Auch an dieser Stelle erfährt ein testerfahrener Schüler einen Vorteil, weil er mit einer solchen testbeschleunigenden Konstruktion vertraut ist. Der testunerfahrene Schüler wird eher irritiert sein, weil in der Schriftsprache normaler Texte, insbesondere bei schulischen Texten, Wörter in Großbuchstaben eine derart starke Exponierung erzeugen, dass ein Nachdenken über den Grund der Exponierung angezeigt ist. Sucht man nach Beispielen einer adäquaten Anwendung von Großbuchstaben, so trifft man auf HILFE oder SPINNST DU oder WIE OFT SOLL ICH ES NOCH SAGEN!? Testfähigkeit bedeutet hier, den Grund der Exponierung bereits zu kennen: Vermeidung naheliegender Fehler. Der Fehlervermeidungs- bzw. Beschleunigungsvorteil durch diese Exponierung gilt natürlich nur für jenen, der der latenten Aufforderung widersteht, über den Grund der Exponierung nachzudenken.

Auch die Wahl des Verbs „schattieren“[43] ist zweischneidig. Einerseits ist es eine klare Bezeichnung: Verbindungen mit „färben“ könnten zu Diskussionen bzw. gedanklichen Irritationen darüber führen, dass hier nichts gefärbt wird, weil Schwarz oftmals gar nicht als Farbe angesehen wird - man sieht bereits, welchem spitzfindigen und besserwisserischen Einwand hier Vorschub geleistet wird. „Ausmalen“ ist kindlich und verweist auf wenig strenges Vorgehen. „Schwärzen“ ist ein in der Schülersprache eher ungewöhnliches Verb und ist mit Verdeckung und Verschleierung konnotiert. Andererseits ist „schattieren“ ein im Deutschen so selten gebrauchtes Wort, dass es einem bei der Generierung von Möglichkeiten der Weiterführung der Aufgabe wahrscheinlich gar nicht in den Sinn kommt, wenn man die Aufgabenstellung nicht schon gelesen hat.[44] Auch hier kann man von einem Irritationspotential ausge-

[43] Wahrscheinlich übersetzt aus dem englischen shade = dunkel tönen, schraffieren, schattieren, ausmalen; im Englischen eigentlich gemeint im Sinne von „darken with parallel lines“ bzw. Schattieren von Flächen zur Erzeugung eines räumlichen Eindrucks beim Zeichnen.

[44] Im Deutschen benutzt man das Wort „schattieren“ für das Abtönen einer Farbe bzw. im englischen Sinn zur Erzeugung eines räumlichen Eindrucks. In unserem Zusammenhang läßt es außer einem Übersetzungsproblem keinerlei Interpretation zu.

hen, wenn ein so selten genutztes Wort verwendet wird. Wiederum werden testgeschulte Schüler bevorteilt, die gewohnt sind, solche sprachlichen Irritationen nicht zuzulassen. Anders als im Falle der scheinbaren Vereindeutigungen „kleine Quadrate“ und ZUSÄTZLICH ist hier aber eine Alternative abzusehen: Die Verwendung von „gefärbt“ oder von „schraffiert“, wobei die Quadrate in der Zeichnung wirklich schraffiert werden.

[...], damit $\frac{4}{5}$ *der kleinen Quadraten schattiert sind?*

Hier ist die Bruchzahldarstellung und nicht die verbale Darstellung „vier Fünftel“ gewählt. Dies bedeutet eine Vereinfachung der Aufgabe und beschleunigt die Lösung, weil der Übersetzungsschritt aus der verbalen in die Bruchzahldarstellung entfällt (außerdem muß die Notwendigkeit gerade dieser Übersetzung nicht erkannt werden) und weil der Schüler auf den Lösungsweg geleitet wird: Die Betrachtung der Quadrate als Bruchteile und speziell die Betrachtung in Fünfteln. Hier liegt also der Fall vor, dass der Lösungsgedanke bereits in die Aufgabenstellung eingebettet ist.
Der schon aufgezeigten Logik der schnellen Bearbeitbarkeit folgt nun die wiederholte Nutzung der *kleinen Quadrate* und des *schattieren.* Es wird in dieser Konstruktion nicht gefordert, dass $\frac{4}{5}$ *der Gesamtfläche schattiert ist*. Damit wird einerseits die Schwierigkeit der Bestimmung des Ganzen, der Bezugsmenge, vermieden. Andererseits wird die Aufgabe dadurch weiter vereinfacht, weil das Problem bereits „kardinalisiert“ ist.

Multiple Choice

Multiple Choice bietet hier die Möglichkeit des Rückwärtsarbeitens. Man kann die angebotenen Werte einsetzen und ausprobieren, ob dann $\frac{4}{5}$ der Quadrate schattiert sind. Dazu ist aber notwendig zu erkennen, dass das acht Quadrate sind. Die Anforderung an das Umgehen mit Bruchzahlen bleibt also bestehen, es wird lediglich statt einer Differenz- eine Summenbildung verlangt. Aufgrund der kleinen Zahlen ist auch beim Vorwärtsarbeiten eine Summen- statt einer Differenzbildung möglich.
Eine Zuordnung von bestimmten fehlerhaften Antwortmöglichkeiten zu bestimmten Fehlvorstellungen ist wenig ertragreich. Es liegt lediglich nahe, dass der Schüler, der nur noch *ein* Quadrat schraffieren will, mit vier Quadraten statt mit $\frac{4}{5}$ der Quadrate arbeitet.

Zusammenfassung

Die Aufgabe fordert das Erkennen von Bruchteilen aus einer Zeichnung heraus, das Umgehen mit einem Bruch als Darstellung für einen Bruchteil und die Ergänzung des einen Bruchteils zum anderen bzw. die Bildung einer Differenz zwischen beiden. Aus dem im Bild eingeführten Problem heraus wird durch die „zusätzlich“-Konstruktion eine komplexere Aufgabenversion entwickelt, die sich aber nicht ins unsinnig Komplexe versteigt. In der Aufgabenstellung ist der Lösungsgedanke bereits eingebettet.
An mehreren Stellen wird das Erfassen und die Lösung der Aufgabe durch die Aufgabenstellung dy-

namisiert und beschleunigt sowie scheinbar vereindeutigt. Es wird auf eine bildungsbürgerlich verschwierigende Sprache (Genitiv) verzichtet, Fachsprache spielt keine Rolle.
An mehreren Stellen wird der Schüler durch die Aufgabenstellung irritiert. An den Textstellen „kleine Quadrate" und „ZUSÄTZLICH" sind diese Irritationen einer Art „Testerhabitus" geschuldet: Die irritierenden Textstellen dienen offenbar dazu, den Text eindeutig zu machen. Der Text ist aber bereits eindeutig. Diese Diskrepanz zwischen manifester Textbedeutung und objektivem Inhalt (Der Text tut manifest so, als ob etwas eindeutig gemacht werden müsse, er ist aber bereits eindeutig.) ist nur über den Wunsch zur Lösungsbeschleunigung erklärbar.
Die Irritation des „schattieren" ist durch die Nutzung von „schraffieren" vermeidbar. Diese Textstelle ist ausschließlich als Übersetzungsproblem interpretierbar. Mit allen drei Irritationen ergibt sich allerdings eine Bevorzugung von testgeschulten Schülern, denn diese sind darin geschult, solche Irritationen nicht zuzulassen. So wird auch nur für sie die Beschleunigung beim Erfassen der Aufgabenstellung wirksam. Hier wird also Testfähigkeit statt mathematischer Fähigkeit gemessen.
Multiple Choice bietet hier die Möglichkeit des Rückwärtsarbeitens. Man muß aber die mathematischen Anforderungen, die die Aufgabe stellt, auch im Falle des Rückwärtsarbeitens erfüllen.

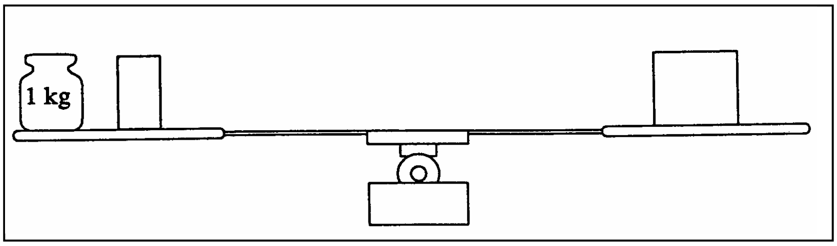
Alternative Aufgabenstellung

Betrachte die Figur. Wie viele von den Quadraten muß man noch schraffieren, damit $\frac{4}{5}$ der Quadrate schraffiert sind? (In der Zeichnung sollen die Quadrate schraffiert sein.)

A. 5 B. 4 C. 3 D. 2 E. 1

4.4.2. Die TIMSS-Aufgabe A2

Die Gegenstände auf der Waage halten sich im Gleichgewicht. Auf der linken Waagschale befinden sich ein Gewicht (eine Masse) von 1 kg und ein halber Ziegelstein. Auf der rechten Seite befindet sich ein ganzer Ziegelstein.



Welches Gewicht (welche Masse) hat ein ganzer Ziegelstein?

A. 0,5 kg
B. 1 kg
C. 2 kg
D. 3 kg

Lösungswege und Struktur des Problems

- argumentatives Vorgehen 1: Wenn sich die Gegenstände auf der Waage im Gleichgewicht halten, dann befindet sich auf beiden Seiten die gleiche Masse. Wenn ein Kilogramm **und** ein halber Ziegelstein die gleiche Masse haben wie ein ganzer, also zwei halbe Ziegelsteine, hat ein halber Ziegelstein eine Masse von einem Kilogramm. Ein ganzer Ziegelstein hat dann eine Masse von 2 kg.

- argumentatives Vorgehen 2: Wenn sich die Gegenstände auf der Waage im Gleichgewicht halten, dann befindet sich auf beiden Seiten die gleiche Masse. Ein halber Ziegelstein wiegt so viel wie ein halber Ziegelstein. Die zweite Hälfte des ganzen Ziegelsteins muß also so viel wiegen wie der 1 kg schwere Körper. Wenn ein halber Ziegelstein also 1 kg wiegt, wiegt ein ganzer Ziegelstein 2 kg.

- rechnerisches Vorgehen: Ich nehme die Waage als Sinnbild für eine Gleichung auf:

$1\text{ kg} + x = 2x$, also $x = 1\text{ kg}$ Hier ist die Masse eines Ziegelsteins 2x.

bzw. $1\text{ kg} + \frac{1}{2}x = x$, also $1\text{ kg} = \frac{1}{2}x$, also $x = 2\text{ kg}$ Hier ist die Masse eines Ziegelsteins x.

Die Aufgabe erfordert zunächst ein Umgehen mit dem Problem des Gleichgewichts der Waage. Eine Waage befindet sich im Gleichgewicht, wenn die Masse und somit unter günstigen Umständen (gleiche Beschleunigungswirkung auf beide Körper, die man hier annehmen kann) das Gewicht der Körper in beiden Waagschalen gleich ist. Diese Information der Gleichheit muß der Schüler aus der Aufgabenstellung bzw. aus dem Bild ziehen.

Bei argumentativem Vorgehen muß der Schüler nun die gegebenen Gewichtsverhältnisse verarbeiten, indem er die Massegleichheit eines halben Ziegelsteins und des 1 kg-Körpers erkennt.

Bei rechnerischem Vorgehen gilt es, die Situation passend zu modellieren, die aufgestellte Gleichung zu lösen und das Ergebnis richtig zu interpretieren. Hier zeigt sich besonders deutlich die Gefahr, dass man die Masse des halben Ziegelsteins mit einem Kilogramm richtig bestimmt, aber nicht beachtet, dass man die Masse eines ganzen Ziegelsteins bestimmen soll. Die Gefahr ist dadurch verstärkt, dass im Bild der halbe Ziegelstein in der Seitenansicht wie ein gebräuchlicher ganzer Ziegelstein aussieht und man einen im Querschnitt quadratischen Ziegelstein im allgemeinen nicht wie im Bild längs, sondern quer halbiert.

Die bisher benannte Struktur des Problems benutzt die implizite Annahme, dass das Problem als reales Problem nicht ernst genommen werden darf, sondern innerhalb der Logik von mathematischen Schulaufgaben der Mittelstufe zu deuten ist. Wenn man das Problem als reales löst, ergibt sich nämlich ein anderes Resultat: Dann stehen auf der linken Waagschale ein Körper mit einer Masse von 1 kg 5,9 cm vom Waagenmittelpunkt entfernt und ein Körper mit einer Masse von x kg 4,5 cm vom Waagenmittelpunkt entfernt. Auf der rechten Seite steht ein Körper mit einer Masse von 2x kg 4,5 cm vom Waagenmittelpunkt entfernt. Dann ergibt sich mit dem Schwerpunkterhaltungssatz oder mit dem Hebelgesetz oder mit einer Drehmomentbetrachtung oder mit einer Rechnung, bei der die Masse mittels der Entfernung gewichtet wird, eine Masse 2x des Ziegelsteins von 2,62 kg. Für diesen Wert ist kein Multiple-Choice-Angebot vorhanden, wenn man rundet, muß man 3 kg wählen. Hier liegt also ein Fall vor, in dem ein Schüler zum im Sinne der Tester falschen Resultat gelangen würde, obwohl er ein anspruchsvolles Problem löst und wahrscheinlich auch das kann, was die Tester zu messen glauben. Man könnte vereinfacht sagen: Ein Schüler, der „zu klug“ ist, gelangt zum falschen Resultat. Defizitär formuliert: Dieser Schüler erkennt nicht, auf welcher Ebene er hier argumentieren soll. Er verschwendet Zeit und gelangt eventuell sogar zu einem „falschen“ Resultat. Es fehlt ihm an Testfähigkeit.

Interpretation des Aufgabentextes

Bei dieser Aufgabe muß auch das Bild interpretiert werden, da die Aufgabe sich auf das Bild bezieht. Das Bild zeigt ein Gebilde, welches aus mehreren symmetrisch angeordneten Teilen besteht - lediglich zwei verschieden große Rechtecke und eine Figur, der „1 kg“ einbeschrieben ist, verletzen die Symmetrie. Ich möchte die Interpretation nicht ins Skurrile gleiten lassen, indem ich mögliche Deutungen dieses Gebildes herausarbeite, aber das Gebilde ist ohne die textliche Erläuterung nicht sonderlich gut als Waage zu erkennen. Erkennt man das Gebilde als Waage, so sind ihre Waagschalen im Vergleich zu den Waaghebeln extrem ausgedehnt. Man findet solche Waagen selten, da mit dieser Konstruktion eine große Abhängigkeit der Wiegegenauigkeit vom Standort der Gegenstände verbunden ist. Die Waage befindet sich nicht mehr unbedingt im Gleichgewicht, wenn die Masse der Körper in beiden Waagschalen gleich ist, sondern nur, wenn gleichzeitig der Schwerpunkt der Körper auch den gleichen Abstand vom Drehpunkt der Waage hat. Umgekehrt kann sich die Waage im Gleichgewicht befindet, wenn die Masse der Körper ungleich, aber die Beträge der Drehmomente gleich ist, man kann verschieden schwere Körper also so lange verschieben, bis die Waage im Gleichgewicht ist. Im Falle der Verteilung der Körper wie im Bild würde die Waage sich bei einer Ziegelsteinmasse von 2 kg zum Beispiel nach links neigen.

Der Schüler soll hier offenbar keine komplizierteren Berechnungen anstellen, und die durch Multiple

Choice angebotenen Lösungen schließen eine Berücksichtigung dieser Betrachtungen fast aus. Der Schüler, der eine gewichtete bzw. Hebelwirkungs- bzw. Drehmomentbetrachtung anstellt, erfährt also auf dem Wege über Multiple Choice von der Irrelevanz seiner Überlegungen für die vorliegende Aufgabe - es sei denn, seine Testfähigkeit ist so gering entwickelt, dass er zu den „falschen“ 3 kg aufrundet. In jedem Fall liegt durch das Bild ein Irritationsmoment vor; der Schüler mit Testfähigkeit, der exakte Überlegungen von vornherein ausspart, erfährt einen Vorteil. Das Bild erweckt den äußeren Anschein, dass hier ein physikalisches bzw. ein alltägliches Meßproblem - in jedem Fall ein Problem mit einer technischen Konnotation - zu betrachten ist. Durch die Angebote von Multiple Choice wird dies dementiert. Es wird sich zeigen, dass auch der Text diese Orientierung aufbaut und dementiert:

Die Gegenstände auf der Waage halten sich im Gleichgewicht.

Es ist von Gegenständen die Rede, nicht von Körpern. Das verweist auf einen alltagssprachlichen, nicht physikalischen Charakter der Aufgabe.

Sie „halten sich“ im Gleichgewicht, sie „sind“ oder „befinden sich“ nicht darin. Sie haben also eine aktive Rolle und rücken damit in den Mittelpunkt der Aufmerksamkeit. Die Gegenstände halten auch nicht die Waage im Gleichgewicht, sondern „sich“. Dies verstärkt die latente Orientierung auf die Gegenstände, also weg von der Waage und somit vom äußerlich physikalischen Kontext.

Gleichgewicht ist ein Wort mit physikalisch-fachsprachlichem und alltagssprachlichem Charakter. Der (durch das vorliegende Problem naheliegende manifest) fachsprachliche Charakter wird hier nicht bedient, denn dann müßten die Körper *im Gleichgewicht sein* oder es müßte *Gleichgewicht vorliegen* - die Summe der angreifenden Kräfte wäre dann Null. Der nahezu personifizierende Charakter von *sich im Gleichwicht halten* (man verdeutliche sich, wer oder was *sich* wohlgeformt *im Gleichgewicht hält* - es sind Menschen, Charaktere und Prinzipien) verweist auf den hier alltagssprachlichen Charakter von *Gleichgewicht*, konnotiert mit Ausgeglichenheit, Ausgewogenheit, dynamischem Miteinander. Außerdem konstruiert diese Formulierung die Gegenstände zu einem Ganzen, zu einer Gesamtheit. Da aber niemals Körper *sich* wohlgeformt *im Gleichgewicht halten*, liegt ein eigentümlicher Widerspruch zum manifest fachsprachlichen Charakter von *Gleichgewicht* vor.

Auf der linken Waagschale befinden sich [...]

Hier wird ein artifizieller Kontext konstruiert: Man findet im „wirklichen“ Leben diese Formulierung nicht. Das zeigt sich, wenn man nach Geschichten sucht, in denen diese Formulierung auftaucht. Will man zweischalige Waagen im Zusammenhang mit einem Wiegevorgang betrachten, so benutzt man nie diese Formulierung, weil sie zwangläufig die Beschreibung des Inhalts der rechten Waagschale nach sich zieht, und bei einem echten Wiegevorgang liegen da immer die Wägestücke, noch dazu ist egal, auf welcher der Waagschalen sie liegen. Selbst wenn man sich eine Novelle vorstellt, in der die Formulierung vorkommt, kann nur eine Zweckentfremdung der Waage oder wenigstens ein außergewöhnlicher Wiegevorgang (z.B. Aufwiegen eines Menschen in Gold) beschrieben werden. Die artifizielle Konstruktion des Problems findet hier ein inhaltliches Pendant.

[...] ein Gewicht (eine Masse) von 1 kg und [...]

Bis hierher haben Bild und Text widersprüchliche Signale bezüglich der Frage gegeben, ob die Aufgabe unter mathematischen, physikalischen oder alltäglichen Gesichtspunkten zu bearbeiten sei. Den äußerlichen auf Physik bzw. Meßtechnik orientierenden Signalen wird latent widersprochen.
An dieser Stelle nun tritt die Unsicherheit in offene Konfusion über. Die Unklarheit wird im Text manifest. Um dies deutlich herauszuarbeiten, diskutiere ich verschiedene regelgemäße Möglichkeiten der Beschreibung der Situation „Auf der linken Waagschale...“, zunächst innerhalb physikalischer Fachsprache:
- „... befinden sich ein Körper mit einer Masse von 1 kg und ein halber Ziegelstein.“ Dies wäre die allgemein-exakte Wortwahl eines Physikers. (Alternativ könnte auch von einem Körper mit einem Gewicht von x Newton die Rede sein, was die Abkürzung dafür ist, dass zwischen dem Körper und der Erde eine Gewichtskraft von x Newton wirkt.) Für ihn würde auch „... befindet sich eine Masse von 1 kg und ...“ genügen, aber selbst der Physiker würde dies als weniger wohlgeformt ansehen.
- „... befinden sich ein Wägestück mit einer Masse von 1 kg und ein halber Ziegelstein.“ Dies wäre die nicht mehr so allgemeine Wortwahl. Der Konkretisierungsschritt hin zum Wägestück bezieht aber das Bild ein und entspricht einem Anspruch größerer Anschaulichkeit.
- „... befinden sich ein Gewichtsstück mit einer Masse von 1 kg und ein halber Ziegelstein.“ Hier wird die Ebene physikalischer und physikdidaktischer Regelhaftigkeit bereits verlassen. Den Alltagsausdruck „Gewichtsstück“ gibt es in diesen Disziplinen nicht mehr, da - kurz gesagt - Gewicht eine Bezeichnung für Gewichtskraft ist und eine Kraft nicht stückchenweise auf die Waage gepackt werden kann. Wir befinden uns hier also bereits auf dem Gebiet der Alltagsprache. Die Bezugnahme auf die „Masse von 1 kg“ würde diesen Exkurs aber entschärfen, weil das Problem mit der (Gewichts-)Kraft anerkanntermaßen ein schwerverständliches ist und der physikalischen Exaktheit wenigstens im Umgang mit der Masse genüge getan ist. Dieser Fall bewegt sich also auf der Grenze der Regelhaftigkeit physikalischer Fachsprache.

Welche regelhaften Weiterführungen von „Auf der linken Seite ...“ wären nun in der Alltagssprache denkbar?
- „... befinden sich ein Körper mit einem Gewicht von 1 kg und ein halber Ziegelstein.“ Die Verwendung von Masse statt Gewicht wäre schon eine Regelabweichung, da es in der Alltagssprache für die Bezeichnung von Schwere kaum benutzt wird, eher für die Bezeichnung einer großen Menge oder von Trägheit.
- „... befinden sich ein Gewicht von 1 kg und ein halber Ziegelstein.“ Auch an dieser Stelle wäre die Verwendung des Wortes Masse eine Regelabweichung.

Schauen wir nun auf den Aufgabentext: *Auf der linken Seite befinden sich ein Gewicht (eine Masse) von 1 kg und ein halber Ziegelstein.*
Die Konfusion, die man schon beim oberflächlichen Lesen dieser Konstruktion erkennt, ist mit den eben diskutierten Kontrastierungen erklärbar. Der Text bewegt sich, solange die Klammer fehlt, wohlgeformt in der Alltagssprache und verweist damit klar darauf, dass hier kein physikalischer Kontext vorliegt. Die Klammer dementiert das, indem sie - sprachlich wenig wohlgeformt - doch noch jene physikalische Größe einbringt, die die Schwere beschreibt. Physikalisch gesehen ist das überflüssig,

da bei einer Balkenwaage Massengleichheit ohnehin über Gewichtsgleichheit festgestellt wird. (Das Wort „Gleichgewicht" verweist darauf, dass die Summe der angreifenden Kräfte bzw. Drehmomente Null ist.) Nur in der Verwendung der Einheit Kilogramm für das Gewicht widersprechen sich Alltags- und Fachsprache ernstlich. Es wird also lediglich ein schulmeisterliches Bedürfnis nach korrektem Gebrauch von Fachsprache bedient. Die Struktur kann man als äußerliche Verfachsprachlichung bei gleichzeitiger inhaltlicher Dementierung von Fachlichkeit bezeichnen. Gleichzeitig wird ein Irritationsmoment geschaffen, denn der Schüler muß einen Umgang mit der offenen sprachlichen Verwerfung finden. Konkret muß er entscheiden, ob die begriffliche Doppelung für die Lösung wichtig ist oder nicht. Ein Schüler, der das Schulmeisterliche des Textes erfassen und übergehen kann, erhält hier einen Zeitvorteil.

Auf der rechten Seite befindet sich ein ganzer Ziegelstein.

Die beschriebene Logik des Artifiziellen, Besonderen wird hier „vollendet". Der im Kern mathematische Probleminhalt ist eingeführt: Ein Kilogramm und ein halber Ziegelstein sind gleichzusetzen mit einem ganzen Ziegelstein. Der physikalische Teil des Probleminhalts bleibt implizit.

Der „ganze Ziegelstein" auf dem Bild hat einen quadratischen Querschnitt. Der „halbe Ziegelstein" hingegen hat den Querschnitt eines klassischen ganzen Ziegelsteins. Die Ziegelsteine mit quadratischem Querschnitt - zumindest diejenigen, die die Größe eines 1kg-Massestücks haben - werden üblicherweise quer und nicht längs halbiert und lassen sich auch nur sehr schwer sauber halbieren. Dies ist nicht nur ein Moment der Irritation, sondern ein echter Stolperstein, denn das Denken muß hier quer zum Gewohnten arbeiten, ohne dass dafür aus dem Problem heraus eine Notwendigkeit besteht. Diese artifizielle Konstruktion bewegt sich aber auch sehr klar weg vom Realen. Dies ermöglicht einerseits in seiner Abstraktion ein Absehen von Aspekten des Realen. Das verweist andererseits auf einen hier transportierten Charakter von Mathematik: Diese Mathematik erlaubt uns, uns nicht um das Halbieren von Ziegelsteinen und um die Konstruktion von Waagen zu kümmern. Diese Mathematik nimmt aber gleichzeitig das Problem nicht ernst, dem sie sich zuzuwenden behauptet.

Welches Gewicht (welche Masse) hat ein ganzer Ziegelstein?

Die Konfusion durch die Einbringung der physikalischen Größe Masse in die alltagssprachliche Verwendung des Wortes Gewicht wiederholt sich hier. Die Verwendung von Multiple Choice unterstreicht die Absurdität dieser Konstruktion, da auch die Lösungen in Kilogramm angegeben sind, so dass der Alltagsgebrauch des Wortes Gewicht nochmals unterstrichen wird.

Multiple Choice

Die Multiple-Choice-Konstruktion erlaubt bei dieser Aufgabe das Rückwärtsarbeiten. Dazu muß der Schüler genau wie beim Vorwärtsarbeiten die Gleichheit der Gewichte auf beiden Seiten erkennen. Er muß im Gegensatz zum vorwärts arbeitenden Schüler nicht die Gewichtsgleichheit des Kilogramm-Körpers mit einem halben Ziegelstein erkennen, sondern kann die jeweiligen Gewichte auf beiden Seiten aufsummieren. Dabei hat er so wie der vorwärts arbeitende Schüler die Möglichkeit, zunächst mit dem halben oder mit dem ganzen Ziegelstein zu arbeiten. Auch für den rückwärts arbeitenden Schüler ergibt sich die Gefahr, einen halben Ziegelstein schon für einen ganzen zu halten.

Der rückwärts arbeitende Schüler muß also alle Klippen nehmen, auf die auch der vorwärts arbeitende Schüler trifft, verwendet aber ein anderes Rechenverfahren.
Kombinationen von Vorwärts- und Rückwärtsarbeiten sind möglich, bringen aber keinen zeitlichen Vorteil. Sie ermöglichen demjenigen, der große Schwierigkeiten hat, die Aufgabe zu strukturieren, eine Lösungschance.

Zusammenfassung

Die Aufgabe erfordert zunächst ein Umgehen mit dem Problem des Gleichgewichts der Waage. Die Information der Gleichheit der Massen auf beiden Seiten der Waage muß der Schüler aus der Aufgabenstellung bzw. aus dem Bild ziehen.
Bei argumentativem Vorgehen muß der Schüler nun die gegebenen Gewichtsverhältnisse verarbeiten, indem er die Massegleichheit eines halben Ziegelsteins und des 1 kg-Körpers erkennt.
Bei rechnerischem Vorgehen gilt es, die Situation passend zu modellieren, die aufgestellte Gleichung zu lösen und das Ergebnis richtig zu interpretieren. Hier zeigt sich besonders deutlich die Gefahr, dass man die Masse des halben Ziegelsteins mit einem Kilogramm richtig bestimmt, aber nicht beachtet, dass man die Masse eines ganzen Ziegelsteins bestimmen soll. Die Gefahr ist dadurch verstärkt, dass im Bild der halbe Ziegelstein in der Seitenansicht wie ein gebräuchlicher ganzer Ziegelstein aussieht und man einen im Querschnitt quadratischen Ziegelstein im allgemeinen nicht wie im Bild längs, sondern quer halbiert. Hier wird eine Klippe für den Schüler konstruiert, der in der Testsituation nicht mit der durch diese Konstruktion geforderten Aufmerksamkeit arbeitet.
Ein weiterer Irritationspunkt ist die ungewöhnliche Form der Waage, die dazu führt, dass bei exakter Betrachtungsweise mit der geforderten Lösung die Waage gar nicht im Gleichgewicht wäre. Hier erfährt der Schüler mit Testerfahrung einen Vorteil, der exakte Betrachtungen von vornherein ausspart.
Auch die aufgezeigte Grundkonfusion bewältigt der Schüler mit Testerfahrung am besten: Nur wer stur auf die Aufgabenlösung hinarbeitet, ohne den Kontext zu beachten, den die Aufgabe stellt, kann die Konfusion irritationslos übergehen. Der Text verweist ebenso wie das Bild manifest mehrfach auf physikalische Probleme, die latent aber immer dementiert werden und die dann auch inhaltlich keine Rolle spielen. Die Aufgabe stellt ein mathematisches Problem und kann sich deshalb bezüglich ihres physikalischen Kontextes der Alltagssprache bedienen. Sie verwirrt aber durch den mehrfachen Versuch, doch noch physikalische Fachsprache zu verwenden.
Die beschriebene Konfusion wurde erklärt durch die Rekonstruktion einer Struktur der äußerlichen Verfachsprachlichung bei gleichzeitiger inhaltlicher Dementierung von Fachlichkeit. Die Struktur zeigt sich in der Art der Verwendung von „Gleichgewicht“ und Gewichts-Massen-Verwerfung. Sie korrespondiert mit einer sich mehrfach zeigenden Struktur des Nichternstnehmens des (angeblichen) realen Problems.
Multiple Choice bietet bei dieser Aufgabe die Möglichkeit zum Rückwärtsarbeiten. Der Schüler hat die gleichen Klippen wie beim Vorwärtsarbeiten zu nehmen, nämlich Erkennen der Massegleichheit auf beiden Seiten der Waage, Überwindung der textlichen Irritationen und Überwinden der Aufmerksamkeitsfalle des halben bzw. ganzen Ziegelsteins. Allerdings kann der Schüler durch Rückwärtsar-

beiten die Strukturierung der Aufgabe umgehen und mit einem anderen Rechenverfahren arbeiten (Summieren statt Vergleichen bzw. Halbieren bzw. Gleichung modellieren und lösen).

Alternative Aufgabenstellung

(Das Bild zeigt eine Balkenwaage, bei der die Waagschalen hängen und alle Körper in der Mitte der Waagschalen stehen. Die Körper sind räumlich gezeichnet, so dass der ganze und der halbe Ziegelstein in üblicher Weise als solche erkennbar sind.)

Die Gegenstände auf der Waage halten sich im Gleichgewicht. Auf der linken Waagschale befinden sich ein Körper mit einem Gewicht von 1 kg und ein halber Ziegelstein. Auf der rechten Seite befindet sich ein ganzer Ziegelstein.
Welches Gewicht hat ein ganzer Ziegelstein?

Bei Einsatz von Multiple Choice sollten die Lösungsvorgaben gleich bleiben. Der rückwärts arbeitende Schüler hat nun allerdings nach Wegfall der künstlichen Klippen den klaren Vorteil, die Aufgabe nicht mehr strukturieren zu müssen und mit einfachem Summieren arbeiten zu können. Ein Verzicht auf Multiple Choice erscheint vorteilhaft.

4.4.3. Die TIMSS-Aufgabe A3

Die Länge einer Schachtel ist auf den nächsten Zentimeter gerundet 9 cm.
Welche Länge könnte die Schachtel tatsächlich haben?

A. 10 cm
B. 9,9 cm
C. 9,6 cm
D. 8,6 cm

Lösungswege und Struktur des Problems

Der Text fragt nach einer Anwendung der Rundungsregeln, allerdings in „umgekehrter" Weise: Es soll nicht eine gegebene Zahl gerundet werden, sondern zu einer bereits gerundeten Zahl soll angegeben werden, welche Zahl gerundet worden sein könnte. Die im deutschen Mathematikunterricht gelehrten Rundungsregeln besagen für diesen Fall, dass alle Längenangaben über 8,5 cm und unter 9,5 cm zu 9 cm gerundet werden, wenn auf den Zentimeter genau gerundet werden soll. (Im Banken- und Geschäftsbereich sowie nach den mathematischen Rundungsregeln würde auch 8,5 auf 9 gerundet werden; es handelt sich bei Rundungsregeln um begründete Setzungen. Die Randproblematik steht in dieser Aufgabe aber nicht zur Debatte.)

Der wesentliche Unterschied zu den bisher besprochenen Aufgaben liegt darin, dass hier - sieht man nur den Aufgabentext ohne Multiple Choice - kein Problem mit einer einzigen Zahl als Lösung vorliegt, sondern dass es unendlich viele Zahlen bzw. alle Zahlen aus einem Intervall als Lösungen gibt. Multiple Choice verändert den Charakter der Aufgabenstellung: Man muß nicht mehr eine Zahl aus dem Intervall wählen, sondern man muß entscheiden, welche der vorgeschlagenen Zahlen im Intervall liegt. Demzufolge kann man die Aufgabe nicht mehr von ihren Multiple-Choice-Angeboten getrennt betrachten. Einen besonderen Charakter erhält die Konstruktion dadurch, dass aufgrund der Teststruktur nur eine Lösung richtig sein kann. Dadurch muß im Prinzip nicht jede Zahl daraufhin untersucht werden, ob sie im Intervall liegt.

Interpretation des Aufgabentextes

Die Länge einer Schachtel ist [...]
Manifest liegt eine kurze, knappe Einführung in einen Kontext vor. Zu erwarten ist nun eine Längenangabe, eventuell ergänzt um eine Genauigkeitsangabe wie „genau, exakt, rund, etwa".
Die Länge einer Schachtel ist 9 cm. Diese sprachliche Sequenz soll zunächst für sich interpretiert werden: Diese Sequenz ist nicht wohlgeformt. Wohlgeformte Varianten zu diesem Inhalt wären
i) Eine Schachtel ist 9 cm lang.
ii) Eine Schachtel sei 9 cm lang.
iii) Die Länge einer Schachtel beträgt 9 cm.
iv) Die Länge einer Schachtel betrage 9 cm.
v) Die Länge einer Schachtel sei 9 cm.
vi) Die Länge einer Schachtel ist/sei mit 9 cm gegeben.
Mit anderen Inhalten wären z.B. auch folgende Weiterführungen möglich:
vii) Die Länge einer Schachtel ist nicht wirklich bestimmbar.
viii) Die Länge einer Schachtel sei bekannt.
ix) Die Länge einer Schachtel ist bekannt.
x) Die Länge einer Schachtel ist/sei auf 9 cm gerundet.

Der Begriff *Länge* verweist im Vergleich zu *lang* stärker auf Fachsprachlichkeit bzw. auf die physikalische Größe Länge. Innerhalb dieses fachsprachlichen Charakters kann die Länge nun als Idealisiertes oder als Empirisches angesehen werden. Der idealisierte Charakter wird in ii, iv, v, vi, vii, viii und x versprachlicht, der empirische Charakter in i, iii und ix. Die Tatsache, dass mit ii eine idealisierte Länge angesprochen wird zeigt, dass es selbst mit dem eher alltagssprachlichen Begriff *lang* möglich ist, wohlgeformt über die idealisierte physikalische Größe Länge zu sprechen.
Der zu interpretierende Text formuliert: *Die Länge einer Schachtel ist 9 cm.* Der Vergleich mit den dargestellten Varianten zeigt, dass hier auf den empirischen Charakter der Länge rekurriert wird. Es geht also nicht um irgendeine idealisierte Betrachtung, sondern um einen alltagsnahen, eher groben Umgang mit der Länge. Es werden aber nicht die wohlgeformten Varianten i oder iii verwendet, sondern es liegt ein nicht wohlgeformter Text vor.
Der Satz iii ist offenbar ein alltagssprachlicher Satz, obwohl er sowohl die auf Fachsprachlichkeit verweisende *Länge* benutzt als auch vom *betragen* spricht, also mit dem Betrag einer Länge ein ebenfalls fachsprachliches Konzept anspricht. Die „Vergröberung" des idealisierenden *betrage* hin zum *beträgt* führt zwar in stärkere Alltagssprachlichkeit, ist aber wohlgeformt. Dies liegt offenbar daran, dass das „betragen" einer Länge, also ihre Eigenschaft als einen Betrag tragend, in die Alltagssprache übergegangen ist. Das *ist* verweigert nun die Formulierung dieser Eigenschaft des Betrag-Tragens. Es

verweigert also die Fortführung der durch *Länge* eingebrachten fachlichen Dimension durch *beträgt*. Die durch *Länge* eingeführte Fachlichkeit wird durch *ist* also dementiert. Dies erfolgt ohne Not, denn das alltagssprachliche *die Länge beträgt* ist sprachlich nicht schwieriger als *die Länge ist*.
Im Vergleich zu Satz i vermeidet der vorliegende Text die Alltagssprachlichkeit des *ist ... lang*, indem das fachsprachlichere *Länge* eingeführt wird. Eben wurde aber gezeigt, dass eben diese Fachsprachlichkeit mit dem *ist* statt *beträgt* verweigert wird. Es wird also ein Zwischending zwischen Fachsprachlichkeit und Alltagssprachlichkeit konstruiert, das in beiden Sprachlichkeiten nicht wohlgeformt ist.
Sowohl die völlig fachsprachliche als auch die völlig alltagssprachliche Variante wären dem Problem adäquat und sprachlich nicht schwieriger als die gewählte verworfene Variante. Die vom Text gewählte nichtwohlgeformte Variante hingegen bläst die Alltagssprachlichkeit des *ist* durch *Länge* auf und zerstört gleichzeitig das Fachsprachliche an *Länge* durch das *ist*.

[...] auf den nächsten Zentimeter gerundet 9 cm.
Hier wird äußerlich klargestellt, dass es nicht um ein alltägliches, sondern um ein mathematisches Problem geht: hier wird auf Zentimeter genau gearbeitet. Im Alltag gibt es „millimetergenau“ oder „auf Hundertstel Millimeter genau“ als Hinweis auf besonders genaue Angaben. „Zentimetergenau“ gibt es dort als Wort ebenso wenig wie „auf den Zentimeter gerundet“; diese Formulierungen verweisen auf einen Ort außerhalb des „normalen“ Lebens, z.B. Mathematikunterricht. Das liegt wahrscheinlich daran, dass eine Zentimetergenauigkeit gar nicht als Genauigkeit wahrgenommen wird, sondern als Ungenauigkeit, und dass Zentimetergenauigkeit selten eine Rolle spielt, z.B. beim Weitsprung oder beim Malern, wo sie kaum thematisiert werden muß.

Welche Länge könnte die Schachtel tatsächlich haben?
Hier soll zunächst die Sequenz *Welche Länge könnte die Schachtel haben?* untersucht werden. Ich wende die Kontrastierungsmethode an: Welche Fragestellungen wären denkbar?
i) Welche Länge hat die Schachtel?
ii) Welche Länge kann die Schachtel haben?
iii) Welche Länge der Schachtel wäre denkbar?
iv) Welche Länge kommt/käme für die Schachtel in Frage?
v) Welche Länge der Schachtel könnte gemessen worden sein?
Das *hat* in Formulierung i würde durch seinen festlegenden Charakter dem Intervallcharakter der Lösung nicht gerecht werden. Die verschärfte Formulierung *Welche der folgenden Längen hat die Schachtel?* wäre nicht mal wohlgeformt.
Die Formulierung ii habe ich angegeben, um zu zeigen, dass der Indikativ des *könnte* nicht wohlgeformt wäre (wahrscheinlich weil das *kann haben* hier eine konjunktive Funktion hat).
Im Vergleich zu iii und iv erscheint *könnte haben* weniger beliebig, stärker fokussierend und sprachlich einfacher. Es erhöht also das Tempo der Erfassung der Aufgabenstellung. Gleiches gilt im Vergleich mit v, hier zeigt sich aber auch, dass nicht der Unterschied zwischen der gemessenen und der gerundeten Länge thematisch sein soll, sondern der Unterschied zwischen der gerundeten und einer „seienden“ Länge. Entweder existiert diese Schachtel nur theoretisch oder ihre gemessene Länge wird

mit ihrer wirklichen Länge gleich gesetzt.

Es fällt auf, dass die Formulierung *Welche Länge könnte die Schachtel haben?* sich nicht wohlgeformt zur Antworterwartung verhält. Es sind ja unendlich viele Antworten auf diese Frage richtig. Angesichts der Multiple-Choice-Angebote wäre folgende Formulierung wohlgeformt:

vi) Welche der folgenden Längen könnte die Schachtel haben?

Man könnte die gewählte Formulierung lediglich als Abkürzung von vi ansehen. Das Wörtlichkeitsprinzip verlangt aber, dem Text keine Intention zu unterstellen, sondern den wirklich vorhandenen Text zu interpretieren. Der Satz *Welche Länge könnte die Schachtel haben?* stellt eine klare Frage, auf die es unendlich viele richtige Antworten gibt. Von diesen unendlich vielen richtigen Antworten wird hier aber nur eine einzige als richtig gewertet: 8,6 cm. Innerhalb der unendlich vielen richtigen Antworten erscheint diese Zahl beliebig. Motiviert erscheint sie lediglich in einem mathematischen Universum, das nur Dezimalzahlen mit einer Stelle hinterm Komma kennt. In diesem Universum ist 8,6 cm die erste eindeutig richtige Lösung, denn die Zahl davor (8,5) wird bei den schulischen Rundungsregeln auf 8 gerundet. In diesem Universum ist dann 9,5 die erste eindeutig falsche Lösung nach oben hin. Der Satz *Welche Länge könnte die Schachtel haben?* eröffnet aber selbst in diesem mathematischen Schrumpfuniversum noch mehr Lösungsmöglichkeiten: 8,7; 8,8; 8,9; 9,0; 9,1; 9,2; 9,3; 9,4. Von diesen 8 richtigen Lösungen wird nur eine akzeptiert, obwohl die Fragestellung alle Lösungen zuzulassen behauptet. (Es fällt auf, dass dieses Problem bei der umgekehrten Fragestellung nicht stehen würde: Sollte man 8,6 cm auf den nächsten Zentimeter runden, so gäbe es das Problem nicht.) Der Text zerstört also nicht nur das Mathematische, indem er unendlich viele Lösungen auf acht Lösungen einer Schrumpfmathematik verkürzt, er läßt nicht mal alle Lösungen des Schrumpfuniversums zu. Gleichzeitig verweigert der Text, die Aufgabenstellung sprachlich zu dem zu wenden, was die wirkliche Aufgabe ist; nämlich zu identifizieren, welche der angebotenen Zahlen in der Menge der Lösungen liegt. Dies würde durch vi sprachlich umgesetzt werden.

Welche Länge könnte die Schachtel ***tatsächlich*** *haben?*

Es geht in dieser Aufgabe nicht darum, den Unterschied zwischen gerundeter und gemessener Länge zu thematisieren, das zeigt der Vergleich mit der Alternative v: „Welche Länge der Schachtel könnte gemessen worden sein?“ Hier wird statt dessen nach einer seienden Länge gefragt. Entweder existiert die Schachtel nur theoretisch oder ihre gemessene Länge wird mit ihrer wirklichen Länge gleich gesetzt.

Nun wird mit dem *tatsächlich* gefragt, welche Länge die Schachtel „in Wirklichkeit“ oder „exakt“ hat. Dieses *tatsächlich* ist zusätzlich in den bereits diskutierten vollständigen Fragesatz eingebaut, es genügte also nicht, einfach nach der Länge der Schachtel zu fragen.

„In Wirklichkeit“ könnte die Schachtel wiederum unendlich viele Längen im Intervall [8,5cm;9,5cm] haben. Insofern wiederholt (und damit: verstärkt) sich der Widerspruch von Eröffnung unendlich vieler Möglichkeiten und Einschränkung auf eine „richtige“ Lösung. „Exakt“ oder „wirklich“ kann man die Länge ohnehin nicht angeben, wohl aber auf einer höheren Genauigkeitsstufe. Der latente Sinn des *tatsächlich* wäre dann also gleichzeitig die Aufforderung, sich auf die nächste Genauigkeitsstufe zu begeben. Dies produziert einen Widerspruch zwischen dem manifesten Sinn des *tatsächlich*, der behauptet, die Länge der Schachtel wäre tatsächlich 8,6 cm, und dem latenten Sinn, der lediglich besagt, die Länge der Schachtel wäre auf der nächsten Genauigkeitsstufe 8,6 cm. Diese Verwerfung ist völlig überflüssig, weil das *tatsächlich* textlich nicht benötigt wird.

Zusammenfassung

Die Aufgabe testet die Anwendung der schulmathematischen Rundungsregeln, und zwar in „umgekehrter" Weise: Es wird nach dem „tatsächlichen" Wert einer gerundeten Meßgröße gefragt. Das Besondere dabei ist, dass ohne Multiple Choice ein ganzes Intervall Lösung wäre. Multiple Choice beeinflußt den Charakter der Aufgabe also prinzipiell. Man darf bzw. muß nicht eine Zahl aus dem Rundungsintervall wählen, sondern man muß entscheiden, welche der vorgeschlagenen Zahlen im Intervall liegt – was sprachlich nicht umgesetzt ist. Da nur ein Wert angekreuzt werden kann, wird nicht erfaßt, ob der Schüler das Gesamtproblem des Lösungsintervalls erfaßt hat.

In dieser Aufgabe wird der Schüler in ein mathematisches Schrumpfuniversum gepreßt. Das korrespondiert mit der Verwerfung von Fachsprachlichkeit und Alltagssprachlichkeit bei *Die Länge einer Schachtel ist [...]*. Sowohl die völlig fachsprachliche als auch die völlig alltagssprachliche Variante wären hier dem Problem adäquat und sprachlich nicht schwieriger als die gewählte verworfene Variante. Die vom Text gewählte nichtwohlgeformte Variante hingegen bläst die Alltagssprachlichkeit des *ist* durch *Länge* auf und zerstört gleichzeitig das Fachsprachliche an *Länge* durch das *ist*.
Die Unterminierung des Mathematischen wiederholt sich in der Formulierung *Welche Länge könnte die Schachtel haben?,* die sich nicht wohlgeformt zur Antworterwartung verhält: Es sind ja unendlich viele Antworten auf diese Frage richtig. Der Widerspruch von Eröffnung unendlich vieler Möglichkeiten und Einschränkung auf eine „richtige" Lösung wiederholt (und damit: verstärkt) sich im *tatsächlich.* Wie die anderen Verwerfungen wäre auch diese Verwerfung leicht zu heilen, weil das *tatsächlich* textlich nicht benötigt wird.

Alternative Aufgabenstellung:

Ich halte eine Multiple-Choice-Aufgabe zum Testen der Rundungsregel für ungeeignet, da kaum vernünftig getestet werden kann, ob der Schüler die gesamte Regel einschließlich des Intervallproblems erfaßt hat. Habituell wird dadurch eine Schrumpfmathematik vertreten. Will man aber unbedingt auf diese Weise testen, wäre nachfolgende Variante vertretbar. Der Aufgabencharakter ist allerdings vielfältig variierbar, wie die in der Interpretation dargestellten Formulierungsalternativen gezeigt haben.

Die Länge einer Schachtel sei - auf Zentimeter gerundet - 9 cm.
Welche der folgenden Längen könnte die Schachtel haben?
A. 9,9 cm B. 9,6 cm C. 8,6 cm D. 8,4 cm

4.4.4. Die TIMSS-Aufgabe A4

In der gleichen Zeit, in der Carola 3 Platzrunden läuft, schafft Alice 4 Platzrunden. Wie viele Runden ist Alice gelaufen, wenn Carola 12 Runden geschafft hat?

A. 9

B. 11

C. 13

D. 16

Lösungswege und Struktur des Problems

Die Aufgabe ist (außerhalb von Raten) nur lösbar, wenn man voraussetzt, dass es sich um eine proportionale Zuordnung handelt, dass also das Verhältnis der von den beiden Mädchen gelaufenen Runden sich nicht ändert. Mögliche Lösungswege sind folgende:

- Schrittweises Vorgehen: Wenn Carola 3 Runden geschafft hat, hat Alice 4 Runden geschafft. Wenn Carola 6 Runden hat, hat Alice 8 Runden. Wenn Carola 9 Runden hat, hat Alice 12 Runden. Wenn Carola 12 Runden hat, hat Alice 16 Runden. Durch das Anlegen einer Tabelle kann man dieses Vorgehen unterstützen. Dieses schrittweise Vorgehen kann auch davon ausgehen, dass eine Runde bei Carola eine und eine drittel Runde bei Alice ist. Man kann Alices Runden dann schrittweise aus Carolas 12 Runden „herausaddieren".

- Aus dem letzten Gedanken ergibt sich ein Lösungsweg mit Bruchmultiplikation: Wenn Alice in jeder Carola-Runde eine und eine drittel Runde läuft, dann ist sie nach zwölf Carola-Runden $12 \cdot 1\frac{1}{3} = 12 \cdot (1+\frac{1}{3})$ oder $12 \cdot \frac{4}{3} = 16$ Runden gelaufen.

- proportionales Argumentieren: Carola ist nach 12 Runden viermal soviel gelaufen wie zum ersten Zeitpunkt. Alice ist also auch viermal soviel gelaufen wie zum ersten Zeitpunkt, sie ist also 16 Runden gelaufen.

- Argumentieren mit Verhältnissen: **Fehler! Textmarke nicht definiert.** $\frac{3}{12} = \frac{4}{x}$ oder 3:4 = 12:x, also ist $x = \frac{12*4}{3} = \frac{48}{3} = 16$

In dieser Aufgabe geht es darum, das Verhältnis zwischen zwei Zahlen - der jeweils von den Mädchen gelaufenen Anzahl von Runden, also drei zu vier - zu erfassen und mit diesem Verhältnis dann weiter zu arbeiten.

Die erste Hürde für denjenigen, der die Aufgabe lösen möchte, besteht im Erkennen des Verhältnisses drei zu vier der von den Mädchen gelaufenen Runden. Es ist keineswegs selbstverständlich, dass es sich um ein proportionales Verhältnis handelt. Beim gedanklichen Verbleib im angesprochenen sportlichen Kontext ist es denkbar, dass es sich bei Alice um eine Sportlerin handelt, die am Beginn eines Langlaufes sehr schnell ist, ihren Vorsprung aber dann nicht mehr ausbaut. Der Vorsprung von einer Runde bliebe dann im weiteren Verlauf des Laufes erhalten, Alice wäre 13 Runden gelaufen. Ebenso ist denkbar, dass Carola eine Sportlerin ist, die langsam losläuft und schneller wird. Der Vorsprung von Alice würde sich dann verringern, Carola könnte Alice im Laufe der Zeit sogar überholen. Lediglich das Wissen um die Konstruktion von Schulaufgaben oder die Erkenntnis, dass das Ergebnis ohne Annahme von Proportionalität nur zu erraten ist, zwingen zur Annahme von Proportionalität.

Hat der Schüler das Problem als Verhältnisproblem erkannt und das Verhältnis von drei zu vier erfaßt, so muß er dieses Verhältnis im weiteren in die Beantwortung der Frage umsetzen. Dazu muß er mindestens einen der oben beschriebenen Gedankengänge vollziehen. Dabei ist entweder ein Verständnis von Proportionalität oder von Verhältnisbeziehungen vonnöten. Der erste Lösungsweg setzt ein einfaches und prozessuales Verständnis von Proportionalität voraus. Der Gedankengang des dritten Lösungsweges erfordert ein etwas abstrakteres Proportionalitätsverständnis. Der vierte Gedankengang erfordert Verständnis von Verhältnisbeziehungen und rechnerische Fähigkeiten sowie Fähigkeiten im Umgang mit Gleichungen.

Interpretation des Aufgabentextes

In der gleichen Zeit, in der Carola 3 Platzrunden läuft, schafft Alice 4 Platzrunden.

Im Sinne einer schnellen Aufgabenerfassung wird durch *In der gleichen Zeit* ein klarer Rahmen gesetzt, sodann wird das Verhältnis der von den Mädchen gelaufenen Runden knapp und klar beschrieben. Es handelt sich um Platzrunden und nicht um irgendwelche Runden. Alice *schafft* ihre Runden.

Die Formulierung *In der **gleichen** Zeit ...* ist sprachlich wenig wohlgeformt. Wohlgeformt wäre die Variante *In der Zeit, in der ...* Wenn man die dadurch gegebene Gleichheit der Zeit-Rahmung betonen wollte, müßte es heißen *In der **selben** Zeit, in der ...* Die Verwendung des nicht wohlgeformten *gleichen* betont nicht nur (im Sinne einer Verstehensbeschleunigung) die inhaltliche Gleichheit der Zeit-Rahmung, sondern erscheint im Wohlgeformtheitsvergleich als verschärfter sprachlicher Hinweis auf Gleichheitsbetrachtungen in einem mathematischen Sinn. Folgt man dem in der so nahe gelegten Formalität, müßte man allerdings hinschreiben: 3 Platzrunden $\Leftrightarrow$ 4 Platzrunden oder $t_{Carola} = t_{Alice}$. Im Sinne der Erfassung des Problems wäre aber sinnvoller, einen Rahmen mit dem Titel „gleicher Zeitraum“ zu schaffen, in den man reinschreibt: Carola: 3 Runden, Alice: 4 Runden. Das *gleichen* soll also das Verstehen der Aufgabe beschleunigen, trägt in sich aber das latente Potential der Fehlleitung.

Das Wort *Platzrunden* irritiert zunächst. Es ist in unserem Sprachgebrauch so selten, dass man einen Moment darüber nachdenken muß, worum es sich handelt. Es verweist dann aber eindeutig auf einen sportlichen, leistungsorientierten Kontext. Dies stellt eine erste Kanalisation der möglichen Interpretationen des Wirklichkeitskontextes auf eine Proportionalität, zumindest aber auf ein Gleichbleiben des Unterschiedes hin dar: Bei einer leistungsorientierten sportlichen Betätigung ist die Wahrscheinlichkeit, dass sich das Leistungsverhältnis der beiden Mädchen nicht ändert, größer als z.B. bei einem Spaziergang um den Teich auf dem Schulhof.

Carola *läuft* ihre Runden, Alice *schafft* ihre Runden. Da Alice immerhin ein Drittel mehr gelaufen ist, liegt eine Vorstellung von „schafft es gerade so“ nicht nahe. Eher könnte man sagen, sie „schafft sogar 4 Runden“ (im Vergleich zu Carola, die nur 3 schafft). *Schafft* stellt hier die besondere Leistung von

Alice gesondert heraus. Dies deutet darauf hin, dass Alice (an diesem Tag?) prinzipiell schneller ist als Carola, dass also nicht davon auszugehen ist, dass Carola Alice noch überholt. Hier wird also latent weiter auf Proportionalität bzw. Gleichbleiben des Unterschiedes hingearbeitet.

Wie viele Runden ist Alice gelaufen, wenn Carola 12 Runden geschafft hat?

Hier dreht sich die Bedeutung des *geschafft* nun um. Es bekommt den Sinn: „wenn Carola endlich 12 Runden geschafft hat", wenn sie also endlich auch so weit ist. Diese Interpretation wird gestützt, wenn man sich vorstellt, die Frage lautete der bisherigen Wortzuordnung folgend: *Wie viele Runden hat Alice geschafft, wenn Carola 12 Runden gelaufen ist?* Die Fragestellung wäre nicht wohlgeformt, wenn man weiß, dass Alice auch weiterhin schneller war, weil das *schaffen* im Sinne der Heraushebung einer besonderen Leistung nur bei gleichzeitiger Nennung der Leistung und bei Aussagen mit Vergleichscharakter funktioniert. („Wieviel hat er heute geschafft?" deutet ebenso wenig wie „Hat sie heute 16 Runden geschafft?" an, dass man herauszustellende Leistungen annimmt.) Diese Fragestellung würde also jeden Hinweis darauf verwischen, wer zum zweiten Zeitpunkt vorn liegt.

Was wäre, wenn die Fragestellung lautete: *Wie viele Runden ist Carola gelaufen, wenn Alice 16 Runden geschafft hat?* oder *Wie viele Runden hat Carola geschafft, wenn Alice 16 Runden gelaufen ist?* Klar ist, dass die Aufgabe dadurch einfacher würde, weil die Reihenfolge der Nennung der Namen im Vergleich zum ersten Satz unverändert ist und das Verhältnis leichter zu erfassen ist. Für den ersten Fall wird sofort deutlich, dass das *schaffen* seinen die besondere Leistung von Alice hervorhebenden Charakter verliert: Es fehlt der Vergleichsmaßstab. Alice hat dann eben endlich 16 Runden geschafft und ist jetzt fertig. Die zweite Aussage enthält nur einen schwachen latenten Hinweis auf die Vergleichbarkeit der Leistungen. Es bleibt eher unklar, ob Carola ihre Runden „erst geschafft" oder „schon geschafft" hat. Nur im Zusammenhang mit Satz 1 wäre ein „erst geschafft" wahrscheinlicher. Diese beiden Fragestellungen enthielten also keinen bzw. nur einen schwachen latenten Hinweis auf die Reihenfolge des Einlaufs.

Durch die im Vergleich zu Satz 1 umgedrehte Bedeutung des *geschafft* weist die Fragestellung der Aufgabe also latent darauf hin, dass Alice weiter schneller war. Sie weist dadurch, dass Carola es „nun endlich auch geschafft" hat, sogar darauf hin, dass der Unterschied zwischen beiden Mädchen (weiterhin) groß ist. Die Aufgabenstellung stützt also latent das Verständnis der Aufgabe als Verhältnis- bzw. Proportionalitätsaufgabe bzw. (in weit geringerem Maße) als Aufgabe mit gleich bleibendem Unterschied.

Das Problem bei dieser latenten Stützung beider Sichtweisen ist, dass für die Sichtweise des Gleichbleibens des Unterschiedes eine Lösungsoption vorhanden ist, dass diese Lösung aber als falsch gewertet wird. Als richtig zählt nur die proportionale Sichtweise. Hier liegt also eine Falle vor. Wenn man das Problem inhaltlich betrachtet, liegt nämlich eine Proportionalität nur dann nahe, wenn man bereits weiß, dass hier Proportionalität getestet werden soll. Im Unterricht ist diese eingekleidete Aufgabe deshalb nur sinnvoll verwendbar, wenn angekündigt ist, sie zur Illustration von Proportionalität zu verwenden und wenn die Grenzen dieser Modellierung diskutierbar sind. Selbst in einer Klassenarbeit bestünde ein Teil der getesteten Leistung im erfolgreichen Erraten des Denkens des Lehrers. In einem Test – in dem weder Nachfragen möglich sind noch eine Einbettung in unterrichtliche Deutungsusancen besteht – ist diese Falle problematisch, denn die Schärfe des Meßprozesses wird gestört: Ein Schüler, der proportionale Probleme erkennen und lösen kann, kann hier zum falschen Resultat gelangen, weil das Problem nicht zwingend proportional ist und weil der Test keinen Raum zur Klärung des „Gemeinten" läßt. Das Erkennen des „Gemeinten" wird also mitgemessen, und das ist eher ein bestimmter Umgang mit schulmathematischem Habitus als ein Problem mathematischer Leistungsfähigkeit.

Warum erscheint das vorgelegt Problem als Proportionalitätsproblem konstruiert und unecht? Schließlich wurde ja deutlich, dass die Aufgabe sprachlich massiv auf Proportionalität (bzw. auf Gleichbleiben des Unterschiedes) verweist. Zudem wurde deutlich, dass immer dann eine weitere Information gegeben werden müßte, wenn mit der vorliegenden Frage **nicht** auf Proportionalität oder Gleichbleiben des Unterschiedes gezielt würde.
Das Problem ist, dass „im echten Leben" niemand so eine Frage stellen würde, wie sie hier vorliegt. Erst die Frage als Frage gibt das Signal für die Definition der Situation. Man stelle sich einen Reporter vor, der den Lauf der Mädchen kommentiert. Er würde sagen: Wenn sie so weiter laufen, dann ist Alice 16 Runden gelaufen, wenn Carola 12 Runden geschafft hat. Völlig unvorstellbar wäre, dass er die vorliegende Frage stellt. Dies ist nur in Schulaufgaben möglich, bei denen von sinnhaften Kontexten der Fragestellung abgesehen werden kann. Damit ist im Test derjenige im Vorteil, der dieses Konstruktionsprinzip von Schulaufgaben zumindest latent durchschaut bzw. sich ihm entsprechend verhält, es bedienen kann.

Multiple Choice

Bei dieser Aufgabe ergeben sich durch die Nutzung von Multiple Choice keine Möglichkeiten, zur Lösung zu gelangen, ohne das Grundproblem erfaßt zu haben. Multiple Choice ermöglicht aber den Verzicht auf Rechenschritte und beseitigt Rechenfehlerquellen, eröffnet aber auch „Fehler"quellen. Es ermöglicht auch, die Aufgabe lediglich mit einem Vorverständnis von Proportionalität bzw. Verhältnissen zu bewältigen.

- Der erste Lösungsansatz des schrittweisen Vorgehens, quasi Auszählens der Lösung, funktioniert nur, wenn man vorwärts arbeitet und den Gedanken bis zu Ende führt. Hier schaltet Multiple Choice im Vergleich zum Hinschreiben der Lösung lediglich reine Rechenfehler aus.
- Auch für den zweiten Lösungsansatz des rechnerischen Argumentierens bietet Multiple Choice keine Möglichkeit, das Lösen der Aufgabe zu umgehen. Die Wahrscheinlichkeit, sich zu verrechnen, ist hier ohnedies geringer als beim ersten Ansatz, da weniger Rechenschritte zu vollziehen sind.
- Geht der Schüler den Weg von Lösungsansatz drei, argumentiert er also mit Verhältnissen, so ergibt sich die Möglichkeit des Rückwärtsarbeitens. Stellt er eine Verhältnisgleichung auf, so kann er, wenn er das Prinzip der Proportionalität verstanden hat, sehen, dass die Lösung 16 ist. Hat er die Proportionalität nicht vollständig verstanden (bzw. denkt, dass die Runden*differenz* gleich bleibt), so ist mit der 13 eine Fehlerquelle gegeben. Es ist nicht zu unterscheiden, ob ein Schüler, der die 13

gewählt hat, das *Prinzip* der Proportionalität nicht verstanden hat oder ob er das Problem lediglich im Sinne eines gleich bleibenden Unterschieds interpretiert.
Arbeitet der Schüler an dieser Stelle rückwärts, so spart er den Schritt des Umstellens bzw. Rechnens ein. Das Verständnis des Verhältnisproblems ist auf jeden Fall nötig, um rückwärts arbeiten zu können.

- Multiple Choice bietet eine Möglichkeit für rückwärts arbeitende Plausibilitätsbetrachtungen: Alice ist schneller als Carola, also kommen nur 13 oder 16 in Frage. 13 ist „zu wenig" größer als 12. Alice hatte ja schon früher eine Runde Vorsprung vor Carola, der Vorsprung müßte jetzt größer sein. Also kommt nur 16 in Frage.
  Für diese Vorgehensweise muß der Schüler verstanden haben, dass im Sinne der Aufgabenkonstruktion Alice schneller bleibt als Carola und die 13 als Lösung ausschließen können. Dafür benötigt der Schüler lediglich ein im Alltagswissen vorhandenes Vorverständnis von Proportionalität. Hier ist Cleverness im Umgang mit Testaufgaben gefordert, es wird also für diesen Schüler Testfähigkeit gemessen.

Multiple Choice bietet in dieser Aufgabe also Möglichkeiten, mit Rückwärtsarbeiten leichter zu einer Lösung zu gelangen. Das Grundproblem muß dazu aber erfaßt werden, das bedeutet auch, die Schulaufgabenkonstruktion zu bedienen.

Multiple Choice bietet die Möglichkeit, die Ratewahrscheinlichkeit zu erhöhen, wenn der Schüler wenigstens erfaßt, dass Alice mehr als Carola schafft. Dann verbleiben nur noch 13 und 16 als Lösung. Die richtige Entscheidung zwischen beiden erfordert entweder das Verständnis von Verhältnisproblemen oder ein Vorverständnis von Proportionalität. Hat der Schüler dieses (Vor-)Verständnis, so kann er 16 richtig bestimmen, ohne einen der drei beschriebenen Lösungswege zu beherrschen. Multiple Choice eröffnet somit einer rückwärtsarbeitenden Ratestrategie Erfolgschancen. Sie ist durchaus anspruchsvoll, es wird dann aber nicht das Verständnis von Proportionalität gemessen, sondern das Vorverständnis und Testfähigkeit.

Umgang mit Schülerfehlern

Ich möchte noch auf Schülerfehler verweisen, die bei sinnvollerer Konstruktion mit Multiple Choice diagnostiziert werden könnten, auf die man die Schüler aber durch Multiple Choice auch hinweisen kann, indem man die entsprechenden Lösungen nicht angibt:

- Hat der Schüler das Prinzip der Proportionalität nicht verstanden, sondern zählt die eine Runde Vorsprung von Alice einfach weiter, so kommt er auf 13, also weiterhin eine Runde mehr als Carola. Man könnte bei den Schülern, die 13 angekreuzt haben, also dieses Nichtverstehen von Proportionalität diagnostizieren. Dies wird aber durch das oben beschriebene Problem der Konstruiertheit der Fragestellung unmöglich. Ein Schüler, der sich lediglich auf das Konstruktionsprinzip von Schulaufgaben nicht einläßt, kann hier trotz Kenntnis der Proportionalität scheitern. Die Lösung 13 diagnostiziert damit das Nichtverständnis von Proportionalität nicht mehr sicher.
- Auf die Lösung 9 kommt der Schüler, wenn er zwar die Proportionalität erfaßt hat, aber verwech-

selt, welches der Mädchen den Vorsprung hat. Mit der Angabe der Lösung 9 wird also der Schüler bestraft bzw. diagnostiziert, der unter dem Zeitdruck des Tests nicht „aufmerksam genug“ arbeitet. Wäre die 9 als Lösung nicht angegeben, so hätte dieser Schüler, der über das eigentlich gefragte mathematische Verständnis verfügt, die Chance, seine „Unaufmerksamkeit“ zu berichtigen.

– Auf die Lösung 11 kommt man „sinnvoll“, wenn man sowohl verwechselt, welches der Mädchen schneller war, als auch die Proportionalität mißversteht. Man kommt dann zu dem Ergebnis, dass die vermeintlich Langsamere eine Runde weniger geschafft hat. Dieses Lösungsangebot könnte also Nichtverständnis von Proportionalität diagnostizieren. Auch dieses Potential wird hier verschenkt, denn die lebensweltliche Einbettung gestattet die Annahme, Carola sei eine Sportlerin, die ihre Stärken erst am Ende eines Rennens entfaltet, Alice hingegen eine Sportlerin, die zum Ende schwächer wird.

Zusammenfassung

Es handelt sich um eine Aufgabe, die das Erfassen und Verarbeiten von Verhältnissen bzw. Proportionalitäten verlangt. Durch den gewählten Kontext und den Aufgabentext wird diese Aufgabenstellung allerdings nicht klar. Nur die Kenntnis der Strukturiertheit von Schulaufgaben ermöglicht es hier, relativ eindeutig eine Entscheidung darüber zu treffen, was die Tester für richtig halten. Es werden jene Schüler belohnt, die die Struktur von Schulaufgaben bedienen können. Selbst der Schüler, der die Proportionalität verstanden hat, kann daran scheitern, dass der Kontext und der Aufgabentext die Proportionalität nicht zwingend nahelegen. Dieses Manko wird dadurch abgeschwächt, dass die latente Sinnstruktur der Aufgabenstellung stark auf Proportionalität verweist.

Multiple Choice liefert in dieser Aufgabe Möglichkeiten, die Aufgabe auch ohne Verständnis – sondern nur mit Vorverständnis - für Proportionalität bzw. Verhältnisse zu lösen. Es bietet Rechenerleichterungen für diejenigen, die die Aufgabe durch Proportionalitäts- bzw. Verhältnisbetrachtungen angehen. Es liefert Fallstricke für jene Schüler, die die mißverständliche Aufgabenstellung nicht im Sinne der Aufgabensteller interpretieren. Die hier vorliegenden Möglichkeiten zur Diagnose von Schülerfehlern werden durch die Aufgabenstellung verschenkt.

Es liegt insgesamt also die Fehlkonstruktion einer Testaufgabe vor: Einerseits kann ein Schüler scheitern, der über die zu messende Fähigkeit verfügt, andererseits kann ein Schüler zum Erfolg gelangen, der nicht über die zu messende Fähigkeit verfügt.

Alternative Aufgabenstellung:
Die vorliegende Aufgabenstellung scheint mir zur Betrachtung von Proportionalität in einem Test ungeeignet. Man sollte einen Kontext wählen, der proportionales Denken zwingend erfordert bzw. die proportionale Modellierung explizieren.

4.4.5. Die TIMSS-Aufgabe A5

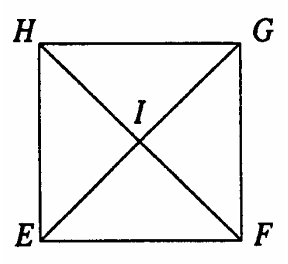
Welche der Aussagen über das Quadrat EFGH ist FALSCH?

A. Δ EIF und Δ EIH sind kongruent (deckungsgleich).

B. Δ GHI und Δ GHF sind kongruent (deckungsgleich).

C. Δ EFH und Δ EGH sind kongruent (deckungsgleich).

D. Δ EIF und Δ GIH sind kongruent (deckungsgleich).



Lösungswege und Struktur des Problems

- Ein Quadrat wird durch seine Diagonalen in zwei jeweils kongruente Dreiecke geteilt. (Die Argumentation dazu kann z.B. über den Kongruenzsatz sss laufen.) Es sind sogar alle vier Dreiecke, die durch diagonale Halbierung des Quadrates entstehen können, kongruent.
Außerdem sind die vier „kleinen” Dreiecke , die jeweils I als einen Eckpunkt haben, kongruent. (Für die Argumentation kann man die Tatsache benutzen, dass sich die Diagonalen gegenseitig halbieren und dass die vier Seiten des Quadrates gleich lang sind.)
Mit diesem Wissen muß man nun die vier Angebote von Multiple Choice durchsuchen. Die Aussagen A., C. und D. sind richtig, die Aussage B. ist falsch. Zum schnelleren Durchsuchen der Multiple-Choice-Angebote bietet sich an, nach der Aussage zu suchen, in der der Punkt I in dem einem Dreieck vorkommt, im zweiten nicht. Diese zwei Dreiecke können nicht kongruent sein.
- Vorgehen nach Augenschein: Nach Augenschein sind - wenn man von A. nach D. vorgeht, die Dreiecke EIF und EIH kongruent, die Dreiecke GHI und GHF nicht kongruent. An dieser Stelle kann man bereits abbrechen, weil die Aufgabe bereits gelöst ist. Man kann auch noch C. und D. betrachten, auch diese Aussagen sind nach Augenschein wahr.

Bei dieser Aufgabe zeigen schon die zwei vorgeschlagenen Vorgehensweisen, dass man die Struktur des Problems auf verschiedene Weisen beschreiben kann. Man kann diese Vorgehensweisen auf die Fähigkeiten des zu Testenden beziehen. Entweder er ist mit dem Kongruenzbegriff (mehr oder weniger) vertraut oder er ist es nicht.
1. Man kann das Problem als Problem betrachten, welches die Vertrautheit des Schülers mit dem Kongruenzbegriff testet. Um den ersten Weg zur Lösung gehen zu können, benötigt man nicht nur Wissen um den Kongruenzbegriff, man muß sogar mindestens einen Kongruenzsatz für Dreiecke kennen. Um den zweiten Weg zu gehen, muß man nur eine etwaige Vorstellung davon haben, was es bedeutet, wenn zwei Dreiecke kongruent sind - sie sollen „gleich groß“ oder „irgendwie gleich“ sein; über den Inhalt des Kongruenzbegriffs als Deckungsgleichheit muß man hier nichts wissen. Man gelangt schnell zur Erkenntnis, dass die Dreiecke GHI und GHF nicht mal ansatzweise gleich groß sind.

Man muß sich nicht weiter damit auseinandersetzen, ob die anderen Dreieckspaare jeweils wirklich gleich groß sind bzw. kann sich damit begnügen, dass sie dem Augenschein nach „so ziemlich" gleich groß sind. Der Kongruenzbegriff spielt dabei kaum noch eine ernstzunehmende Rolle.

2. Man kann das Problem als Problem betrachten, welches den Schüler herausfordert, effektiv herauszufinden, was „man von ihm will". Jeweils zwei Dreiecke sollen „deckungsgleich" sein. Wem sich der Begriff nicht von selbst erschließt, der muß sich die in den Lösungen beschriebenen Dreiecke ansehen. So kann er herausfinden, dass Deckungsgleichheit nichts anderes ist als „Gleichheit" im Alltagsverständnis und er kann sehen, dass B. dabei herausfällt. Für ihn würden dabei verbale und Kombinationsfähigkeiten und wohl auch eine gewisse Unverfrorenheit gegenüber dem mathematischen Impetus der Aufgabe gemessen. In jedem Fall ist das, was hier gemessen wird, eher Testfähigkeit als die Fähigkeit, mit dem Kongruenzbegriff umzugehen.

Interpretation des Aufgabentextes

Da die Multiple-Choice-Antworten Bestandteil der Aufgabenstellung sind, müssen sie mit der Aufgabenstellung interpretiert werden.

Welche der Aussagen über das Quadrat EFGH [...]

Hier fällt zunächst auf, dass im Vergleich zu anderen interpretierten Aufgaben die Hinführung auf das Problem weniger direkt und unvermittelt erfolgt. Nach der sich dort zeigenden Struktur der schnellen Erfaßbarkeit und Hinführung würde hier z.B. stehen: „Betrachte das Bild! Welche der Aussagen über das Quadrat EFGH ..." Perfekter auf diese Strukturlogik zugeschnitten wäre „Betrachte das Quadrat EFGH! Welche der folgenden Aussagen ..." Hier wäre die schnelle Hinführung mit dem Verweis auf den Quadratcharakter von EFGH verbunden. Der Verweis auf die Aussagen wäre nicht nur von dieser Charakterisierung getrennt, was die Übersichtlichkeit erhöhte, sondern die Reihenfolge der Bearbeitung (1. Analyse der Situation, 2. Analyse der Aussagen) wäre bereits durch die Reihenfolge in der Aufgabenstellung vorweggenommen.

Im Vergleich gerade zur letzten Version verschafft die im Text gewählte Konstruktion wenig Orientierung, obwohl sie durch äußerliche Kürze den Anschein erweckt, der genannten Strukturlogik zu folgen.

Welche der [...]: Geschichten:

a) Welche der Hosen willst du anziehen?
b) Welche der Aussagen stammen von Ihnen?
c) Welche der Mädchen werden heiraten?
d) Welches der Mädchen wird als erste heiraten?
e) Welche der Lampen würdest du eher bevorzugen?

Lesart: Es gibt eine Anzahl gleichkategorialer Entitäten, die vorher vorgestellt wurde oder nachher vorgestellt wird oder gerade offensichtlich ist. Es muß eine Frage folgen. *Welche der* ... stellt sogar eine Zuspitzung der Fraglichkeit dar, denn man ist aufgefordert, sich zu entscheiden. Die Stärke des Drängens zu einer Entscheidung erkennt man durch e), denn selbst der ansonsten wenig nachdrücklich formulierte Satz verlangt durch das *Welche der* ... eine klare Entscheidung. Die Entscheidung muß allerdings nicht eindeutig sein, wie ein Vergleich von c) und d) zeigt. Die Eindeutigkeit - und damit das Beschränken auf ein einziges Kreuz - wird in der Aufgabe erst am Ende des Satzes durch *ist* eingefordert.

[...] Aussagen über das Quadrat EFGH [...]: Die Kategorie, innerhalb derer zu entscheiden ist, heißt „Aussagen über das Quadrat EFGH“. Es werden also noch Aussagen zu treffen sein. Der Charakter der Figur EFGH als Quadrat ist damit eingeführt. Es handelt sich um ein spezielles Quadrat, nämlich *das* Quadrat EFGH, es muß für den Schüler also mental vorhanden sein, er muß es also aus dem Text und der Zeichnung heraus repräsentieren.

[...] ist FALSCH?: Zur Interpretation dieser Verbindung bietet sich eine Betrachtung kontrastierender Möglichkeiten an:
Welche Aussagen ... kannst Du machen? Dies wäre eine völlig offene Aufgabenstellung für den Schüler, der sich unter allen möglichen Blickwinkeln mit der gegebenen Situation auseinandersetzen soll, um sich dann z.B. einem interessanten Gesichtspunkt näher zuzuwenden.
Welche Aussagen ... kannst Du über bestehende Kongruenzen machen? Auch diese Aufgabenstellung wäre noch offen, würde den Blickwinkel aber bereits einschränken.
[...] ist richtig? Dies wäre die erste Multiple-Choice-taugliche Aufgabenstellung und ist wohl die am ehesten zu erwartende Aufgabenstellung für unseren Test. Es wäre aus vier Aussagen die richtige zu wählen. Aus der Betrachtung der Struktur des Problems wurde aber schon oben deutlich, dass in Bezug auf Kongruenzfragen hier nur eher triviale Aussagen möglich sind, da per Augenschein ersichtlich ist, welche der Dreiecke für Kongruenz überhaupt in Frage kommen.
[...] ist falsch? Dies ist nun die Umkehrung der eben besprochenen Variante. Die Kontrastierung zeigt, dass sich der mathematische Anspruch nicht erhöht, sehr wohl aber der kognitive: Nun ist eine Negativaussage zu bearbeiten, was eine gedankliche Umkehrung erfordert. Das erhöht den kognitiven Anspruch der Aufgabe.
[...] ist FALSCH? Die Umkehrungsanforderung erklärt die Hervorhebung des FALSCH durch Großbuchstaben: Das Ungewöhnliche muß hervorgehoben werden, um eine Verwechslung mit der erwartbaren Anforderung zu vermeiden. In der Variante “...ist richtig” käme der Gedanke, RICHTIG groß zu schreiben, wohl eher nicht auf. Die Großschreibung in Aufgabentexten stellt allerdings ein Irritationsmoment dar - für testunerfahrene Schüler. Der Schüler mit Testerfahrung ist mit diesen Mitteln zur Vereindeutlichung des Textes und Beschleunigung der Aufgabenbearbeitung vertraut.
Die sprachliche Verkomplizierung durch das *“... ist FALSCH”* nimmt die Strukturlogik auf, nach der eine an sich mathematisch anspruchslose Fragestellung künstlich verkompliziert wird: Hier sind acht verschiedene Dreiecke auseinanderzufummeln (jedes andere Wort wäre beschönigend), es ist reine Fleißarbeit zu verrichten, deren Anspruch aus dem zu lösenden Problem heraus nicht zu begründen ist. Mathematisch wird hier wenig vollbracht.[45]

[45] Dies ist eine Stelle für den üblichen Einwand, dass „man doch nicht päpstlicher als der Papst” sein solle und dass dieses kleinkarierte Vorgehen doch gar nichts damit zu tun hat, wie der Schüler mit der Testaufgabe umgeht oder wie die Intention der Tester war. In der Objektiven Hermeneutik geht es zunächst lediglich darum, die objektive Textstruktur zu erfassen. Die Frage lautet nur: Was sagt der Text? In dieser Aufgabe zeigt sich dabei, dass „der Text sich selbst nicht ernst nimmt” und dass dies die inhaltliche Strukturlogik aufnimmt, nach der

ΔEIF und ΔEIH sind kongruent (deckungsgleich). Zunächst fällt eine Diskrepanz zur Ausgangsfrage ins Auge: *Welche der Aussagen* ***über das Quadrat EFGH*** *ist FALSCH?*
Die Antworten von Multiple Choice treffen gar keine Aussagen über das Quadrat EFGH selbst. Sie treffen Aussagen über Dreiecke, die dem Quadrat EFGH einbeschrieben sind. Die Frage ist also nicht so ganz ernst zu nehmen. Diese Fehlkonstruktion erfolgt ohne Not, denn man könnte ja schreiben: „EFGH ist ein Quadrat. Welche der folgenden Aussagen ist FALSCH?“ oder „EFGH sei ein Quadrat. Welche der folgenden Aussagen ist dann falsch?“
Wenn man sich vorstellt, dass man die oben genannte offene Konstruktion „Welche Aussagen ... kannst Du machen?” ausfüllt, so könnte dort kaum „Welche Aussagen über das Quadrat EFGH kannst Du machen?” stehen, wenn man die Hoffnung hat, die Schüler mögen sich mit Kongruenzen befassen. Dafür müßte dort stehen: „Welche Aussagen über die Dreiecke im Quadrat EFGH ...” oder „Welche Aussagen über die dem Quadrat EFGH eingeschriebenen Dreiecke ...” oder auch nur „Welche Aussagen über die Dreiecke kannst Du machen?”
Die Tatsache, dass die Fragestellung sich selbst nicht ernst nimmt, nimmt wiederum die inhaltliche Logik der Aufgabe auf, nach der es zwar äußerlich um das Testen des Wissens von Kongruenz geht, man aber gar nichts oder wenig über Kongruenz wissen muß, um die Aufgabe zu lösen. Hier wird der mathematische Inhalt nicht ernst genommen.
In der Verwendung von *kongruent* und *deckungsgleich* spiegelt sich ein wahrscheinlich nicht lösbarer Konflikt von „zentralen“ Tests. Beide Wörter sind - bezogen auf das hier in Rede stehende Problem - gleichbedeutend. Beide Wörter gehören zur Fachsprache von Mathematikunterricht. Es gibt Klassen, in denen der Begriff der Deckungsgleichheit der Lernstoff ist. Es gibt auch Klassen, in denen der Kongruenzbegriff der Lernstoff ist. Bei einigen von diesen wird wiederum der Begriff der Deckungsgleichheit zur Erklärung des Kongruenzbegriffs herangezogen - das bezieht sich auf ein gewisses Selbsterklärungspotential des Begriffs der Deckungsgleichheit. Die in der Aufgabe gewählte Formulierung *kongruent (deckungsgleich)* nimmt nun vorrangig diesen letzten Aspekt auf: *kongruent* wird - quasi zur Erinnerung - als *deckungsgleich* erläutert. Gleichzeitig wird *deckungsgleich* als begriffliche Alternative für diejenigen, die den Begriff *kongruent* nicht kennen, angeboten. Für diese Gruppe ist ein neuer Begriff - quasi als Hauptbegriff, weil nicht in der Klammer stehend - aufgetaucht. Für diejenigen, die nur den Begriff kongruent kennen, ist ebenfalls ein neuer Begriff eingeführt, und zwar in einer Klammer. Für alle drei Gruppen entsteht durch die Klammer ein Irritationspotential: Entweder wird man mit einem neuen Begriff konfrontiert, oder es wird plötzlich in der Klammer an die Bedeutung eines Begriff erinnert - für einen Tests ein seltsames Unterfangen. Erklärbar ist diese Begriffsverwirrung für denjenigen, der das - im Kern didaktische - Problem der zwei Begriffe kennt. „Normal“ ist es für denjenigen, der mit solchen Konstruktionen in Tests vertraut ist und sie übergehen kann - ein Bestandteil von Testfähigkeit.

mathematische Anspruchslosigkeit mit verbaler Verkomplizierung verbunden wird. Es wird nicht behauptet, dass dieses Irritationsmoment die Intention der Tester war oder dass es zwingend zu einem Scheitern des Schülers führt, es wird lediglich behauptet, dass es da ist. Damit wird dann allerdings auch die Behauptung verbunden, dass es wirkt und dass ein Teil der Anforderung des Tests darin besteht, diesem Irritationsmoment zu begegnen.

Multiple Choice

Wie bereits in der Betrachtung über mögliche Alternativen zur Fragestellung besprochen, generiert die Verwendung von Multiple Choice im gemeinsamen Zusammenhang von Kongruenzproblematik und konkret gewählten Figuren eine hohe Abgeschlossenheit der Aufgabenstellung. Diese Abgeschlossenheit scheint mit der Trivialität der Anforderung verbunden zu sein.
Die bei den anderen Aufgaben gestellte Frage, ob Multiple Choice hier Rückwärtsarbeiten ermöglicht, erübrigt sich: Die Antwortalternativen müssen einzeln untersucht werden, es kann also nur von den Multiple-Choice-Angeboten ausgehend gearbeitet werden. Da schon an der zweiten Stelle die Falschaussage erscheint, kann der testgeschulte Schüler seine Arbeit bereits an dieser Stelle beenden. Ein testunerfahrener oder ein unsicherer Schüler wird sich auch noch an den beiden verbleibenden Alternativen der Richtigkeit seiner Wahl vergewissern.

Zusammenfassung

Man kann das vorliegende Problem als Problem betrachten, welches die Vertrautheit des Schülers mit dem Kongruenzbegriff testet. Um zur Lösung zu gelangen, benötigt man in diesem Sinne Wissen um den Kongruenzbegriff. Wenn man einen Kongruenzsatz für Dreiecke kennt, so kann man sich der Kongruenz begründeter vergewissern. Man kann aber auch einen Lösungsweg gehen, bei dem eine etwaige Vorstellung davon genügt, was es bedeutet, wenn zwei Dreiecke kongruent sind: sie sollen gleich groß sein. Der Kongruenzbegriff spielt dann kaum noch eine ernstzunehmende Rolle, Kongruenzsätze muß man gar nicht kennen. Diese Generierung einer Lösung aus einer Kombination eines Hauches von Wissen mit klugem Umgang mit dem Gegebenen ist mit dem Begriff Testfähigkeit wohl am besten belegt.
Man kann das Problem auch als Problem betrachten, welches den Schüler herausfordert, effektiv herauszufinden, was „man von ihm will". Für ihn würden dabei verbale und Kombinationsfähigkeiten und wohl auch eine gewisse Unverfrorenheit gegenüber dem mathematischen Impetus der Aufgabe gemessen. In jedem Fall ist auch das, was bei dieser Herangehensweise gemessen wird, eher Testfähigkeit als die Fähigkeit, mit dem Kongruenzbegriff umzugehen.
Schon in dieser ersten Betrachtung der Möglichkeiten der Aufgabenlösung zeigt sich die Logik dieser Aufgabe: Es geht zwar äußerlich um das Testen des Wissens über Kongruenz, man muß aber gar nichts oder wenig über Kongruenz wissen, um die Aufgabe zu lösen. Der mathematische Inhalt wird nicht ernst genommen.
Diese Logik des Nicht-ernst-Nehmens des mathematischen Inhalts wiederholt sich in der Trivialität der zu bestimmenden Kongruenzen und in der Tatsache, dass nach Aussagen über das Quadrat EFGH gefragt wird, obwohl gar keine Aussagen über das Quadrat gemacht werden.

An die Stelle von mathematischem Anspruch tritt die Verkomplizierung der Aufgabe durch verbale Hürden:
Die Hinführung zum Problem erfolgt (gerade im Vergleich zu den anderen untersuchten Testaufgaben) wenig direkt und unvermittelt. Die Aufgabenstellung ist unübersichtlich, die zur Lösung erforder-

liche Reihenfolge (1.Analyse des Problems, 2.Analyse der Aussagen) wird durch die Aufgabenstellung gebrochen. Die im Text gewählte Konstruktion verschafft wenig Orientierung, obwohl sie durch äußerliche Kürze den Anschein erweckt, der Strukturlogik von schneller und klarer Hinführung zum Problem zu folgen.
Durch die Forderung, die falsche Aussage auszuwählen, ist eine Negativaussage zu bearbeiten, was eine gedankliche Umkehrung erfordert. Das erhöht den kognitiven Anspruch der Aufgabe – quasi künstlich.
Das Ungewöhnliche der gedanklichen Umkehrung muß hervorgehoben werden, um eine Verwechslung mit der erwartbaren Anforderung zu vermeiden. Die dazu genutzte Großschreibung des *FALSCH* im Aufgabentext stellt allerdings ein zusätzliches Irritationsmoment dar. Dies gilt aber vorrangig für testunerfahrene Schüler. Der Schüler mit Testerfahrung ist mit diesen Mitteln zur Vereindeutlichung des Textes und Beschleunigung der Aufgabenbearbeitung vertraut. Die sprachliche Verkomplizierung findet in der Formulierung der Antworten eine weitere Fortführung: Hier sind acht verschiedene Dreiecke auseinanderzufummeln, hier wird reine Fleißarbeit verrichtet - mathematisch wird hier wenig vollbracht.
Die Strukturlogik dieser Aufgabe besteht also in der mehrfach auftretenden Verbindung mathematischer Anspruchslosigkeit mit künstlicher Verkomplizierung der Aufgabe durch verbale Kapriolen.

Alternative Aufgabenstellung

Das gewählte Problem ist innerhalb des Multiple-Choice-Formats selbst durch die in der Interpretation aufgezeigten möglichen textlichen Veränderungen nicht sinnvoll zu verarbeiten. Die Aufgabe ist als Testaufgabe ungeeignet.

4.4.6. Die TIMSS-Aufgabe A6

Jan hatte in drei Tests die Ergebnisse 78, 76 und 74 Punkte. Nicole dagegen hatte 72, 82 und 74 Punkte. Wenn man Jans Durchschnitt mit dem von Nicole vergleicht, dann ...

A. liegt Jan 1 Punkt höher.
B. liegt Jan 1 Punkt niedriger.
C. sind beide gleich.
D. liegt Jan 2 Punkte höher.
E. liegt Jan 2 Punkte niedriger.

Lösungswege und Struktur des Problems

- Argumentation mit Gesamtpunktzahlen: Jan hat insgesamt 78+76+74 Punkte, das sind 228 Punkte. Nicole hat insgesamt 72+82+74 Punkte, das sind 228 Punkte. Sie haben gleiche Punktzahlen bei gleicher Testanzahl, also sind beider Durchschnitte gleich.

- Man kann die Argumentation auf die Berechnung des Durchschnitts ausdehnen: Beide haben einen Durchschnitt von 228:3, also von 76.

- Man kann auf die Einbeziehung der 74 Punkte verzichten, weil diese Punktzahl für beide vorliegt.

- Wenn man bedenkt, dass beide die gleiche Anzahl Tests zu bestehen hatten, kann man bequemer rechnen, indem man die jeweils über 70 hinausgehenden Punkte betrachtet. Man könnte sich auch eine andere bequeme Basis suchen.
Jan hat jeweils 8+6+4 über 70 hinausgehende Punkte, das sind 18 Punkte. Nicole hat jeweils 2+12+4 über 70 hinausgehende Punkte, das sind ebenfalls 18 Punkte. Sie haben also die gleiche Punktzahl - durchschnittlich 76 Punkte.

In dieser Aufgabe ist das Berechnen und Vergleichen von Durchschnitten gefordert. Die aufgezeigten Lösungsmöglichkeiten unterscheiden sich bezüglich dieser Forderung nur im Weg bzw. in der Vereinfachung durch Weglassen der 74 bzw. in der Erkenntnis der Tatsache, dass man wegen der gleichen Testanzahl gar nicht den Durchschnitt, sondern nur die Gesamtpunktzahl vergleichen muß.

Interpretation des Aufgabentextes

Zunächst fällt die Selbstreferentialität der Aufgabe auf. Es geht um Tests, und offenbar hat die Bildung des Durchschnitts der Testpunkte irgendeine Bedeutung.

Jan hatte in drei Tests die Ergebnisse 78, 76 und 74 Punkte.
Eine kurze, knappe Einführung in das Problem. Hier gibt es noch keinen Hinweis auf die mathematische Aufgabenstellung. Ein real denkbares Problem wird eingeführt, ohne dass es näher inhaltlich bestimmt würde, z.B. durch einen Hinweis darauf, um was für einen Test es sich handelt. Durch das Präteritum des *hatte* wird die Abgeschlossenheit der Punktvergabe und somit des Rahmens der Aufgabe gekennzeichnet. Die Einschiebung *die Ergebnisse* ist leicht redundant, die Formulierung ist nicht wohlgeformt. Es handelt sich um einen Lesebeschleuniger, die Beschleunigung geschieht durch Fokussierung auf und Vereindeutigung des Folgenden. Die gewählten Zahlen machen das Rechnen im Kopf anspruchsvoll.
Um die Bedeutung der etwas seltsamen Formulierung zu verstehen, kontrastiere ich eine alternative Formulierung und Geschichten:

Jan hatte in drei Test ...	In drei Tests hatte Jan ...
... Glück. (wohlgeformt)	... Glück. (Nur wohlgeformt, wenn neben den drei noch andere Tests gemacht wurden.)
... jeweils die höchste Punktzahl. (nicht wohlgeformt)	... jeweils die höchste Punktzahl. (wohlgeformt)
... die höchste Gesamtpunktzahl. (wohlgeformt)	... die höchste Gesamtpunktzahl. (nicht wohlgeformt)
Diese Formulierung schafft eine Gesamteinheit „drei Tests“	Diese Formulierung konstruiert die drei Tests nicht als Gesamteinheit, sondern als drei separate Entitäten.

Konfrontiert man die Lesart mit der Formulierung ... *die Ergebnisse 78, 76 und 74 Punkte*, so erklärt sich die mangelnde Wohlgeformtheit: Der erste Teil des Satzes konstruiert die drei Tests als Gesamteinheit, der zweite Teil des Satzes verweist aber auf die Ergebnisse der Einzeltests, bezieht sich also nicht auf die im ersten Teil konstruierte Einheit. Wohlgeformt wäre der zweite Satzteil hingegen mit dem alternativen Satzbeginn. Latent wird durch die gewählte Formulierung, also durch die Konstruktion einer Gesamteinheit, die Durchschnittsbildung nahegelegt. Die mangelnde Wohlgeformtheit schafft (geringes) Irritationspotential.

Nicole dagegen hatte 72, 82 und 74 Punkte. Das *dagegen* grenzt Nicoles Punktzahl von der von Jan ab, es legt eine Differenz zum Janschen Ergebnis nahe. In ihm scheint eine leichte Tendenz zu stecken, für Nicole eine geringere Punktzahl als für Jan anzunehmen.[46] Um dies zu untersuchen, betrachte ich folgende kontrastierende Varianten für die Formulierung:

1. Nicole dagegen hatte nur ... Punkte.
2. Nicole dagegen hatte sogar ... Punkte.
3. Nicole hatte ... Punkte.

Diese drei Varianten zeigen lediglich den Abgrenzungscharakter des *dagegen*. In Variante 1 und 2 verstärkt es die abgrenzende Wirkung der wertenden „nur“ und „sogar“. Variante drei ist völlig neutral. Dies hat allerdings in einem Test unter Zeitdruck den Nachteil, die Abgrenzung und damit beschleunigende Klarheit des *dagegen* nicht zu leisten.
Um die wertende Tendenz des *dagegen* zu untersuchen, betrachte ich weitere Formulierungsvarianten:

4. Klaus ist Arzt. Franz dagegen ist Buchhalter.
5. Klaus ist Buchhalter. Franz dagegen ist Arzt.

Wie sehr man es auch hin- und herdreht: Auch hier zeigt sich noch kein wertender Charakter des *dagegen*. Erst wenn man Wertungen der Berufe hinzudenkt und einen entsprechenden Tonfall annimmt, kommt eine Färbung der Aussage hinzu. Der Text gibt eine Wertung nicht her. Das *dagegen* an sich scheint symmetrischen Charakter zu haben. Ich versuche es noch mit einer Variante, die so eine Wertung nicht schon durch ihre Sujets in sich trägt.

6. Klaus kennt Schiller. Franz dagegen kennt Hölderlin.

Auch hier zeigt sich kein Anschein von Wertung durch das *dagegen*. Zu guter letzt verändere ich die Reihenfolge des Aufgabentextes selbst nochmals:

7. Nicole hatte in drei Tests die Ergebnisse 78, 76 und 74 Punkte. Jan dagegen hatte 72, 82 und 74 Punkte.

Auch diese Variante zeigt nur, dass das *dagegen* zwar einen gegeneinander setzenden, jedoch keinen wertenden Charakter hat.

Der Leser, der das Gefühl des wertenden Charakters des *dagegen* von Anfang an nicht teilte, wird jetzt vielleicht am Sinn der eben ausgeführten Operation zweifeln. Die Operation der vertieften Interpreta-

[46] An dieser Stelle fordert die Methode der Objektiven Hermeneutik besondere Sorgfalt in der Interpretation der Textstelle: Es gilt zu untersuchen, ob die objektive Textbedeutung ein Gefühl oder ein „es scheint“ wirklich bestätigt und der Text damit latente Signale zur Lösung der Aufgabe gibt. Dies wäre auch dann bedeutend, wenn das latente Signal der wirklichen Aufgabenlösung widerspricht.

tion einer Textstelle, deren Bedeutung unklar ist, die einen stolpern und zweifeln läßt, gehört aber unbedingt zum methodischen Repertoire der Objektiven Hermeneutik.[47] Hier wird nicht nur der Vorzug der methodischen Kontrollierbarkeit des Interpretationsprozesses deutlich. Man kann nun „interpretationspsychologisch“ fragen, worin der Grund für eine oberflächliche (falsche) Deutung liegt: In diesem Fall nämlich in meinem offenbar vorhandenen Verdacht, dass die Aufgabe dem Mädchen eine niedrigere Punktzahl zuordnen würde. Dieser wurde sicherlich durch die Asymmetrie der Anordnung der Namen hervorgerufen.[48]

Wenn man Jans Durchschnitt mit dem von Nicole vergleicht, dann ... Zur Interpretation dieser Textstelle generiere ich zunächst Alternativen:
Wenn du Jans Durchschnitt mit dem von Nicole vergleichst, dann ...
Wird Jans Durchschnitt mit dem von Nicole verglichen, dann ...
Vergleicht man Jans Durchschnitt mit dem von Nicole, dann ...
Vergleiche Jans Durchschnitt mit dem von Nicole.

Man erkennt zunächst, dass die Formulierung mit *Wenn* einen kleinen Abstand zum bisherigen Text schafft, quasi eine kleine Atempause. Die Aufforderung zum Vergleich erfolgt nicht gleich am Satzanfang. Dies nimmt der Aufgabenstellung ein wenig von ihrer Kompaktheit.
Das *man vergleicht* anstelle des persönlichen „du vergleichst“ oder des absolut unpersönlichen „wird verglichen“ stellt eine mittlere persönliche Ansprache des Schülers dar. Diese Form ist der Testform adäquat.
Die Reihenfolge der Nennung der Namen von Jan und Nicole ist beibehalten, so dass die Aufgabe hier nicht zusätzlich verkompliziert wird.
Erst in diesem Satz wird die eigentliche Aufgabenstellung vorgestellt: Der Vergleich der Durchschnitte. Der Schüler ist nun eigentlich vollständig in das Problem eingeführt. Nur ein testunerfahrender Schüler erwartet jetzt noch - für eine Testaufgabe nicht wohlgeformte - Erweiterungen wie
... freut man sich.
... sieht man, dass Mädchen teststärker sind als Jungen.
... kommt man ins Grübeln.
... wird Jan böse.
Der Text signalisiert durch die drei Punkte, dass jetzt Antwortalternativen auftauchen. Die Aufgabenstellung ist also gegeben, es kann losgehen.

[47] An dieser Stelle stellt sich die Frage, was es für eine Interpretation bedeutet, wenn der Interpret leichtfertig über eine Stelle hinweginterpretiert, in der aber mehr steckt; wenn er z.B. über das dagegen hinweggegangen wäre und es sich aber als bedeutsam erweist. Erstens ist dies ein handwerklicher Fehler, vergleichbar mit Fehlern beim Experimentieren, die vielleicht erst von jemandem entdeckt werden, der das Experiment wiederholt. Zweitens versucht die Objektive Hermeneutik, solche Fehler durch den Falsifizierungszwang zu entdecken. Durch jede fehlgeschlagene Falsifikation, die die Interpretation bestätigt, erhöht sich der Grad der „empirischen Sättigung“ der Interpretation. Obwohl die interpretierten Aufgabentexte kurz sind, bieten sich jeweils mehrere Falsifikationsansatzpunkte. Zeigt sich eine Textlogik in einer Aufgabe mehrfach, so erhöht dies die empirische Sättigung der Interpretation.
[48] Der „Eindruck“ von Asymmetrie wurde offensichtlich dadurch verstärkt, dass in Abweichung von sequentiellem Vorgehen die doppelte Erstnennung des Jungen unreflektiert in die Deutung des *dagegen* einbezogen wurde.

Multiple Choice-Angebote

- Die Reihenfolge der Lösungsangebote folgt der Logik der Reihenfolge des bisherigen Textes, dass der erstgenannte weiter als erster genannt wird. Dies schließt Irritationen aus einer Veränderung der Reihenfolge aus. In dieser Logik liegt es aber auch, dass der zuerst genannte als punktmäßig höher liegend angenommen wird.[49] Diese Logik wird allerdings dadurch gebrochen, dass nicht die Reihenfolge 2, 1, 0, -1, -2 für die Punktedifferenz gewählt wird, sondern 1, -1, 0, 2, -2. Hierdurch wird die Asymmetrie vermindert, welche dadurch zustande kommt, dass Jan an erster Stelle genannt wird: Die Logik der Erstnennung von Jan, die irritationsvermeidend über den gesamten Text durchgehalten wird, führt dazu, dass Jan der Maßstab ist, an dem gemessen wird. Dies führt zu einer Asymmetrie zwischen Jan und Nicole. Diese Asymmetrie würde verstärkt, wenn die Reihenfolge 2, 1, 0, -1, -2 gewählt würde. Die der Asymmetrie latent innewohnende Behauptung, dass Jan mehr Punkte hat, würde bestärkt. Im übrigen kann diese Asymmetrie nur um den Preis von Erschwernis verhindert werden, weil immer einer als erster genannt werden muß: Würde nun die Reihenfolge der Nennung ständig verändert werden, dann wäre die Aufgabe schwieriger. Wird die Reihenfolge beibehalten, folgt zwangsläufig Asymmetrie zwischen der Gewichtung der beiden Personen.

Wie könnte der Text nun weitergeführt werden?
Wenn man Jans Durchschnitt mit dem von Nicole vergleicht, dann ...
... stellt man fest, dass ...
... wäre eindeutig klar, dass ...
... sieht man einen Vorsprung von ...
... wird deutlich, dass ...
Die sich aus den Geschichten herauskristallisierende Lesart ist, dass hier eine das Resultat des Vergleichs erfassende Tätigkeit stattfindet: Wenn man die Durchschnitte vergleicht, dann gelangt man zu einer Erkenntnis. Vergleicht man dies mit dem wirklichen Text, so fällt eine seltsame sprachliche Verwerfung ins Auge:

A. liegt Jan 1 Punkt höher.
B. liegt Jan 1 Punkt niedriger.
C. sind beide gleich.
D. liegt Jan 2 Punkte höher.
E. liegt Jan 2 Punkte niedriger.

Die vergleichende Tätigkeit wird ausgespart, es wird sofort das Resultat präsentiert. Dadurch wird sprachlich konstruiert, dass die Resultate nur dann unverändert existieren, wenn man die Durchschnitte vergleicht: *Wenn man Jans Durchschnitt mit dem von Nicole vergleicht, dann sind beide gleich.* Wenn man die Durchschnitte nicht vergleicht, dann ist Jan vielleicht besser oder schlechter. In der wohlgeformten Variante ist es so, dass man lediglich nichts feststellt, wenn man nicht vergleicht.

[49] Im Alltag findet sich dieser Gedanke im Spruch „Der Esel nennt sich immer zuerst.“ wieder. Dabei soll Höflichkeit über sinnvolles Verhalten gestellt werden: Im Konkurrenzkampf ist in Wirklichkeit ja derjenige der Esel, der sich nicht zuerst nennt bzw. nicht zuerst genannt wird.

Man kann das nun als sprachlichen Fehler bewerten und als Irritationspotential zur Seite legen. Textlich wird hier aber eine Realität konstruiert, die der positivistischen Herangehensweise eines Tests erstaunlich adäquat ist: Wesen und Erscheinung werden einander gleichgesetzt. Jans und Nicoles Durchschnitte sind gar nicht „beide gleich“, sondern sie sind es nur, insofern sie verglichen werden. Es gibt gar keine Wirklichkeit, solange sie nicht messend verglichen ist. Dies ist umso bizzarer, als ja der Testprozeß bereits stattgefunden hat, sie aber gleiche Punktzahl in dieser Konstruktion doch erst haben, wenn der Vergleich stattgefunden hat.

Die Selbstreferentialität des Textes bezüglich der eigenen Methode taucht ein weiteres Mal in der Formulierung „sind beide gleich“ auf. In den anderen Formulierungen liegt Jan Punkte höher oder niedriger (als Nicole - eine weitere Textverkürzung). Beim Punktegleichstand steht nicht *haben beide dieselbe Punktzahl, haben beide denselben Durchschnitt, liegen beide gleichauf, ist der Durchschnitt derselbe, ist die Punktanzahl dieselbe, besteht Punktgleichheit* usw. Hier „sind beide gleich“. Wörtlich genommen ist das absurd, denn zwei Personen sind ja nicht beide gleich, nur weil sie den gleichen Punktdurchschnitt in drei Tests haben. Der Text konstruiert aber aus der Durchschnittsgleichheit die Gleichheit der Personen.

Wird durch Multiple Choice Rückwärtsarbeiten möglich?

Rückwärtsarbeiten wird hier durch Multiple Choice nicht ermöglicht. Allerdings liegt mit den angebotenen Alternativen eine Kontrollmöglichkeit vor, falls man sich extrem verrechnet hat oder sich der Aufgabe völlig falsch genähert hat: Die geringen angebotenen Punktedifferenzen ermöglichen, völlig unsinnige Rechenprozeduren zu erkennen. Sollte man sich hingegen beim Zusammenzählen der Punkte um Eins oder Zwei verrechnen, so hilft die Durchschnittsbildung, den Fehler zu erkennen, weil dann gebrochene Zahlen als Durchschnitt herauskommen, die als Lösungen nicht angeboten sind.

Insgesamt bietet Multiple Choice also lediglich einige wenige Kontrollmöglichkeiten bezüglich Rechenfehlern.

Zusammenfassung

In dieser Aufgabe ist das Berechnen und Vergleichen von Durchschnitten gefordert. Dabei machen die gewählten Zahlen das Rechnen im Kopf anspruchsvoll.

In das real denkbare, inhaltlich nicht näher bestimmte, zeitlich abgeschlossene Problem wird kurz und knapp eingeführt. Die Nennung von Jan an erster Stelle beginnt hier und wird bis zum Ende der Aufgabe durchgehalten. Dies führt zu einer Asymmetrie der beiden Personen. Eine nähere Untersuchung zeigte aber, dass diese Asymmetrie durch den sonstigen Text nicht aufgenommen wird. So grenzt das *dagegen* lediglich Nicoles Punktzahl von der von Jan ab, es legt eine Differenz zum Janschen Ergebnis nahe. Hier wird vermieden, dass das Ergebnis der Punktgleichheit schon latent mitgeteilt wird. Gleichzeitig ist die Anordnung der Antwortalternativen so gewählt, dass die nur um den Preis von Inkonsistenz vermeidbare Asymmetrie von Jan und Nicole nur im unvermeidbaren Umfang latent eine höhere Punktzahl von Jan nahelegt.

Innerhalb der kompakten Aufgabenstellung ist mit dem Satzanfang *Wenn man* ... eine kleine Atempause für den Schüler eingefügt. Die Aufforderung zum Vergleich erfolgt nicht gleich am Satzanfang. Die mittlere persönliche Ansprache des Schülers ist der Testform adäquat.
Multiple Choice ermöglicht bei dieser Aufgabe kein Rückwärtsarbeiten. Allerdings liegt mit den angebotenen Alternativen eine Kontrollmöglichkeit vor, falls man sich extrem verrechnet hat oder sich der Aufgabe völlig falsch genähert hat.
Alles in allem liegt mit dieser Aufgabe ein gelungener Fall einer Testaufgabe im vorgegebenen Format vor. Die unsinnige positivistische Weltkonstruktion ist ein Habitusproblem und beeinflußt die Meßeigenschaften der Aufgabe nur marginal.

4.4.7. Zusammenfassung: Der TIMSS-Test

Es wurden die sechs „Kernaufgaben“ des TIMSS-Tests untersucht. Das sind jene sechs Mathematikaufgaben, die in allen Testheften zur Anwendung kommen. Nur diese sechs Aufgaben mußten also von allen getesteten Schülern bearbeitet werden.
Mit der Aufgabe A6 liegt der einzige Gelingensfall einer Testaufgabe vor. Hier kann eindeutig benannt werden, was diese Aufgabe mißt, gleichzeitig ist das Gemessene das, was gemessen werden soll, und der Meßvorgang wird nicht durch Aspekte von Testfähigkeit gebrochen.
In der Aufgabe A1 kann ebenfalls eindeutig benannt werden, was gemessen wird, und es wird gemessen, was gemessen werden soll. Der Meßvorgang wird allerdings durch das Mitmessen von Testfähigkeit gebrochen. Sie besteht hier allerdings lediglich im Überwinden künstlicher Irritationen. Diese entstehen durch die Wortwahl „schattieren“ sowie durch sprachliche Beschleuniger und Vereindeutigungen (kleine Quadrate, ZUSÄTZLICH), welche zwar manifest beschleunigen und vereindeutigen, latent aber Irritationspotential schaffen.
In Aufgabe A2 ist der Meßprozeß direkt gestört: Ein Schüler, der „zu klug“ ist, das Problem also zu ernst nimmt und nicht erkennt, dass er hier eine schlichtere Denkweise bedienen muß, gelangt zur „falschen“ Lösung 3 kg. Verursacht wird das durch die Gestaltung des Bildes, welches insgesamt erhebliches Irritationspotential trägt. Ein derart massiver Konstruktionsfehler macht die Aufgabe als Testaufgabe unnutzbar.
Im Text liegen massive Verwerfungen vor, die daraus resultieren, dass ein alltägliches, ein physikalisches und ein mathematisches Problem miteinander verworfen werden statt zu einem schlüssigen Ganzen zu verschmelzen. Die Authentizität des Realen wird dabei ebenso zerstört wie die Authentizität des Fachlichen. Diese Zerstörung wiederholt sich auf der Ebene von Alltags- und Fachsprachlichkeit. Es liegt eine Struktur von äußerlicher Verfachsprachlichung bei gleichzeitiger Dementierung von Fachlichkeit vor.
Es wird massiv Testfähigkeit mitgetestet. Dies betrifft sowohl das Überwinden erheblicher Irritationen und Verwerfungen als auch das Überwinden von Fallen (Verteilung der Körper auf der Waage, Halber/ganzer Ziegelstein)

In Aufgabe A3 wird die Anwendung der schulmathematischen Rundungsregeln getestet, und zwar in „umgekehrter“ Weise: Es wird nach dem „tatsächlichen“ Wert einer gerundeten Meßgröße gefragt. Das Besondere dabei ist, dass ohne Multiple Choice ein ganzes Intervall Lösung wäre. Multiple Choice beeinflußt den Charakter der Aufgabe also prinzipiell, und es bleibt unklar, ob *eigentlich* die Kenntnis der Rundungsregeln gemessen werden *soll*. Man darf bzw. muß nicht eine Zahl aus dem Rundungsintervall wählen, sondern man muß entscheiden, welche der vorgeschlagenen Zahlen im Intervall liegt – was sprachlich nicht umgesetzt ist. Da nur ein Wert angekreuzt werden kann, wird nicht erfaßt, ob der Schüler das Gesamtproblem des Lösungsintervalls erfaßt hat.

Der Schüler wird dabei in ein mathematisches Schrumpfuniversum gepreßt. Das korrespondiert mit der Verwerfung von Fachsprachlichkeit und Alltagssprachlichkeit und mit einer Fragestellung, die nicht die zu beantwortende Frage stellt, weil unendlich viele Antwortmöglichkeiten auf die Frage durch die Multiple-Choice-Auswahl ausgeschlossen werden.

In Aufgabe A4 wird ein Problem gestellt, welches nicht wirklich eindeutig zu lösen ist, weil verschiedene Deutungen der Situation möglich sind. Nur die Kenntnis der Strukturiertheit von Schulaufgaben ermöglicht es hier, relativ eindeutig eine Entscheidung darüber zu treffen, was die Tester für richtig halten. Dieses Problem wird dadurch abgeschwächt, dass die latente Sinnstruktur der Aufgabenstellung stark auf Proportionalität verweist, weniger stark auf ein Gleichbleiben des Abstands zwischen den Läuferinnen. Trotzdem bleibt die Aufgabe als Testaufgabe unnutzbar.

In dieser Aufgabe liefert Multiple Choice die Möglichkeit, die Aufgabe auch ohne Verständnis für Proportionalität bzw. Verhältnisse zu lösen. Dies kann über Plausibilitätsbetrachtungen geschehen, welche lediglich ein alltägliches Vorverständnis von Proportionalität und Cleverneß im Umgang mit Testaufgaben erfordern. Mit diesen Fähigkeiten ist auch die Ratewahrscheinlichkeit auf fünfzig Prozent zu erhöhen. Es liegt also die Fehlkonstruktion einer Testaufgabe vor: Einerseits kann ein Schüler scheitern, der über die zu messende Fähigkeit verfügt, andererseits kann ein Schüler zum Erfolg gelangen, der nicht über die zu messende Fähigkeit verfügt.

In der Aufgabe A5 soll äußerlich das Wissen um Kongruenz getestet werden. Man muß aber nichts oder wenig über Kongruenz wissen, um die Aufgabe lösen zu können. Allerdings muß man klug mit dem Gegebenen umgehen können, denn die Aufgabe ist kognitiv und konstruktiv verkompliziert: Die an sich triviale Aufgabe ist umgedreht (die falsche Aussage muß herausgefunden werden) und man muß sich durch acht verschiedene Dreiecke durcharbeiten, um rauszufinden, worum es geht. Diese Verkomplizierung ist begleitet durch verschiedenste Irritationen (FALSCH, kongruent (deckungsgleich), Aussage über das Quadrat). Die Strukturlogik dieser Aufgabe besteht in der mehrfach auftretenden Verbindung mathematischer Anspruchslosigkeit mit verbalen Kapriolen zur künstlichen Verkomplizierung.

In den sechs „Kernaufgaben“ des TIMSS-Tests findet sich also mit der Aufgabe A6 ein Gelingensfall. Die Aufgabe A1 muß lediglich von Komponenten von Testfähigkeit befreit werden. Die Aufgabe A5 läßt zu, dass man sie auch lösen kann, ohne über die zu testenden Fähigkeiten zu verfügen, sie und die

Aufgabe A3 sind für das Multiple-Choice-Format ungeeignet. Die Aufgaben A2 und A4 sind als Testaufgaben unnutzbar, da sie – neben anderen Probleme – Fallen stellen. Das kann dazu führen, dass selbst ein Schüler, der über die zu testende Fähigkeit verfügt, scheitert. In A4 kommt zusätzlich hinzu, dass die Aufgabe gleichzeitig gelöst werden kann, ohne über die zu testende Fähigkeit zu verfügen.

Für den TIMSS-Test muß also konstatiert werden, dass er als mathematischer Leistungstest nicht geeignet ist: Die aufgezeigten Probleme lassen sich nicht als Einzelfälle deuten. Eher deutet sich der Gelingensfall als Einzelfall an.
Es zeigen sich verschiedene Charaktere und Ursachen der aufgezeigten Probleme. Durch das Vorgehen der Objektiven Hermeneutik nachvollziehbar und methodisch kontrollierbar aufzuzeigen sind jene Probleme, die durch Verwerfungen von manifestem und latentem Aufgabentext zustande kommen. Diese Probleme sorgen für Mehrdeutigkeiten im Meßprozeß und für das Mitmessen verschiedener Dimensionen von Testfähigkeit. - Diese Themenbereiche werden im Schlußkapitel in Zusammenführung der Erkenntnisse aus den TIMSS- und den PISA-Interpretationen diskutiert.

Für das Aufzeigen des problematischen Bildes von Mathematik, in dem der TIMSS-Test arbeitet, war Objektive Hermeneutik hilfreich, weil mit ihrer Hilfe Zerstörungen des Mathematischen und das Gegeneinanderlaufen von Fachsprachlichem und Nichtfachsprachlichem expliziert werden konnten. Ein Test aber, in dem das Mathematische unterminiert wird, kann kein mathematischer Leistungstest sein. Dass diese Unterminierungen kein Privileg mathematischer Laien ist, wird sich im PISA-Kapitel zeigen: PISA wurde - im Gegensatz zu TIMSS - von Mathematikdidaktikern erstellt und bringt vergleichbare Unterminierungen auf, die im PISA-Kapitel unter dem Stichwort eines mathematikdidaktischen Habitus vertieft erörtert werden.

Als TIMSS-Spezifikum erweist sich eine gewisse Hölzernheit der Aufgabentexte, deren Charakter in den Interpretationen expliziert wurde: Die Hölzernheit erwächst aus der Konzentration auf den Testvorgang selbst statt auf die zu testende Fähigkeit bzw. den zu testenden Inhalt. Der Schüler wird geradezu durch den Test „getrieben". Es gibt sehr viele Textelemente, die ausschließlich der schnelleren Bearbeitung bzw. der Vereindeutigung der Aufgabenstellung dienen - offenbar ein unüberschaubares Potential für die Produktion von Verwerfungen. Hier liegt mit der Methode der Objektiven Hermeneutik ein Hilfsmittel vor: Diese Verwerfungen entstehen, weil etwas anderes geheilt werden soll - nämlich vorhandene oder angenommene Uneindeutigkeiten (bei Vereindeutigungen) bzw. begrenzte Informationsverarbeitungspotentiale (bei Lese- bzw. Testbeschleunigern). Objektiv-hermeneutische Aufgabeninterpretationen können Testern helfen zu verhindern, dass sie ihre Probleme nur reproduzieren, wenn sie sie zu heilen versuchen. Allerdings deutet sich an, dass einige ihrer Probleme nicht zu heilen sind. - Informationsverarbeitungskapazitäten sind nun mal begrenzt, und sprachliche Eindeutigkeit benötigt oftmals eher mehr als weniger Worte.

Zwei wesentliche Erkenntnisse der Interpretationen wären auch ohne Objektive Hermeneutik generierbar gewesen:

i. Der Meßprozeß bei TIMSS ist ein unscharfer. Die Aufgaben sind auf verschiedene Weise zu lösen. Das Kreuz im Testheft erzählt uns, dass der Schüler das Kreuz richtig gesetzt hat. Die Deutung dessen, welche Fähigkeiten ihn dazu befähigt haben, ist eine unscharfe. Gleiches gilt für die Diagnose von Schülerfehlern. Die objektiv-hermeneutische Interpretation gibt uns Hinweise darauf, welche Lösungswege oder welche Fehlwege gestärkt oder geschwächt werden - die Unschärfe, die uns die verschiedenen Lösungswege zeigen, kann sie aber nicht heilen.

ii. Die Tester scheinen sich nicht die Frage zu stellen: „Welche Fähigkeit ist vonnöten, um diese Aufgabe zu lösen?“ Sie scheinen sich zu fragen: „Mit Hilfe welcher Fähigkeit soll der Schüler die Aufgabe lösen?“ Sie konzentrieren sich also auf ihre eigenen Wünsche an den Schüler bzw. an die Aufgabe statt auf den Charakter der Aufgabe. Das Problem wurde bereits in Kapitel 1 diskutiert und zeigt sich auch in den Interpretationen. So bleibt bei A3 und A5 unklar, was getestet werden soll, so entstehen die Probleme bei A2 und A4. Ich nehme an, dass dieses Problem entstanden ist, weil keine Operationalisierung eines Fähigkeitskonstrukts stattgefunden hat. Die TIMSS-Gruppe scheint dem Glauben zu unterliegen, dass sich durch die Versammlung von Inhaltskategorien quasi von selbst ein sinnvolles Testkonstrukt ergibt. Das wäre der Fall, wenn sich die Unschärfen sinnvoll und nachvollziehbar überlagern würden. Dafür gibt es aber keinen Anhaltspunkt.

Kapitel 5: Der PISA-Test

In den Aufgabeninterpretationen der vorhergehenden Kapitel wurden Fragen beantwortet, die sich anhand theoretischer Überlegungen zu Testaufgaben und aus methodologischen Problemkonstellationen ergeben hatten. Mit PISA liegt nun ein Test vor, der auf fachdidaktischer Grundlage erarbeitet wurde. Diese theoretische Grundlage ist verbunden mit einen normativen Anspruch - das PISA-Konsortium spricht sogar von „inhaltlichem Benchmarking" (PISA 2000, S.19)[50]. An diesem normativen Anspruch sollten auch die Testitems gemessen werden können: Einerseits müßten die zu messenden Fähigkeiten beschrieben und in ihrer Bedeutung für das Meßziel begründet werden. Andererseits müßten Operationalisierungen von diesen Fähigkeiten auf Testitems hin beschrieben werden. Es müßte also nachvollzogen werden können, welche Items in welcher Kombination welche Fähigkeit messen sollen. Die Aufgabe einer Iteminterpretation liegt dann darin zu überprüfen, ob die zu messenden Fähigkeiten durch die Items bzw. ihre Kombination wirklich gemessen werden. Die Interpretation der Aufgaben ist damit ein wesentlicher Teil der Überprüfung der Geltung der Testaussage.
Dabei ist wegen des normativen Anspruchs implizit immer auch die Brauchbarkeit des Items im Prozeß des Erlangens bzw. zum Training dieser Fähigkeit thematisch. Bei PISA wird allerdings bereits in der Grundkonstruktion des Tests der Charakter von Aufgaben als didaktische Instrumente mit dem Charakter von Aufgaben als Meßinstrument vermengt. Dies scheint „subjektiv" darin begründet zu sein, dass die beteiligten Didaktiker den Test nicht nur als Meßinstrument sehen, sondern als Instrument zur didaktischen Instruktion von Lehrern[51]. Durch diese Vermengung ist bei PISA-Aufgaben von vornherein mit Verwerfungen zu rechnen. Die Interpretationen müssen das Vorhandensein, den Charakter und die Konstruktion dieser Verwerfungen näher beschreiben.

Theoretisch ideal wäre, wenn man den Operationalisierungsvorgang der PISA-Gruppe objektiv-hermeneutisch nachvollziehen und überprüfen könnte. Dazu müßte ein Katalog vorliegen, der beschreibt, welches Item (warum) welche Fähigkeit messen soll. Da wir es bei mathematischen Leistungstests mit komplexen Fähigkeitskonstrukten zu tun haben, würden oftmals mehrere Items zum Testen einer bestimmten Fähigkeit notwendig sein und die meisten Items würden mehrere Fähigkeiten gleichzeitig messen - eine Operationalisierung müßte diese Mehrfachmessungen miteinander zu einem Gesamtmeßkonstrukt verzahnen. Wir haben solche Mehrfachmessungen bereits in den bisherigen In-

[50] Die deutsche PISA-Gruppe beschreibt als „Zielorientierung von PISA, die Qualität der mathematischen Fähigkeiten gegen Ende der Pflichtschulzeit zu beschreiben und aus den Testresultaten schließlich positive Anregungen für eine Weiterentwicklung des Mathematikunterrichts zu gewinnen." (Neubrand u.a. 2001, S.46) Zwar bleibt unklar, wie aus Testresultaten „positive Anregungen" gewonnen werden, aber immerhin wird hier textlich klar, dass „positive Testresultate" von vornherein nicht vorgesehen waren.

[51] Das Fehlen des Bewußtseins für diesen Unterschied trägt wesentlich zu der Gefahr der technokratischen Verkürzung von Lehren durch verstärktes Testen bei: Wenn die Tester behaupten, in ihrem Test würde ein normativ hochzuhaltendes didaktisches Konzept in eine Aufgabensammlung gerinnen, dann ist Testschulung für den Lehrer nicht mehr nur ein notwendiges Übel, um seine Schüler auf zukunftsentscheidende Testsituationen vorzubereiten. Den Beteiligten wird dann zusätzlich suggeriert, hier würde „didaktisch wertvoll" unterrichtet, ohne dass die technokratische Verkürzung noch thematisch ist. Wenn man diese Vorstellung zu Ende denkt, dann muß der Lehrer eigentlich nur noch intensiv genug Tests trainieren, und die normativ geforderten Fähigkeiten werden sich quasi von selbst einstellen.

terpretationen gesehen, und dort wurde auch bereits gezeigt, wie man diese verschiedenen Fähigkeiten, die ein Item mißt, herausarbeiten bzw. durch Textveränderung manipulieren kann.

Die Resultate der PISA-Aufgabeninterpretationen erfordern es, in diesem Kapitel einen neuen Interpretationsfokus einzuführen: Ursprünglich ging es in dieser Arbeit lediglich darum, die Eignung von TIMSS und PISA als Meßinstrumente zu untersuchen. Bereits im TIMSS-Kapitel wurde dabei das von den Aufgaben vermittelte Bild von Mathematik betrachtet. Dieses Bild berührt den Charakter als Meßinstrument nur am Rande, es erzählt vielmehr etwas über den Aufgabenersteller. Das latente Mathematikbild der TIMSS-Gruppe ist aber kein Schwerpunkt einer mathematikdidaktischen Analyse. Anders verhält es sich bei PISA: An der Erstellung von PISA waren direkt und indirekt anerkannte Vertreter der deutschen Mathematikdidaktik beteiligt. Es handelt sich bei PISA offenbar um mathematikdidaktische Reflexion auf höchstem Niveau. Auch in der Rezeption von PISA sind kaum kritische Stimmen zu den Aufgaben zu vernehmen, was auf eine verbreitete Zustimmung hinweist. Gerade die veröffentlichten internationalen Aufgaben werden als „vorbildliche“ und produktive Aufgaben (zumindest für den Unterricht) wahrgenommen. Es handelt sich bei den PISA-Aufgaben also zweifellos um exponierte Protokolle eines mathematikdidaktischen Habitus. In den PISA-Aufgaben sind nun zwei Charakteristika dieses Habitus sichtbar: Erstens zeigt sich in mehreren Aufgaben eine Zerstörung des Mathematischen bei gleichzeitiger Überbetonung des Fachsprachlichen. Zweitens mißlingt die Annäherung an das Reale. Letzteres erweist sich als unmittelbar mit dem Konzept der „Mathematical Literacy“ zusammenhängend, welches PISA zugrunde liegt.
Diese Habituselemente entfalten sich in den Interpretationen so deutlich und ohne dass nach ihnen gesucht wurde, dass sie nicht ignoriert werden konnten. Das führt zu der eigentümlichen Situation, dass hier nicht nur eine wissenschaftliche Debatte mit der PISA-Gruppe geführt wird, sondern gleichzeitig die PISA-Gruppe quasi ein Proband ist, anhand von dessen Handlungsprotokollen ein Habitus rekonstruiert wird. Diese Doppelung läßt sich aber nicht vermeiden. Es wird zu diskutieren sein, inwieweit dieser mathematikdidaktische Habitus auch ein pädagogischer Habitus ist.

In den Interpretationen läuft die Untersuchung der Meßeigenschaften der Aufgabe parallel zur Rekonstruktion des in den Aufgaben geronnenen mathematikdidaktischen Habitus´. Das liegt daran, dass sich beide Untersuchungsfokusse auf denselben Text beziehen.

5.1. Theoretische Grundlagen von PISA und ihre Verwendung bei der Analyse der Testaufgaben

Im Zusammenhang mit PISA wird man mit einer Vielfalt von Theorien, theoretischen Versatzstücken und Ansätzen, deren theoretischen Status man schwer beschreiben kann, konfrontiert: Man trifft unter anderem auf das Reussersche Modell der mentalen Situationsmodelle, auf das Kompetenzstufenmodell, auf den Ansatz zur Beschreibung mathematischer Modellierungen („Modellierungskreislauf“) und auf den Katalog der Testaufgaben.

- Bereits im Abschnitt 1.2. wurde für das Reussersche Modell gezeigt, dass die TIMSS-Gruppe dieses Modell zwar als theoretische Grundlage vorschiebt, aber keineswegs damit arbeitet. Bei PISA wiederholt sich das: Zwar wird der kognitive Ansatz als theoretische Grundlage für PISA angegeben (siehe PISA 2000, S.71 ff., 144 f.), er wird aber bei der Interpretation der Aufgaben nicht herangezogen, diese erfolgt über unsystematische Alltagsdeutungen (siehe z.B. PISA 2000, S.147-154; Knoche u.a. 2002, S.174-180).
- Das Kompetenzstufenmodell ist speziell für PISA ausgearbeitet worden. Zur Konstruktion dieses Modells werden die Aufgaben nach ihren Lösungshäufigkeiten geordnet und in fünf Klassen unterteilt. Diese fünf Klassen werden als Kompetenzstufen bezeichnet. Dann werden die Aufgaben in einer Kompetenzstufe daraufhin untersucht, welche Kompetenzen zu ihrer Lösung benötigt werden. Diese Kompetenzen beschreiben nun das Anforderungsniveau bzw. den Inhalt dieser Kompetenzstufe.
Eine Argumentation, dass dieses Modell theoretisch hinfällig ist, habe ich in meinem Artikel „Das Kompetenzstufenmodell von PISA – eine empirische Dekonstruktion" (Meyerhöfer 2004) geführt. Die Argumentationslinie ist dabei, dass es nicht möglich ist, mit den PISA-Aufgaben die Stufen inhaltlich zu beschreiben. Das liegt daran, dass die PISA-Aufgaben viele Möglichkeiten verschiedener Lösungswege mit sehr weit auseinander liegenden Ansprüchen bieten. Man kann dadurch keine Aussage darüber treffen, welche Kompetenzen *wirklich* benötigt werden, um die Aufgabe zu lösen. Mit den vorliegenden inhaltlichen Beschreibungen der Kompetenzstufen ist z.B. eine Aufgabe mit Kompetenzen aus Stufe 1 zu lösen, aber auch mit Kompetenzen aus Stufe 2 oder Stufe 3 oder 4.
- Die PISA-Gruppe verändert die herkömmlichen mathematikdidaktischen Modellierungstheorien hin zu einer Modellierungstheorie ohne Modellierung. Das wird in Abschnitt 5.2. beschrieben. Die Ersetzung einer Operationalisierung durch die Erstellung eines Testaufgabenkatalogs beschreibt Abschnitt 5.3.

Nun liegt das Anliegen dieser Arbeit nicht in der Untersuchung des kulturindustriellen Umgangs mit Wissenschaft. Hier sollen die theoretischen Konstrukte rund um den Test nur insoweit betrachtet werden, als sie zur Beantwortung der Frage beitragen, was der Test mißt. Im kulturindustriellen Prozeß der Testerstellung, -durchführung und -auswertung unterliegen alle theoretischen Elemente einerseits der Dynamik der wissenschaftlichen Debatte, andererseits der oftmals kurzatmigen Dynamik von Reparatur und Resultatproduktion. Wenn sich z.B. der erste Teil der Aufgabe „Fläche eines Kontinents" (siehe 5.9.)[52] als unlösbar erwiesen hat, so mußte dies in der statistischen Verarbeitung berücksichtigt werden, ebenso mußte aber entschieden werden, ob der zweite Teil der Aufgabe noch gewertet werden kann. Dies hätte eigentlich eine eigene, neue Untersuchung erfordert. Der Reparaturmechanismus der PISA-Gruppe bestand lediglich darin, die erste Aufgabe zu streichen und die zweite unkommentiert weiter zu verwenden. In den „Beispielaufgaben aus dem Mathematiktest" (OECD 2001) ist die erste Aufgabe noch aufgeführt. In den „Beispielaufgaben Mathematik aus der nationalen Ergänzung zu

[52] In dieser Arbeit werden ausschließlich Aufgaben präsentiert, welche von der PISA-Gruppe veröffentlicht wurden. Alle veröffentlichten internationalen Aufgaben (Fläche eines Kontinents, Äpfel, Bauernhöfe, Dreiecke, Behälter, Geschwindigkeit eines Rennwagens) finden sich in OECD (2001) sowie in dieser Arbeit und (bis auf die letzten beiden Aufgaben) in PISA 2000. Einige nationale Aufgaben wurden zunächst in PISA 2000 und Deutsches PISA-Konsortium (2002) veröffentlicht. Eine vollständige Aufstellung der veröffentlichten nationalen Aufgaben findet sich in OECD (2003).

PISA in Deutschland“ (OECD 2003), die auch internationale Aufgaben enthält, ist nur noch die zweite Aufgabe verzeichnet.

In den folgenden Darlegungen findet sich diese Dynamik wieder und wirkt vielleicht zunächst verwirrend: So stellt sich heraus, dass das PISA-Basiskonzept „Realistic Mathematical Education“ keineswegs auf Realität zielt. Die Testkonstruktion erweist sich als eine Art „Zusammenstückeln“ von halbwegs praktikablen Aufgabenkategorien und die Kategoriensysteme ändern sich gelegentlich. Auch das Kompetenzstufenmodell wurde im Nachhinein verfeinert, allerdings ohne dabei das Problem zu lösen, dass man die dort konstruierten Kompetenzstufen inhaltlich nicht beschreiben kann - die Reparatur verschärft das Problem sogar noch.
Es entstehen auf diese Weise „Theorieversionen“, deren Lebensdauer man nach Monaten bemessen kann. Es ist kaum davon auszugehen, daß diese Dynamik bereits ihren Endpunkt erreicht hat, das verringert aber nicht den Wert eines kritischen Hinterfragens der bisherigen Theorieelemente. Die im Folgenden aufgezeigten Probleme erweisen sich als so gravierend, dass sie mit einfachen Reparaturen nicht zu bewältigen sind. Im gemeinsamen Kern laufen die Probleme darauf hinaus, dass nicht genug darüber reflektiert wird, wie (und wo) die Geltung der Testaussagen erzeugt wird. Der in dieser Arbeit im Mittelpunkt stehende Ort der Geltungserzeugung ist die Aufgabe selbst, und damit die Operationalisierung eines (bei PISA nicht wirklich vorhandenen) Konstrukts von „Mathematischer Leistungsfähigkeit“.

„Mathematical Literacy” und Realistic Mathematics Education

Die mathematischen Tests von PISA wurden unter Berufung auf eine fachdidaktische Ausrichtung erarbeitet (PISA 2000, S.139). Dem *internationalen* Test liegt das Konzept der *Mathematical Literacy* zugrunde:

> „PISA richtet sich dezidiert *nicht* auf die Beherrschung von mathematischen Verfahren und Faktenwissen und auch nicht auf ein nur schematisches Anwenden von Mathematik zur Lösung „eingekleideter“ Aufgabenstellungen. Vielmehr untersucht der PISA-Mathematiktest, inwieweit mathematisches Wissen funktional, flexibel und mit Einsicht zur Bearbeitung vielfältiger kontextbezogener Probleme eingesetzt werden kann (...). Diese Fähigkeit wird als *Mathematical Literacy* bezeichnet.“ (ebenda)
>
> Der Begriff der Mathematical Literacy wird als Zuordnung eines Fähigkeitskonzept zum Freudenthalschen didaktischen Konzept der „Realistic Mathematics Education“ beschrieben. (siehe ebenda S.141-143)

Die Frage, was PISA mißt, beinhaltet also die Frage, ob PISA wirklich Mathematical Literacy mißt. Umgekehrt wird das theoretische Konstrukt „Mathematical Literacy“ immer auch durch seine Konkretisierung in Aufgaben mitkonstruiert. Die Interpretationen erzählen uns also auch etwas darüber, was Mathematical Literacy ist.

Die PISA-Gruppe gibt als Charakteristikum für den *Realistic Mathematics Education*-Ansatz an, „dass eine außermathematische Situation (Foto) sogleich durch eine schematische Zeichnung ergänzt wird, sodass beides, außer- und innermathematische Zusammenhänge, gleichzeitig angezeigt sind.“ (PISA 2000, S. 151)

Die Aussage, dass die außermathematische Situation durch die schematische Zeichnung *ergänzt* wird, erzeugt durch ihre Mehrdeutigkeit theoretische Fragen: Eine Ergänzung ist ein Akt des Nebeneinanderstellens zweier Sujets, die gleichwohl miteinander verbunden sind. Eine Aussage über *die Art der Verbindung* wird hier aber vermieden - das „gleichzeitige Anzeigen“ von inner- und außermathematischen Zusammenhängen bedeutet nicht notwendig, dass irgendeine Verbindung zwischen ihnen gezogen wird. Ein normatives Ideal, an dem eine Aufgabe gemessen werden kann, die sich diesem Ansatz verschreibt, kann daraus nicht abgeleitet werden. Eine Interpretation wird aber untersuchen müssen, wie Inner- und Außermathematisches miteinander verbunden wird: Ist es eine produktive Verbindung oder lediglich ein Nebeneinander? Wird der Gefahr begegnet, durch die gleichzeitige Präsentation von inner- und außermathematischem Aufgabeninhalt das Außermathematische von vornherein abzustreifen? Aus den Aufgaben selbst wird zu erschließen sein, wie eine produktive Verbindung aussehen kann - PISA selbst hat kein Konzept für so eine produktive Verbindung. Es werden aber Stichworte gegeben, an denen sich Freudenthals didaktischer Ansatz und damit das Konzept der Mathematical Literacy orientieren: Mathematik als System begrifflicher Werkzeuge, mit dem sich Schülerinnen und Schüler Phänomene ihrer natürlichen, technischen, geistigen und sozialen Umwelt erschließen können, Beziehungshaltigkeit, Orientierung an der Welt, aber kein instrumentelles Verständnis, sondern Ausbildung mentaler Modelle für mathematische Begriffe. (ebd., S. 142 f.) Man bemerkt aber sofort, dass diese Stichworte *habituell* etwas völlig anderes vermitteln als das oben beschriebene „gleichzeitige Anzeigen“ einer außermathematischen Situation und einer schematischen Zeichnung. *Habituell* hat sich das Konzept der Realistic Mathematics Education offenbar weit von Freudenthal entfernt.

Nun kann man von den Freudenthalschen Grundideen ausgehend einerseits Aufgaben entwickeln, die einen Lernprozeß begleiten. Die Aufgabenfolgen in der niederländischen Konzeption von Realistic Mathematics Education verstehen sich als eine solche Konkretisierung: „Diese Aufgaben gehen häufig von einer „realistischen“ Problemstellung aus, an der, entlang systematisch aufgebauter Teilprobleme, mathematische Begriffe entwickelt werden. Der Kontext muss freilich nicht „realistisch“ im Sinne von handlungs- oder alltagsbezogen sein, vielmehr werden oft gleichzeitig mit realitätsnahen Problemen auch innermathematische Beziehungen eingebracht, etwa wenn bei geometrischen Mustern auch nach algebraischen Beschreibungen gefragt wird oder reale Bilder und schematische Skizzen schon von Anfang an gegenübergestellt werden. Das Adjektiv *realistic* ist daher wohl eher missverständlich. Es geht wesentlich um die Aneignung mathematischer Begriffe [...]“ (PISA 2000, S. 142 f.) Man muß sich also bereits an dieser Stelle von dem Gedanken befreien, Realistic Mathematics Education hätte allzu viel mit Realität zu tun. Das eben genannte Konzept liegt nah am Begriff der eingekleideten Aufgabe. Diesem Aufgabentyp wird in der mathematikdidaktischen Debatte mit einem gewissen Unbehagen begegnet: Es entwickelte sich aus der Kritik an der technischen und auf einen bestimmten - dem Inhalt oftmals nicht organisch bzw. zwingend innewohnenden - Lösungsweg gerichteten Orien-

tierung eingekleideter Aufgaben. Dieser Aufgabentyp hat seine Berechtigung aber nicht dadurch, dass dabei ernsthafte praktische Probleme gelöst werden bzw. vollständige Modellierungsprozesse durchlaufen werden. Eingekleidete Aufgaben dienen der Illustration eines mathematischen Gedankens, als Veranschaulichungsmittel. Sie sollen nicht vordergründig Mathematik und nichtmathematische (objektive) Welt verknüpfen, sondern sie sollen Mathematik und die Vorstellungswelt der Schüler verknüpfen, und zwar unter Rückgriff auf die als vorhanden angenommene Verbindung dieser Vorstellungswelt mit der nichtmathematischen (objektiven) Welt. In dieser Funktion haben sie ihre Berechtigung an jenen Stellen im Unterricht, in denen es um genau diese Verknüpfung geht. Man trifft aber auf eingekleidete Aufgaben, die „so tun", als seien sie Anwendungsaufgaben - dann liegt eine Debatte über ihre Nähe zur behaupteten Anwendung natürlich nahe. Das Problem ist hierbei aber, dass die Sache (also die Realsituation oder der mathematische Inhalt) nicht ernst genommen wird.
Eine adäquate normative Forderung an eine Aufgabe kann nur sein, dass sie das zu bearbeitende Problem und damit den Aufgabenlöser ernst nimmt. Das heißt insbesondere, dass die Aufgabe deutlich macht, welche Anforderung zu erfüllen ist. Die häufigste Verwerfung ist dabei sicher, dass so getan wird, als ob ein (eventuell praktisches) Problem zu lösen ist, obwohl lediglich ein technischer Akt zu vollziehen ist. Die dargestellte Kombination der Begriffsbildung „*Realistic* Mathematics Education" mit der Dementierung des Realitätsanspruchs in der Begriffsklärung läßt befürchten, dass wir es bei den PISA-Aufgaben mit solchen Verwerfungen zu tun bekommen.

Diese Befürchtung verstärkt sich, weil nicht zwischen einer guten unterrichtlichen und einer guten Testaufgabe unterschieden wird und weil keine Operationalisierungen vorgenommen werden. Testaufgaben sind dann nur Verschneidungen von Aufgaben, die ursprünglich vielleicht sogar reichhaltig waren:

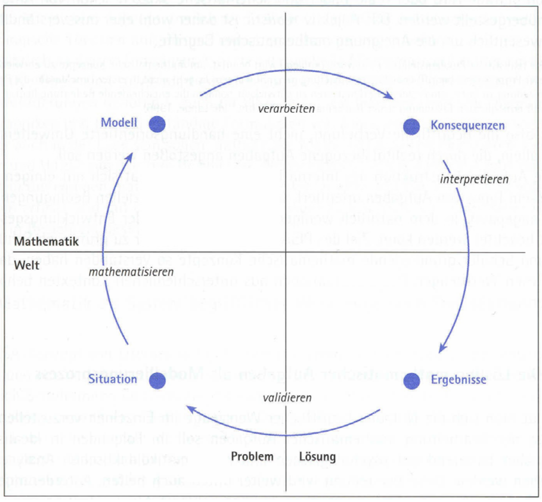
> „Die Aufgabenkonstruktion des internationalen PISA-Tests hat sich mit einigen Items an diesem [niederländischen] Typus von Aufgaben orientiert, sie allerdings den speziellen Bedingungen eines Tests angepasst, in dem natürlich weniger als im Unterricht der Entwicklungsgesichtspunkt beachtet werden kann. Ziel des PISA-Tests ist es vielmehr zu prüfen, ob Schülerinnen und Schüler grundlegende mathematische Konzepte so verstanden haben, dass sie mit diesen Werkzeugen Problemsituationen aus unterschiedlichen Kontexten behandeln können." (ebd., S. 143)

Damit ist der Begriff der *Realistic Mathematics Education* für den Test auf etwas sehr Schlichtes geschrumpft, zumal wegen der fehlenden Operationalisierung nie klar wird, worin die Umsetzung der Grundgedanken bestehen soll.

Modellbildung

Aus der PISA-Umsetzung des Konzepts der *Realistic Mathematics Education* heraus ist also kein normatives Ideal für die Beurteilung von Testaufgaben abzuleiten. In Bezug auf „Reales" in den Testaufgaben verbleibt damit nur der an jede Aufgabe zu stellenden Anspruch: Die Realsituation, auf die sich die Aufgabe bezieht, ist ernst zu nehmen (womit der Schüler als Problemlöser ernst genommen wird), und der mathematische Inhalt ist ernst zu nehmen (womit der Schüler als Lernender ernst genommen wird). Dieser Anspruch kann aus PISA heraus aber erweitert werden. Mit der dort verfolgten **Theorie des Modellierungsprozesses** wird nämlich implizit eine Norm gesetzt:

> „Im Kern geht es um das Verknüpfen von Situationen - diese können realitätsbezogen oder auch innermathematisch sein - mit mathematischen Ansätzen. Der mathematische Ansatz dient als Modell der Ausgangssituation. Die Bearbeitung und Lösung der Aufgabenstellung kann daher als Prozess der Erstellung, Verarbeitung und Interpretation eines mathematischen Modells beschrieben werden. [... Der Prozess des Modellierens] besteht aus einigen Teilprozessen, die ... mit mathematisieren → verarbeiten → interpretieren → validieren bezeichnet werden können. Diese Teilprozesse vermitteln zwischen „der Welt" und „der Mathematik" bzw. zwischen „dem Problem" und „der Lösung"..." (PISA 2000, S.143) Grafik zum zugehörigen Modellierungskreislauf: (ebenda, S. 144)



Diese Teilprozesse und die Verknüpfung von Situationen mit mathematischen Ansätzen liefern weitere Orientierungspunkte für die Interpretation. Es wird zu untersuchen sein, wie die Teilprozesse und die durch sie geleistete „Vermittlung" in der Aufgabenstellung angelegt sind bzw. gefördert oder behindert werden.
Dabei wird zu untersuchen sein, inwieweit die PISA-Modellierungstheorie sinnvoll ist: Hier wurde zunächst jedes Lösen einer mathematischen Aufgabe als Modellierungsprozeß angesehen (ebenda). Erst im Nachhinein wurden zumindest von der nationalen PISA-Gruppe die technischen Items aus dieser Sichtweise herausgenommen, für diese Items wird also im weiteren auf eine Theorie zur Lösung solcher Aufgaben verzichtet. (Knoche u.a. 2002, S. 162)
Man kann versuchen, die PISA-Modellierungstheorie anzuwenden, mit der Theorie also zu Erkenntnis über das Aufgabenlösen zu gelangen. Man bemerkt dann, dass man zwar für eine Aufgabe benennen kann, wie die Teilprozesse (mathematisieren → verarbeiten → interpretieren → validieren) konkret aussehen können. Das hat aber wenig mit der wirklichen Aufgabenlösung zu tun. In den Interpretationen werden viele Aufgabenlösungen generiert, keine einzige davon benötigt so einen Modellierungskreis. Der Modellierungskreis hilft weder bei der Lösung noch bei der Interpretation – also der didaktischen Erschließung – der Aufgabe. Auch die PISA-Gruppe verwendet diese Theorie bei der Testdeutung nicht. Die PISA-Modellierungstheorie ist also lediglich ein theoretisches Mäntelchen, welches dem Test umgelegt wird.
Der Grund für diese seltsame Nicht-Passung von Test und Rahmentheorie mag in Defiziten der Rahmentheorie liegen, aber auch darin, dass die PISA-Aufgaben keine realen Situationen zum Gegenstand haben: Ursprünglich sollte der Modellierungskreislauf einen Modellierungsprozeß beschreiben, dabei war wesentliches Element das Übersetzen aus der Realsituation in ein mathematisches Modell. Eckpunkte dieses Herangehens finden sich bei Schupp (1988), Kaiser (1995, eine ausführliche Darstellung der nationalen und internationalen Debatte), und Zais/Grund (1991, dort ein recht kleinteilig ausgear-

beiteter Ansatz), eher kurz die deutsche Diskussion zusammenfassend auch bei Blum (1985, S. 200-206; 1996, S. 16-20). Der potentielle Erkenntnisfortschritt dieser Ansätze liegt darin, die Übergänge zwischen der Welt der Realsituation und der Welt der Mathematik sowie das Tun innerhalb dieser Welten konkret für eine Aufgabe ausbuchstabieren zu können. (Es ist allerdings unklar, ob Modellierungsprozesse wirklich so ablaufen wie in diesen Konzepten angenommen.) Eine objektiv-hermeneutische Aufgabeninterpretation könnte die Passung des Modellierungskreislaufs für eine Aufgabe dahingehend untersuchen, ob die Aufgabe überhaupt prinzipiell ein Durchlaufen des Kreislaufes ermöglicht. Dazu wäre zunächst herauszuarbeiten, ob überhaupt eine Modellierung zu leisten ist, also ein Übergang zwischen der Welt des Realen und der Welt des Mathematischen nötig ist. In dieser Arbeit soll aber die Erklärungskraft der verschiedenen Modellierungsansätze nicht weiter untersucht werden.

Die nationale PISA-Gruppe ist zur Erkenntnis gelangt, dass es vom kognitiven Standpunkt aus unerheblich ist, ob „sich eine Modellierung auf eine innermathematische oder auf eine außermathematische Situation bezieht [...] (Cohors-Fresenborg 1996). Es sind die gleichen kognitiven Prozesse (formalisieren, übersetzen, strukturieren, im Modell arbeiten, zurückübersetzen), die auch bei der Behandlung einer innermathematischen Aufgabe ablaufen". (Neubrand u.a. 2001, S.50 f.) Auch Kurt Reussers Theorie der mentalen Situationsmodelle wird in den Theorie-Topf geworfen (PISA 2000, S.144 f.) Ob diese drei völlig verschiedenen Ansätze in irgendeiner Weise zusammenpassen, wird nicht diskutiert. Es wird auch kein neues theoretisches Konzept entworfen, das diese drei Ansätze zu einem schlüssigen Ganzen integriert. Es wird einfach suggeriert, dass diese drei Theorien etwas mit dem Inhalt des Tests zu tun hätten. Was genau sie mit der Erstellung des Tests zu tun haben, wird nicht ausgeführt. Inwiefern der so entstandene eklektische „Theoriemix" Erklärungen für die Aufgabenlösungen, also für die Deutung der Testresultate liefert, wird sich bei den Aufgaben zeigen.

Die PISA-Gruppe liefert einen weiteren normativen Anspruch:

> „In Schulbüchern und im Schulunterricht findet man allerdings oft die so genannten eingekleideten Aufgaben, die den Mathematisierungsprozess praktisch ausblenden oder weitgehend trivialisieren, weil sie den Eindruck erwecken, genau eine Weise der Mathematisierung sei „richtig". Dann wird also der für den Erwerb von Mathematical Literacy zentrale, ja charakteristische Vorgang des Mathematisierens abgeschnitten und die Aufgabe erscheint unmittelbar auf Modellebene." (S. 145)

Ich halte die hier mitschwingende Kritik an den eingekleideten Aufgaben im Kern für ein theoretisches Mißverständnis. Der oben dargestellte spezifische Ort dieser Aufgaben im Unterricht wird dabei nicht beachtet. Damit wird die eingekleidete Aufgabe verworfen – ohne dass es eine ernsthafte Alternative gäbe – statt sie entsprechend ihrer Funktion „ehrlicher" (also ohne die beschriebenen Verwerfungen) zu gestalten. Hier wird aber in der Distanzierung von eingekleideten Aufgaben für PISA-Aufgaben der Anspruch formuliert, den Vorgang des Mathematisierens nicht abzuschneiden. Die Einlösung dieses Anspruchs wird zu überprüfen sein. In der Zuspitzung dieses Anspruchs ist eine Aufgabe gefordert, die lediglich ein reales Problem vorstellt. Der Schüler muß dann den Modellierungsweg selbst gehen. Solche Aufgaben wurden von der PISA-Gruppe meist verworfen, weil sie im Vortest sehr selten gelöst wurden. Das heißt, dass man dem Schüler in seinen Modellierungsbemühungen entgegen kommen muß. Nun soll dabei aber der „zentrale, charakteristische Vorgang des Modellierens" nicht abgeschnitten werden, weil dies dem Konzept von Mathematical Literacy widerspräche. Man muß dem Schüler also Modellierungsanforderungen abnehmen, ohne gleich alle Modellierungsanforderungen zu zerstören, ohne die Vielfalt der Lösungswege allzu sehr einzuschränken und ohne die mathematische Anforderung zu zerstören. Die von der PISA-Gruppe angestrebte „Vermittlung zwischen „der Welt" und „der Mathematik", zwischen „dem Problem" und „der Lösung"" (ebenda, S.143) muß also in der Aufgabenkonstruktion bereits angelegt sein. Dies ist ein weiterer Maßstab der Interpretation.

Zur Umsetzung des Fähigkeitskonstrukts in Testitems

Bei der Testerstellung ist nun die Frage, wie man die Grundideen solcher Konzepte wie „Mathematical Literacy" in Testitems umsetzt:

> „In der angelsächsischen Diskussion ist dieser Begriff bildungstheoretisch breit entfaltet und bis hinein in die Empfehlungen des *National Council of Teachers of Mathematics* (NCTM) der Vereinigten Staaten konkretisiert worden (NCTM, 1989, 2000). Wesentliche Zieldimensionen bilden demnach über fachliche Kenntnisse und Fähigkeiten hinaus das mathematische Denken und die Anwendung auf innermathematische und außermathematische Aufgabenstellungen. Die Schülerinnen und Schüler sollen befähigt werden, die Anwendbarkeit mathematischer Konzepte und Modelle auf alltägliche - vor allem auch offene, nicht gut definierte - Problemstellungen zu erkennen, die einem Problem zu Grunde liegende mathematische Struktur zu sehen und Aufgabenstellungen in mathematische Operationen zu übersetzen. Den Kern von *Mathematical Literacy* bildet somit das Anwenden, verstanden als Prozess des Modellierens von Situationen mithilfe mathematischer Begriffe. [...] Zusätzlich zu diesen kognitiven Aspekten umschließt *Mathematical Literacy* eine kommunikative Dimension sowie mathematikbezogene Einstellungen."[53] (PISA 2000, S. 141 f.)

Für die Testerstellung ist nun aus diesen immer noch groben Fähigkeitsbeschreibungen ein konkreteres Fähigkeitskonstrukt zu erarbeiten und zu operationalisieren. Praktisch scheint es sinnvoll, die Erarbeitung des Konstrukts und seine Operationalisierung teilweise zusammen laufen zu lassen. Es kann dabei punktuell durchaus sinnvoll sein, eine mathematische Fähigkeit lediglich durch ein zu lösendes Problem oder ein wiederzugebendes Wissenselement zu beschreiben. – Dieses Vorgehen kann aber nicht ein ganzes Testkonstrukt hervorbringen.
Die PISA-Gruppe beschreibt an keiner Stelle, dass sie ein Fähigkeitskonstrukt erarbeitet hat, welches konkret wiedergibt, was Mathematical Literacy ist.

[53] In diesem Zitat findet sich die in Kapitel 3 beschriebene Gefahr der technokratischen Verkürzung von Lernen auf Tests hin wieder: Mathematical Literacy ist die Umsetzung eines didaktischen Konzepts in ein Fähigkeitskonzept. Wenn diese Verkürzung in den NCTM-Standards „konkretisiert" wurde, dann liegt hier die Konkretisierung des Fähigkeitsaspektes eines didaktischen Konzepts vor. Wenn die deutsche Arbeitsgruppe zu mathematischen Bildungsstandards in der Sekundarstufe I wie geplant vom PISA-Konzept ausgehend arbeitet, dann praktiziert sie eine ebensolche Verkürzung.

Eine Vorentscheidung über den Charakter ihres Testkonstrukts hat die PISA-Gruppe mit der Entscheidung für *schulnahe Mathematikaufgaben* als Testitems getroffen: Damit hat sie sich gegen andersartige Items und für eine gewisse Orientierung an einer bestimmten Art der Vermittlung und Prüfung mathematischen Wissens entschieden. Es wird - wahrscheinlich durch die methodische Voreinstellung der beteiligten Forscher bzw. durch die politische Vorgabe des Vergleichs von Schulsystemen - als selbstverständlich vorausgesetzt, dass das Konstrukt „mathematische Fähigkeiten" schulmathematische Fähigkeiten meint, von denen man sich dann bewußt auf jene konzentriert, die durch das Konzept der Mathematical Literacy angestrebt bzw. abgedeckt werden. Die Vorentscheidung für Schulmathematik und für die Itemform „Aufgabe" prädiziert damit ein bestimmtes Bild davon, was Mathematik ist. Mathematik ist dann nämlich das, was in schulischen Mathematikaufgaben stattfindet. Hier bestimmt das Ganze - nämlich die Grundform des Tests - bereits den Teil - nämlich die jeweilige Aufgabe, genauso wie der Teil das Ganze bestimmt. Würden wir also in den Interpretationen plötzlich Anforderungen aus mathematischen Intelligenztests, Projektentwicklungsaufforderungen, Beweisanforderungen, Aufforderungen zur Darstellung der Struktur eines schulmathematischen Teilgebiets oder zu tiefergehenden Auseinandersetzungen in Form kurzer Aufsätze finden, dann läge ein Strukturbruch vor, der einer Interpretation bedürfte.

Es steht bei der Testentwicklung nicht nur das herkömmliche didaktische Problem, die eigenen didaktischen Gedanken in Aufgabenstellungen fließen zu lassen. Die didaktischen Gedanken müssen auch noch „operationalisiert" werden. Das heißt, dass die didaktischen Gedanken in ein Fähigkeitskonstrukt umgesetzt werden müssen, welches dann operationalisiert werden muß, um das Vorhandensein der mit den didaktischen Gedanken verbundenen Fähigkeiten messen zu können. Ideen unterrichtlichen Handelns müssen also klar vom Problem der Fähigkeitenmessung getrennt werden.
Die Aufgabeninterpretation muß andersherum fragen: *Ist der didaktische Gedanke in die Aufgabe umgesetzt und ist er zusätzlich so operationalisiert, dass die Aufgabe wirklich das Vorhandensein der entsprechenden Fähigkeit mißt?* Die Interpretation vollzieht dabei in gewisser Weise den Prozeß nach, den auch die Testersteller bei PISA gehen, denn sie haben keine „wirkliche" Operationalisierung vorgenommen: Bei der Testerstellung wurde *praktisch* nicht so vorgegangen, dass aus den didaktischen Ideen der Freudenthalschen Didaktik oder der Mathematical Literacy und ihren Konkretisierungen heraus neue Aufgaben entworfen wurden, die dann Operationalisierungen von konkretisierten Ideen waren. Statt dessen wurden (im besten Fall) die didaktischen Ideen konkretisiert und dann *aus bereits vorhandenen Aufgabensammlungen* Aufgaben daraufhin ausgewählt, ob sie wie Operationalisierungen der Ideen aussehen. Das kann man so zugespitzt sagen, weil genau dieser Prozeß - der ja ein Kernprozeß der Herstellung von Geltung der Testaussage ist - alltagshermeneutisch, also ohne methodische Kontrolle ablief. Da bei der Testerstellung kein systematisches, geschweige denn ein methodisch kontrolliertes Operationalisierungsverfahren benutzt wird, setzt man an seine Stelle den Gedanken „Mehrere geschulte Augen sehen mehr als zwei Augen". Die Testaufgaben durchlaufen also mehrere Ebenen von Expertenbesichtigungen. Dies bringt einerseits das ganz pragmatische Problem, dass der jeweilige Experte mehrere Dutzend Aufgaben besichtigen und bewerten muß. Das viel schwerwiegendere

Problem ist aber, dass die Experten zwar Fehler ausmerzen und auf Probleme aufmerksam machen können, dass aber auch sie keinen methodisch kontrollierbaren Prozeß vollführen. Offensichtlich steht dahinter der Gedanke, dass allein durch eine große Anzahl von Experten eine Art methodische Kontrolle hergestellt wäre. Die Argumentation bezieht sich wohl deshalb nur auf die Anzahl und das Renommee der beteiligten Experten. In einer wissenschaftlichen Arbeit erwartet man aber eine Mitteilung darüber, was diese Experten konkret tun, wie sich zum Beispiel verschiedene methodische Herangehensweisen in der Aufgabenbeurteilung zu einem schlüssigen Ganzen vereinen. Natürlich nimmt ein „guter“ Experte auch latente Probleme einer Aufgabe wahr, das ist wesentlicher Bestandteil seines Expertentums – und in diesem Sinne erhöht die Anzahl beteiligter Experten auch die Qualität des Tests. Aber auch seine Wahrnehmung unterliegt keiner methodischen Kontrolle und kann die (natürlich immer nur relative, aber die Grundstruktur abdeckende) Vollständigkeit einer kontrollierten und methodisch ausgearbeiteten Interpretation nicht erreichen. Zudem hat er im kulturindustriellen, machbarkeitsorientierten Arbeitskontext ein schwieriges Argumentationsumfeld, wenn er die latenten Probleme dem Kollektiv nachvollziehbar ausargumentieren will.
Die Interpretationen zeigen, dass manche Aufgaben wie Operationalisierungen bestimmter Fähigkeiten aussehen, es aber nicht sind oder nicht nur sind. Darin zeigt sich die Begrenztheit des methodisch nicht kontrollierten Vorgehens. Objektive Hermeneutik erweist sich gleichzeitig als eine Möglichkeit der Auswahl von Aufgaben, weil sie Entscheidungen darüber ermöglicht, ob eine Aufgabe zur Operationalisierung bestimmter Ideen dienen kann.

5.2. Das internationale PISA-Konzept

Die PISA-Gruppe hat ein Papier veröffentlicht, das beschreibt, wie die Grundidee von Mathematical Literacy inhaltlich konkretisiert ist, so dass dann ein Testinstrument entstehen kann: *Measuring Student Knowledge And Skills. The 2000 PISA Assessment of Reading, Mathematical and Scientific Literacy.* (OECD 2000) Seine wesentlichen Gedanken sind auch in Neubrand u.a. (2001) zusammengefaßt.

Es zeigt sich, dass es bei der inhaltlichen Konkretisierung von Mathematical Literacy nicht um die Vorbereitung einer Operationalisierung im strengen Sinne geht: Die PISA-Gruppe will nicht bestimmten Aufgaben bestimmte Fähigkeiten zuordnen. Eine Interpretation kann also - entgegen dem oben dargelegten theoretischen Ideal - eine solche Zuordnung nicht überprüfen, sie kann lediglich untersuchen, ob aus dem Aufgabentext selbst eine solche Zuordnung vorgenommen werden kann. Die PISA-Gruppe will vorrangig eine gewisse Aufgabenbreite absichern. Sie eröffnet mehrere Aufgaben- bzw. Fähigkeitskategorien, die abgedeckt werden sollen:

1. Ebene: Mathematical Processes/Competencies

> „PISA tasks are designed to encompass a set of general mathematical processes that are meant to be relevant and pertinent to all education levels:
> 1. Mathematical thinking ...

2. Mathematical argumentation ...
3. Modelling ...
4. Problem posing and solving ...
5. Representation ...
6. Symbols and formalism ...
7. Communication ...
8. Aids and tools ...“ (OECD 2000, S. 46 f.)

Eine Vorstellung von der Art der zugehörigen Erläuterungen erhält man durch folgendes Beispiel:

„6. Symbols and Formalism, which includes:

- decoding and interpreting symbolic and formal language and understanding its relationship to natural language;
- translating from natural language to symbolic/formal language;
- handling statements and expressions containing symbols and formulae;
- using variables, solving equations and undertaking calculations.“ (OECD 2000, S. 47)

Man erkennt, dass für die Frage dieser Arbeit, was der Test bzw. einzelne Aufgaben messen, diese Kriterien durchaus Hilfe bzw. Normen beisteuern. Operationalisierungsgedanken oder wenigstens Zuordnungen der Anforderungen zu Aufgaben finden sich nun im weiteren nicht. Die Angaben zu den „Processes“ geben eine *Vollständigkeitsorientierung* für die gewünschten mathematischen Fähigkeiten, lassen aber keine Schlußfolgerung zu, was die Tester mit einer einzelnen Aufgabe messen wollen, welche Fähigkeiten auf welchem Niveau der Schüler also haben soll, damit er eine Aufgabe oder eine Itemkombination erfolgreich lösen kann. Diese Frage bleibt auch im weiteren unbeantwortet, im Verlauf der Testentwicklung wurde also lediglich versucht, vorhandene Aufgaben auf diese groben Kategorien aufzuteilen.

So eine *kategoriale* Zuordnung funktioniert natürlich nicht, und so stellt die PISA-Gruppe fest: „When doing „real mathematics“ it is necessary to draw simultaneously upon many of these skills.“ (ebd.) Zur Zuordnung der Testaufgaben vergröberte sie also die Fähigkeitsniveaus wieder und wählte für diese neue Einteilung den Namen „Levels of mathematical Competency“ oder „Competency Classes“:

> „PISA organises processes into three classes, defining the type of thinking skill needed: i) reproduction, definitions and computations; ii) connections and integration for problem solving; and iii) mathematisation, mathematical thinking, generalisation and insight. In general, these processes are in ascending order of difficulty, but it does not follow that one must be mastered in order to progress to the other: it is possible for example to engage in mathematical thinking without being good at computations.“ (ebd.)

Diese Kategorien sind nun groß und grob genug, um sie und damit die Breite der zu testenden mathematischen Fähigkeiten mit genügend vielen Aufgaben abdecken zu können und jede Aufgabe auch halbwegs einer Kategorie zuordnen zu können. (Wir werden Ausnahmen kennenlernen.) Diese Zuordnung führt natürlich noch weiter weg von einer Zuordnung von Anforderungen zu Aufgaben.

2. Ebene: „Mathematical content: strand and „big ideas“

School mathematics curricula are usually organised in strands. These strands compartmentalise mathematics and often over-emphasise computation and formulae.“ (ebd., S. 48) Statt durch diese Gebiete - z.B. Arithmetik, Geometrie, Algebra, Infinitesimalrechnung - organisiert PISA die Inhalte in „big ideas“. „Darunter sind miteinander stark vernetzte mathematische Konzepte verstanden, die unter einem gemeinsamen übergeordneten Gesichtspunkt gesehen werden können.“ (Neubrand u.a. 2001, S. 55) Die „big ideas“ von PISA sind: change and growth; space and shape; quantitative reasoning; uncertainty; dependency and relationships. Im Jahr 2000 wurden nur die ersten beiden Ideen getestet. Die PISA-Gruppe erläutert die Ideen, ich zitiere nur die Anforderungen:

> „1. Change and growth
> ... PISA examines students` ability to represent changes in a comprehensible form, to understand the fundamental types of change, to recognize particular types of change when they occur, to apply these techniques to the outside world, and to control a changing world to our best advantage. ...
> 2. Space and shape
> ... In understanding space and shape, students need to look for similarities and differences as they analyse the components of form and recognise shapes in different representations and different dimensions. This means that they must be able to understand the relative positions of objects. They must be aware of how we see things and why we see them as we do. They must learn to navigate through space and through constructions and shapes.
> Students should therefore be able to understand the relations between shapes and images or visual representations, such as that between a real city and photographs and maps of the same city. They must also understand how three-dimensional objects can be represented in two dimensions, how shadows are formed and must be interpreted, what „perspective“ is and how it functions.“ (OECD 2000, S. 49)

Die zweite Ebene von Fähigkeitskategorien ist also die Ebene der mathematischen Inhalte, die quer zu den klassischen schulmathematischen Gebieten konstruiert ist. Die hier beschriebenen Fähigkeiten sind - wie das auch bereits auf der ersten Ebene der Fall war - relativ allgemein gehalten. Die Aufgabeninterpretationen können damit zwar auf diese Anforderungen rückbezogen werden, aber überprüfbare Operationalisierungen oder wenigstens spezifizierte Anforderungen liegen auch hier nicht vor. Auch hier wird lediglich eine Inhaltsbreite abgesteckt, die durch die Aufgaben abgedeckt werden soll.

3. Ebene: „Situations and contexts

> Students` mathematical insight and understanding need to be assessed in a variety of situations, partly to minimize the chance of students finding that tasks are not culturally relevant.
> One can think of a situation as being at a certain distance from the student. The closest is private life (daily life), next is school life, work and sports, followed by the local community and society as encountered in daily life, and furthest away are scientific contexts. In this way one can define a more or less continuous scale of situations.“ (ebd., S. 50)

Hier geht es also wieder um die Abdeckung einer Breite eines für wichtig gehaltenen Aufgabencharakteristikums: der Schülernähe. Es steht zu befürchten, dass sich die recht schlicht gehaltene Vorstellung

eines Grades von Schülernähe in den Aufgaben wiederfindet. Bereits in der Konstruktion deuten sich Potentiale für Verwerfungen an: Es ist ja zum Beispiel keineswegs anzunehmen, dass Schulleben oder kommunale Angelegenheiten den Schülern „näher“ sein müssen als wissenschaftliche Kontexte. Auch ein Einfluß auf die Schwierigkeit liegt nicht nahe. Hier liegt offenbar eine recht technisierte und unzulässig vereinfachte Deutung der Erkenntnis vor, dass man sich leichter auf Kontexte einlassen kann, die einem selbst emotional oder kognitiv näher sind, und dass subjektiv nähere Kontexte die subjektive Schwierigkeit verringert.
Die PISA-Gruppe formuliert mit den Vokabeln *authentisch* und *glaubhaft* die Normen, die an die Betrachtung der Kontexte anzulegen ist:

> „Whatever the distance from the students, PISA aims to ensure that tasks are based on „authentic“ contexts which are likely to occure in a real world setting. If mathematics education is to prepare students to be active and informed citizens it has to deal with „real“ contexts such as pollution problems, traffic safety and population growth. This does not exclude, however, artificial fictional contexts based on the stylised representation of problems - such as a traffic situation in a non-existent town.“ (ebd.)

Es hat sich herausgestellt, dass die drei Ebenen, auf denen im internationalen PISA-Teil die Aufgaben bzw. die zu testenden Fähigkeiten kategorisiert sind, nur in grober Form Normen vorgeben, an denen bei den Aufgabeninterpretationen gemessen werden kann. Die Beschreibung der Ebenen diente offenbar vorrangig der Erstellung eines Vollständigkeitskatalogs. In der Logik der PISA-Erstellung müßte man nun überprüfen, ob jede Kategorie gleich stark mit Aufgaben gefüllt ist. Diese Frage ist aber für diese Arbeit nicht von Interesse. Es ist ebenfalls nicht von Interesse, ob diese Kategorisierungen sinnvoll sind: Der Rückblick auf die bisherigen Interpretationen (und der Blick auf die folgenden) zeigt bereits, dass die Zuordnung und die Zuordenbarkeit zu Kategorien in Frage steht. Diese Fraglichkeit wird nicht erst durch Objektive Hermeneutik bewußt und spielt in Debatten um Testkonstruktionen immer wieder eine Rolle. Gegenargumente klingen meist pragmatisch und orientieren sich lediglich an der technischen Machbarkeit des Tests, nicht an der Geltungskraft der mit ihm getroffenen Aussagen: „So genau interessiert das gar nicht.“; „Das ist nicht so schlimm, die Zuordnung zu Kategorien dient nur der Orientierung.“; „Das Problem verliert sich durch die Menge der Items.“ usw.
Die Kategorienzuordnung ist keine Operationalisierung und sie wird im internationalen Teil auch nicht für theoretische Schlußfolgerungen genutzt. Für die Fragestellung dieser Arbeit muß sie also nicht untersucht werden. Ich werde mich in den Interpretationen also lediglich auf die normativen Aussagen beziehen. Da keine Operationalisierungen erfolgt sind, ergibt sich die Frage, was getestet werden soll, ausschließlich aus dem Aufgabentext selbst. Für die internationalen Aufgaben stellt das Konzept der Testerstellung also eine eher geringe Bereicherung für die Interpretationen dar.

5.3. Nationale Ergänzung des internationalen PISA-Mathematik-Tests in einer deutschen Zusatzerhebung

Die deutsche PISA-Expertengruppe Mathematik (siehe Neubrand u.a. 2001) hat ein Papier zur Begründung und Beschreibung einer nationalen Ergänzung des PISA-Tests vorgelegt. Für diese nationale Ergänzung, die insbesondere den Vergleich zwischen den deutschen Bundesländern in Mathematik und Naturwissenschaften differenzieren sollte, wurde ein zweiter Testtag angesetzt.

Gründe für eine nationale Ergänzung

Die deutsche PISA-Expertengruppe Mathematik konstatiert eine sehr spezifische - sich an Freudenthal anlehnende - Auslegung des Begriffs „mathematical literacy" im internationalen PISA-Framework. Sie begründet die nationale Ergänzung mit der Notwendigkeit einer Widerspiegelung der realen Kalkülorientierung im deutschen Mathematikunterricht und der Berücksichtigung von Aspekten „mathematischer Grundbildung", welche im freudenthalorientierten Begriff von „Mathematical literacy" nicht stark genug vertreten seien:

> „Offenbar stehen also in der „realistic mathematics education", und demnach abgeleitet auch bei „mathematical literacy" gemäß dem internationalen PISA-Framework, mathematische Konzepte als zu erreichendes Ziel im Vordergrund. Es wird nun verständlich, dass die in den traditionellen Curricula oft vorherrschend unterrichteten mathematischen Fertigkeiten nicht den Hauptfocus von PISA bilden können. Diese „skills and routines" werden daher im internationalen PISA-Test nicht isoliert erfasst, sondern sind stets eingebunden in konzeptuell geprägte und kontextgebundene Aufgaben und Aufgabenkomplexe. ...
>
> Wünschenswerte Ziele des Mathematikunterrichts, nämlich praktische Nutzbarkeit und systematischen Aufbau der Mathematik so zu verbinden, dass nicht lediglich isoliertes Lernen von Fähigkeiten entsteht, sondern kontextverankerte mathematische Begriffe ausgebildet werden, sind in dieser Konzeption konkret, allerdings in einer sehr spezifischen Weise umgesetzt. Es sind also hierin Komponenten „mathematischer Grundbildung" (Winter 1995) mit angesprochen. ... Allerdings gehört zu den Zielen einer „mathematischen Grundbildung" ... zusätzlich der Hinweis darauf, dass Mathematik auch als „eine deduktiv geordnete Welt eigener Art" (Winter 1995), gelöst von den phänomenologischen Verankerungen, gesehen werden kann und soll. [[54]] Dieser Aspekt ist in der internationalen PISA-Konzeption eher unterrepräsentiert. Betont wird aber hier wie dort, dass es auf die Verbindung inhaltlicher und formaler Kenntnisse ankommt." (Neubrand u.a. 2001, S.47)

---

[54] Hier werden Mathematik und Schulmathematik vermengt, die man deskriptiv als Gebiete mit unterschiedlichen Charakteristika ansehen muß. Die Wintersche Sichtweise auf Mathematik als „eine deduktiv geordnete Welt eigener Art" führt zunächst zu einer Forderung an Mathematikunterricht, diesen Charakter widerzuspiegeln. Wenn die Mathematik heutigen Mathematikunterrichts, also die Schulmathematik, „eine deduktiv geordnete Welt eigener Art" *ist*, dann ist jedenfalls sowohl ihre Deduktivität als auch die in ihr vorherrschende Art des Ordnens als auch ihr eigener Charakter verschieden von dem der Mathematik.
Zu dieser Vermengung paßt, dass sich die nationale PISA-Gruppe zwar auf das Grundbildungskonzept beruft, in der weiteren Argumentation und in der konkreten Umsetzung aber fast nur noch Fertigkeiten übrig bleiben. Der Blick in die nationalen PISA-Testhefte zeigt dann auch, dass fast ausschließlich Kalkülorientierung im Zentrum dieser Testteile steht. Es findet also keine Operationalisierung des Winterschen Theoriegebäudes statt, es bleibt vorgeschobene Theorie zur Rechtfertigung eines Aufgabenkatalogs, der sich nicht aus dieser Theorie herleitet.

Damit wird nun ein neuer Interpretationsfokus eingeführt, nämlich der *Charakter von Mathematik als deduktiv geordnete Welt eigener Art*. Dieser Fokus spielt für die vorliegende Untersuchung nur eine untergeordnete Rolle, da keine der veröffentlichten nationalen PISA-Aufgaben diesen Charakter zum Thema hat. Dieses Defizit verweist – im Zusammenhang mit der Bedeutung dieses Charakters in den theoretischen Rahmensetzungen der nationalen PISA-Gruppe – wiederum auf ein Element des zu rekonstruierenden mathematikdidaktischen Habitus`. Es wäre auch für diese Habitusuntersuchung interessant gewesen, die spezifische Umsetzung dieses Charakters in einer operationalisierten Form zu untersuchen. Die PISA-Gruppe stellt aber keine solche Operationalisierung vor. Anders ist es mit der Kalkülorientierung des deutschen Mathematikunterrichts:

> „Im realen Mathematikunterricht in Deutschland und korrespondierend in der Struktur der Leistungen deutscher Schülerinnen und Schüler bildet sich jedoch diese Verbindung von praktischer Anbindung, formaler Kenntnis und begrifflicher Vertiefung *nicht* ab. Nach vorliegenden empirischen Untersuchungen zeigen sich vielmehr im deutschen Mathematikunterricht Probleme beim verständnisvollen Gebrauch und bei der Vernetzung mathematischer Konzepte, sowie eine Überbetonung kalkülorientierter Fertigkeiten. ... Die dominante begriffliche Orientierung des internationalen PISA-Tests, die oben als Konsequenz der „mathematical literacy"-Definition im Framework herausgearbeitet wurde, trifft also (auch und insbesondere) in Deutschland nicht auf eine entsprechend vorbereitete Unterrichts- und curriculare Realität." (ebd. S.47 f.) Da „Schülerinnen und Schüler diejenigen Aufgaben relativ zu ihrem allgemeinen Leistungsstand besser beherrschen, die ihrem Curriculum sowohl vom Stoff als auch von den impliziten Ausrichtungen, der „Färbung" ... her am ehesten entsprechen" (ebd., S. 48)

und wegen des Bundesländervergleichs wird ein verschärftes Problem der Passung konstatiert, das eine Erweiterung des internationalen Tests erfordert.

> „Daher sind - wenigstens im ersten Zyklus von PISA - auch formale und technische Kenntnisse entsprechend zu berücksichtigen. Längerfristige Konzepte zur Fortentwicklung des Mathematikunterrichts sehen solche Fertigkeiten allerdings mehr und mehr in ihrer Vernetzung, und folglich ist der Stellenwert dieser Kenntnisse auch in den folgenden Zyklen von PISA kritisch zu prüfen." Aber auch innerhalb von Reformbestrebungen ist „diese Ausgangslage [...] zuerst zu dokumentieren, um sie überwinden zu können.
>
> Zu diesem Zweck wird im folgenden eine *Differenzierung* und *Erweiterung* des internationalen Frameworks, das gleichwohl als Ausgangsbasis dient, vorgenommen. Der internationale PISA-Test wird dabei auf der Basis dieser Überlegungen in der deutschen Zusatzerhebung durch Items ergänzt, die in dieser Art im internationalen Test gar nicht vorkommen oder dort in Zusammenhängen und „Färbungen" auftreten, die für deutsche Schülerinnen und Schüler eher ungewohnt sind. Insgesamt ergibt sich dann aus der *Kombination* des internationalen PISA-Tests und der deutschen Zusatzerhebung die Möglichkeit, die „mathematische Grundbildung" deutscher Schülerinnen und Schüler adäquater als es durch den internationalen Tests allein möglich wäre zu erfassen."[55] (ebd.,

[55] Die etwas seltsame Sprachfigur, welche die Kalkülorientierung deutschen Mathematikunterrichts plötzlich dem Konzept der mathematischen Grundbildung zuschlägt, verschleiert den Verlust des Aspekts von Mathema-

S. 48 f.)

Inhalt der nationalen Ergänzung

Die „Deutsche PISA-Expertengruppe Mathematik“ differenziert ihr Framework in mehreren Dimensionen mathematischer Fähigkeiten. Die Differenzierung stellt dabei nicht nur ein Mittel zur Erreichung einer großen Breite von Aufgabenanforderungen bei der Testkonstruktion dar, sie soll auch der Interpretation der Testresultate dienen. Die Anforderungsmerkmale sind so differenziert beschrieben, dass es nahe liegt, sie als „Maßstab“ für objektiv-hermeneutische Interpretationen zu nutzen: Man kann die Aufgaben interpretieren und einerseits die Sinnhaftigkeit der Zuordnung der Aufgabe zu den verschiedenen Einordnungsdimensionen der PISA-Gruppe diskutieren. Dabei wird eventuell die Sinnhaftigkeit der Dimensionen selbst thematisch. - Die Frage nach einer sinnvollen Konstruktion von Meßdimensionen für „mathematische Leistungsfähigkeit“ geht allerdings über die Fragestellung dieser Arbeit hinaus. Ich zeige lediglich, dass objektiv-hermeneutische Aufgabeninterpretationen diesen Konstruktions- und den Operationalisierungsprozeß präzisieren. Die Änderung einer Aufgabenstellung, so dass sie als Operationalisierung einer solchen Dimension dient, ist die verbesserte Herstellung einer Subsumtion. Die Rekonstruktion der Sinnstruktur einer Aufgabe in einer objektiv-hermeneutischen Interpretation kann also zu einer präziseren Subsumtion führen. Sie hat damit aber noch nicht die Frage beantwortet, inwieweit das notwendigerweise subsumierende Vorgehen beim Messen für sich genommen sinnvoll ist, ob also so eine Subsumtion eine adäquate Widerspiegelung dessen ist, was man sinnvollerweise als „mathematische Leistungsfähigkeit“ bezeichnet.

Die *Elemente des deutschen Frameworks* sind im folgenden zusammengefaßt (siehe Neubrand u.a. 2001, S. 50 ff.):

- Ausgewählte mathematische Kompetenzen werden differenzierter betrachtet als im internationalen Framework: Begründungsfähigkeiten, mathematische Tätigkeiten (z.B. induktiver Schluß oder Verallgemeinerung), Modellierungsfähigkeiten (hier wird die eventuelle Dominanz einer der Aspekte „präzisieren“, „mathematisieren“, „interpretieren“ und „validieren“ erfaßt). (ebd., S. 50)
- Die Präsentation eines Items (Text, Graphiken, Diagramme, Funktionsgraphen, geometrische Zeichnungen, Tabellen, reale oder realitätsnahe Bilder, Landkarten) wird nur insofern betrachtet, als sie Aufschluß darüber gibt, aus welchen Repräsentationsformen heraus der Modellierungsprozeß einsetzt. (ebd., S. 51)
- Eine weitere Unterscheidung erfolgt nach Stoffgebieten: Arithmetik, Proportionalität (insbesondere Prozentrechnung), Algebra (insbesondere lineare Funktionen), Geometrie, Stochastik und Umgehen mit Daten. (ebd., S. 55)
- Kontextvariablen werden unterschieden: authentisch, realitätsbezogen, Rechnen mit Größen, innermathematisch, ohne Kontext (ebd., S.55)
- Gesichtspunkte für die Analyse einzelner Items sind darüber hinaus: Art der angesprochenen Grund-

tik als deduktiv geordneter Welt eigener Art.

vorstellungen, sprachlogische und kognitive Komplexität (z.B. Übereinstimmung sprachlicher Fluß und Lösungsgang im mathematischen Modell), formaler Typus des Problemlöseprozesses (Umkehraufgabe, Aufgabe, die gegen den Gedankengang der zugehörigen Grund- oder Standardaufgabe zu lösen ist) (ebd., S.54)

Die Zuordnung aller bisher genannten Differenzierungen zu den einzelnen Aufgaben ist nicht veröffentlicht, kann hier also nur aus dem Aufgabentext selbst heraus diskutiert werden. Es erscheint allerdings auch als wenig interessante Forschungsfrage, ob die PISA-Gruppe die Zuordnung zu all diesen Kategorien - die ja eher eine gewisse Aufgabenbreite garantieren als interessante theoretische Erkenntnisse zu versprechen - richtig vorgenommen hat. Das Problem der Zuordnung von Aufgaben zu solchen Kategorien erschließt sich bereits, wenn wir die Zuordnung nur für das folgende Kategoriensystem betrachten, für welches die Zuordnungen auch teilweise veröffentlicht sind:

Als wesentlichen Kern des nationalen PISA-Frameworks bezeichnet die deutsche PISA-Gruppe die sogenannten **Kompetenzklassen**:

Klasse 1A: Technische Fertigkeiten
Klasse 1B: Einschrittige Standardmodellierungen
Klasse 2A: Begriffliche Modellierungen
Klasse 2B: Mehrschrittige Modellierungen (ggf. integrativ/repetitiv)
Klasse 3: Strukturelle Verallgemeinerung

Im einzelnen werden die Klassen zusammenfassend folgendermaßen beschrieben:

„**Klasse 1A:** Die Klasse erfordert nur technische Fertigkeiten und/oder den Abruf von Faktenwissen. [...]

Klasse 1B: Zur Lösung der Aufgabe ist eine Modellierung erforderlich. Diese ist jedoch unter Rückgriff auf einen einzigen Algorithmus, eine einzige Formel möglich. Es ist also die passende Formel, das passende Verfahren, die passende Prozedur aus dem vorhandenen Wissen auszuwählen und dann anzuwenden. Das zur Modellierung erforderliche Wissen stammt aus einem einzigen mathematischen Gebiet. [...]

Klasse 2A: Der zur Lösung erforderliche Schritt ist überwiegend begrifflicher Art. Die Lösung der Aufgabe kann durch Anwendung eines einzigen begrifflichen Arguments, durch Herstellung eines begrifflichen Zusammenhangs erfolgen. Dabei genügt es auch, den erforderlichen Begriff als Wissensbestandteil abzurufen. [...]

Klasse 2B: Die Struktur der Modellierung ist mehrschrittig, d.h. bei der Lösung der Aufgabe ist entweder Wissen aus mehreren mathematischen Zusammenhängen einzusetzen oder mehrfach gleiche Schritte sind vorzunehmen und zu kombinieren. [...]

Klasse 3: Die Aufgabe beinhaltet Schritte der Verallgemeinerung, des Entwerfens einer allgemeinen komplexen Strategie, der Reflexion über das verwendete mathematische Modell, die Präsentation eines subtilen mathematischen Arguments, etc..“ (Neubrand u.a. 2001, S. 52-54)

Unter Bezugnahme auf diese Klasseneinteilung der deutschen PISA-Gruppe führen Klieme, Neubrand und Lüdtke (Klieme u.a. 2001) eine andere Einteilung ein, nämlich in „rechnerische Modellierungsaufgaben“ und „begriffliche Modellierungsaufgaben“. Die deutsche PISA-Gruppe (Knoche u.a. 2002) bezieht sich nun wiederum auf diese Einteilung, um eine dritte Gruppe „technische Aufgaben“ hinzuzufügen und ordnet ihre „alten“ Klassen diesen „neuen“ Klassen zu. Damit ergibt sich folgende neue Klasseneinteilung:

technische Aufgaben: „Aufgaben, bei deren Bearbeitung überwiegend technische Fertigkeiten oder Faktenwissen benötigt werden.“ (Knoche u.a. 2002, S.164)
rechnerische Modellierungsaufgaben (Kompetenzklassen 1B und 2B): „Dies sind Aufgaben, bei denen die Mathematisierung auf rechnerisch durchzuführende Modelle hinausläuft. Typische Beispiele sind „klassische“ Textaufgaben, von eingekleideten Aufgaben bis hin zu komplexeren Anwendungsproblemen.“ (Klieme u.a. 2001, S.146)
begriffliche Modellierungsaufgaben (Kompetenzklassen 2A und 3): „Hierzu gehören vor allem jene Aufgaben, zu deren Lösung ein begrifflich geprägter Zusammenhang herzustellen ist, was bis zur strukturellen Verallgemeinerung einer Situation oder dem Entwerfen einer umfassenden Strategie reichen kann (Beispiel: Aufgabe „Äpfel 3“ ...). Zum Bereich der begrifflichen Modellierung gehören auch Aufgaben, die im Teilprozess „Verarbeiten“ überwiegend qualitatives Denken und Schlussfolgern erfordern und nicht nur das Abarbeiten von festen Verfahren.“ (ebenda)

Nun deutet sich bereits hier die Schwierigkeit an, Aufgaben eindeutig einer Kompetenzklasse zuzuordnen, denn viele Aufgaben können sowohl durch rechnerische als auch durch begriffliche Modellierung gelöst werden. Die PISA-Gruppe entschied sich nun nicht für eine handhabbarere Kategorisierung: „Um solche Probleme weitgehend zu vermeiden, erfolgte die nationale Einteilung der Aufgaben bei PISA durch ein *Rating*, bei dem sich die *Rater* Aufgabenlöser mit „Expertenwissen“ vorzustellen hatten“. (Knoche u.a. 2002, S.164) Damit scheinen die Tester allerdings von vornherein die Chance zu verspielen, zu wissen, was sie eigentlich messen: Das Gemessene ergibt sich gerade aus der Ganzheit der Wege zum richtigen Kreuz bzw. zur richtigen Zahl. Wenn man die Aufgaben lediglich nach einem einzigen Weg – dem des imaginierten Experten – kategorisiert, dann weiß man hinterher weder, was mit der Aufgabe gemessen wird noch was in der Kategorie gemessen wird.
Für die veröffentlichten Aufgaben liegen die durch die PISA-Gruppe vorgenommenen Einordnungen entweder in die drei- oder sogar in die fünfstufige Klasseneinteilung vor (siehe OECD 2003). Die Sinnhaftigkeit dieser Zuordnung kann somit mit Hilfe der Interpretationen diskutiert werden.

5.4. Untersuchungsdesign: Welche PISA-Aufgaben werden interpretiert?

In dieser Arbeit ist es nicht möglich, eine Gesamtanalyse des nationalen Tests der ersten Testrunde

(Test im Jahr 2000[56]) vorzunehmen. Nur eine Teilanalyse ist möglich. Ein erstes Problem ist ein veröffentlichungspragmatisches: Man benötigt für eine Gesamtanalyse den gesamten Test, auch die unveröffentlichten Teile. Die PISA-Gruppe läßt aber keinerlei Veröffentlichung nichtveröffentlichter Aufgaben zu. Dies liegt daran, dass Items, die als Ankeritems zur statistischen Verzahnung der Testhefte und Testrunden dienen, immer wieder verwendet werden. In der Konzeption von PISA ist außerdem nicht vorgesehen, dass diese Ankeritems sich nicht bewähren: Selbst wenn sie sich in einer Interpretation als problematisch erweisen, können sie nicht verändert werden, weil dann der Ankercharakter wegfällt. Wenn sich also bei den PISA-Ankeritems wie bei den TIMSS-Ankeritems herausstellte, dass fünf von sechs Aufgaben eigentlich verändert werden müßten, dann müßte man dieses Resultat ignorieren, selbst wenn man es einleuchtend fände. Die PISA-Langzeitkonstruktion zwingt also zur Aufrechterhaltung von Konstruktionsfehlern.

Im Lichte des Nichtöffentlichkeitsproblems stellt sich die Frage nach dem Zweck eines solchen aufwendigen Vorgehens in besonderer Schärfe: Eine Gesamtschau würde dazu dienen, *den Gesamttest* einer Bewertung zu unterziehen. Das wäre für die theoretische Seite dieser Arbeit wenig fruchtbringend: Diese Arbeit zeigt zunächst die Probleme der Nichtbeachtung der latenten Textebene für die Meßkraft von Testaufgaben. Sie zeigt Alternativen der Testkonstruktion und der Interpretation von Testergebnissen. Für all das sind keine vollständigen Testanalysen nötig. Bei der Betrachtung von PISA in diesem Kapitel wird sich zeigen, an welchen Stellen die Umsetzung didaktischer Konzepte in Testaufgaben scheitert. Auch dazu genügen Beispiele, die das Prinzip aufzeigen und Alternativen skizzieren.

Eine Gesamtschau des Tests wäre innerhalb der wissenschaftlichen Debatte um PISA scheinbar überzeugender. Dieser Eindruck zerschlägt sich aber: Die in dieser Arbeit vorgebrachten theoretischen Argumente ziehen ihren Überzeugungsanspruch nicht aus der Menge an Aufgabenbeispielen, die untersucht werden, sondern aus dem methodischen Konzept und aus der entweder mehr oder weniger überzeugenden Ausarbeitung jeder einzelnen Interpretation. Niemand, der sich auf die in den wenigen ausführlichen Interpretationen geführte Argumentation nicht einlassen kann, wird dadurch überzeugt, dass der Test als Ganzes durchbuchstabiert wird.

Eine Gesamtanalyse müßte den Test nach bestimmten charakteristischen Polen hin ausleuchten, welche er selbst textlich vorgibt. In der ursprünglich von mir angestrebten Gesamtanalyse hatte ich einen ersten solchen Pol in den (von mir als Startpunkt ad hoc ausgewählten, später aber nicht veröffentlichten) Aufgaben „Schulhof“ und „Matrose“ herausgearbeitet. Dort war es der Pol der Verwerfung: In beiden Aufgaben war die Textstruktur am stärksten dadurch geprägt, dass keine Entscheidung über den Status der Aufgabe getroffen werden konnte, konkret darüber, ob eine reales Anwendungsproblem oder eine eingekleidete Aufgabe oder (bei „Matrose“) gar eine Knobelaufgabe vorliegt - was entsprechendes Irritationspotential birgt. Ausgehend von dieser charakteristischen Verwerfungsstruktur wählte ich die nächste Aufgabe. Dazu begab ich mich an den anderen Pol, suchte also eine äußerlich völlig

[56] Die Entwicklung der PISA-Aufgaben unterliegt einer Dynamik, die wohl erst mit der Erstellung der dritten Testrunde, welche im Jahr 2006 stattfindet, enden wird. Schon deshalb ist es wenig sinnvoll, den jeweiligen Tests „hinterherzuinterpretieren“.

unverworfene Aufgabe. Ich habe dafür z.B. die Aufgabe „Feriengeld“ gefunden, deren Unverworfenheit sich auch in der Analyse bestätigte. Diese Aufgabe erwies sich nun wiederum als inhaltsarm und langweilig im Sinne von intellektueller Armseligkeit. Damit ergibt sich der Gegenpol einer äußerlich interessanten und reichhaltigen Aufgabe, deren Interpretation zum nächsten Charakteristikum führt usw. Dieses Verfahren führt zu einer empirisch gesättigten Gesamtschau des Tests, wobei die Beurteilungskriterien aus den Aufgabentexten selbst rekonstruiert werden, also nicht von außen herangetragen werden müssen.

Sinnvollerweise geht man auch in einer Teilanalyse in der beschriebenen Weise kontrastiv vor. Auch hier ist das Abbruchkriterium für die Fallauswahl wieder die empirische Sättigung: Wenn der durch die Interpretationen geschärfte Blick im Korpus der veröffentlichten Aufgaben keine inhaltliche Herausforderung mehr entdeckt, dann sollte dem Test vorerst genug entlockt sein.

Da die veröffentlichten Aufgaben nur einen Teil des Gesamttests ausmachen, könnte es sein, dass die Teilanalyse weniger Erkenntnis generiert als eine Gesamtanalyse. Schließlich könnte es passieren, dass sich unter den unveröffentlichten Aufgaben noch einige befinden, die auf neue Aspekte verweisen. Es ist aber deutlich, dass auch die Beschränkung auf veröffentlichte Aufgaben eine Struktur untersucht, die wiederum auf ihre Weise den Gesamttest (und damit seine Struktur) widerspiegelt: Die PISA-Gruppe hat entschieden, welche Aufgaben sie als exemplarisch für das von ihr vorgestellte PISA-Konzept und für die Interpretation der Ergebnisse öffentlich zur Diskussion stellen möchte. Man kann also davon ausgehen, dass hier besonders aussagekräftige Aufgaben ausgewählt wurden. Ich beschränke mich natürlich auf jene Aufgaben, die auch wirklich im Test genutzt wurden.

Aufgabeninterpretationen

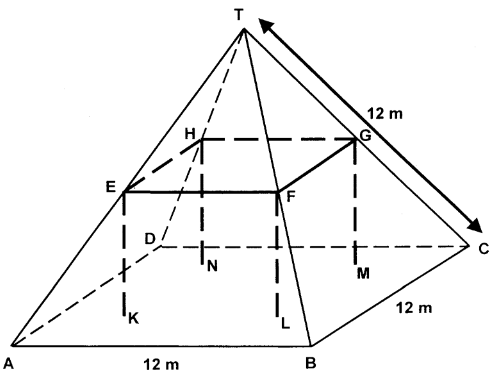
5.5. Die (internationale) PISA-Aufgabe „Bauernhöfe“

BAUERNHÖFE

Hier siehst du ein Foto eines Bauernhauses mit pyramidenförmigem Dach.

Nachfolgend siehst du eine Skizze mit den entsprechenden Maßen, die eine Schülerin vom ***Dach*** *des Bauernhauses gezeichnet hat.*

Der Dachboden, in der Skizze ABCD, ist ein Quadrat. Die Balken, die das Dach stützen, sind die Kanten eines Quaders (rechtwinkliges Prisma) EFGHKLMN. E ist die Mitte von $\overline{AT}$*, F ist die Mitte von* $\overline{BT}$*, G ist die Mitte von* $\overline{CT}$ *und H ist die Mitte von* $\overline{DT}$*. Jede Kante der Pyramide in der Skizze misst 12 m.*



Bauernhöfe 1. Berechne den Flächeninhalt des Dachbodens ABCD.
Der Flächeninhalt des Dachbodens ABCD = ______ m^2.

Bauernhöfe 2. Berechne die Länge von $\overline{EF}$, einer der waagerechten Kanten des Quaders.
Die Länge von $\overline{EF}$ = ______ m.

Lösungswege zu „Bauernhöfe 1“:

Alle Lösungswege laufen darauf hinaus, die Maßzahl der Seitenlänge von 12 m des Quadrats zu quadrieren und 144 m^2 zu erhalten. (Die Einheit des Flächeninhalts ist bereits vorgegeben.) Dies kann im Sinne eines Rechtecks mit zwei gleich langen Seiten geschehen oder im Sinne einer separaten Formel für das Quadrat. Es kann mit oder ohne Betrachtung der Einheit geschehen. Im Prinzip kann man sogar eine Skizze machen und „die Quadratmeter auszählen“. Aus der richtigen Lösung läßt sich ablesen, dass der Schüler entweder die technische Anforderung der Multiplikation bzw. des Quadrierens für eine Flächeninhaltsberechnung bewältigt oder das Konzept des Flächeninhalts im Sinne des Auszählens realisiert hat. Das technische Können kommt ohne das inhaltliche Verstehen aus und umgekehrt, es kann aber auch beides zusammenlaufen.

Lösungswege zu „Bauernhöfe 2“ und ihre Zuordnung zu den PISA-Kompetenzstufen[57]:

- Intuition[58]: E, F, G und H sind Mittelpunkte von Pyramidenseiten. Es liegt nahe anzunehmen, dass

[57] Hier wird versucht, die grundsätzlich denkbaren Wege zur Erlangung des Punktes für die richtige Lösung zu betrachten. Vollständigkeit wird angestrebt, ist aber für die Argumentation nicht notwendig. Die einzelnen Wege werden durch den Aufgabentext zwar gestärkt bzw. geschwächt, das Testziel ist aber lediglich die Erlangung des Lösungspunktes. Jede Möglichkeit, diesen Punkt zu erhalten, verweist auf Bestandteile dessen, was die Aufgabe mißt.
Die Explizierung von Lösungswegen ist immer idealtypisch zu verstehen. Empirisch werden alle möglichen Mischformen vorkommen.

[58] In der mathematikdidaktischen Debatte um Aufgaben spielt Intuition kaum eine systematische und theoretisch reflektierte Rolle und sie findet auch im Kompetenzstufenmodell keinen Platz. Sie ist aber nicht nur Bestandteil vieler Problemlöseprozesse, sondern kann auch als eigenständiger Lösungsweg auftreten. Dabei ist zu beachten, dass es sich bei intuitivem Vorgehen im Kern um Hypothesenbildung handelt. Das macht auch die Unsicherheit intuitiven Vorgehens aus, denn es ist eine Hypothesenprüfung notwendig, die außerhalb des Intuitiven stattfinden muß. Es gibt Aufgaben, bei denen die Hypothesenprüfung so einfach ist, dass man auch sie (fälschlich) als

EF halb so lang wie die gegebenen 12 m der Grundseite AB ist. Man kann das ausgehend von der Intuition auch nachmessen bzw. rechnerisch überprüfen, aber das intuitive Lösen stellt einen eigenen Weg der Lösungsgenerierung dar. In der Kompetenzstufentheorie hat dieser Weg keinen Platz, obwohl auch die PISA-Gruppe ihn sieht. (ebd.; S. 151)

- Messen: Hier kann direkt in der Zeichnung gemessen werden, man kann aber auch vorher eine maßstäbliche Zeichnung anfertigen und in ihr messen. Das Messenkönnen bewegt sich zwar außerhalb dessen, was man mit „mathematisch“ bezeichnet, hat aber seine eigene Bedeutung. Das Messen als empirisches Vorgehen ist hier keine Annäherung an geometrische Operationen, sondern eine Abkehr von ihr, weil es ihre Unnötigkeit zur Erlangung des Lösungspunktes unterstreicht. In jedem Fall ist das Ausmessen einer Strecke eine Fähigkeit, die man auf Stufe I (Wissen auf Grundschulniveau) verweisen würde. Setzt man den Fokus auf die Modellierungsanforderung, könnte man aber die Grenze zu Stufe II schon überschritten sehen.

- Lösen über den Satz von der Mittellinie des Dreiecks: Dazu muß man EF als Mittellinie des Dreiecks ABT erkennen und wissen, dass die Mittellinie eines Dreiecks halb so lang wie seine Basis ist. Dieser Weg verlangt lokales mathematisches Fakten- bzw. Satzwissen. Man würde ihn wohl auf Stufe II (wegen der Einfachheit der Modellierung) oder III (wegen der Verankerung des Wissensinhalts in der Sekundarstufe) einordnen.

- Lösen über den 2.Strahlensatz: Dieser Weg setzt die Erkenntnis voraus, dass die Strecken AT und BT durch ihre Mittelpunkte halbiert sind. Außerdem muß erkannt werden, dass eine Konstruktion vorliegt, die über den Strahlensatz bearbeitet werden kann. Das kann z.B. durch die Anwendung der Umkehrung des Strahlensatzes oder durch eine entsprechende Modellannahme geschehen. Über eine allgemeine Betrachtung oder über das Einsetzen von Zahlen bzw. Variablen gelangt man mit Hilfe des zweiten Strahlensatzes zur Lösung. Dieser Weg verlangt genuin mathematisches Denken. Man würde ihn auf Stufe IV (umfangreiche technische Verarbeitungsprozesse, wenn man mit den Werten arbeitet) oder V (wenn man mit Hälften argumentiert) ansiedeln.

- andere innermathematische Argumentationen: Man kann zum Beispiel die drei Mittelpunkte der Dreiecksseiten miteinander verbinden und dann Kongruenzen der entstehenden Dreiecke betrachten. Man kann aber auch das obere Teildreieck klappen oder spiegeln. Hier sind auf jeden Fall komplexe Argumentationen nötig, die aber z.T. auch rein intuitiv abgesichert sein können. Sie können auch auf einen Beweis des Satzes von der Mittellinie des Dreiecks führen. Dieses Vorgehen verweist auf Stufe V.

Bereits an dieser Stelle wird deutlich, dass die Aufgabe inhaltlich nicht einer einzigen Kompetenzstufe zugeordnet werden kann. Für diese Aufgabe kann eine Kompetenzstufenzuordnung (die PISA-Gruppe ordnet sie Stufe III zu) also nicht bestätigt werden.

intuitiv wahrnimmt. Bei der vorliegenden Aufgabe erfolgt die Hypothesenprüfung auf einem der nachfolgend dargestellten Lösungswege. Man erkennt auch, dass die Unsicherheit des *rein* intuitiven Lösens darin besteht, dass keine Hypothesenprüfung stattgefunden hat. Da es beim Lösen aber nur um das richtige Resultat geht, liegt trotzdem eine vollwertige Lösung vor.

Textanalyse

„Bauernhöfe"

Diese Überschrift (es handelt sich der Form im Testheft nach um eine Überschrift, nicht nur um eine Aufgabenbezeichnung wie im nationalen Testteil) konstruiert einen Widerspruch zum Inhalt, da es in der Aufgabe nicht um Bauernhöfe geht. Man darf etwas erwarten, das mit Bauernhöfen zu tun hat: Statistiken über Bauernhöfe, irgendwelche Rechenaufgaben zu dortigen Lebens- bzw. Arbeitsweisen, vielleicht auch architektonisch-geometrische Betrachtungen zu Bauernhöfen. Wir wissen, dass nichts davon passiert. Es geht lediglich um das **Dach eines** Bauern**hauses**. Es geht also nicht um Bauernhöfe, sondern um **einen** Bauerhof. Konkreter: Es geht nicht um einen Bauernhof, sondern um ein Bauern**haus**. Es geht nicht um ein Bauernhaus, sondern um das **Dach** eines Bauernhauses. Wenn es sich überhaupt um ein Bauernhaus handelt, so hat die Bauernzuschreibung nichts mit dem Problem zu tun. Die Überschrift verspricht also größere Reichhaltigkeit, als die Aufgabe liefert. Der Schüler wird nicht nur - quasi dreieinhalbfach - in die Irre geführt, sondern auch an der Nase herumgeführt. Die Ernsthaftigkeit des Gegenstandes der Aufgabe wird dementiert. Dies kennzeichnet einerseits einen mathematikdidaktischen Habitus, ist aber gleichzeitig wegen seiner Irritationshaltigkeit auch ein Meßproblem.
Habituell ist nun die Frage, warum man hier nicht einfach sagen kann, worum es in der Aufgabe geht: Es ist ja kein „Versehen", dass über einer Aufgabe mit einem Hausdach die Überschrift „Bauernhöfe" steht. Hier zeigt sich zunächst eine romantisierende Tendenz. Es liegt aufgrund der theoretischen Konstruktionen (siehe 3.Ebene „situations and contexts") nahe anzunehmen, dass das bereits kritisierte Konzept der Schülernähe dem gleichen Habitus folgt. Die Romantisierungstendenz erzeugt hier allerdings eine zusätzliche Verwerfung, weil – nähme man das Schülernähekonzept ernst – es viel näher läge, ein Stadthaus zu betrachten. Scheinbar projizieren die Aufgabenersteller die eigenen romantisierenden Bauernhofvorstellungen auch noch auf die Schüler, denen es näher zu kommen gilt. Man würde darüber mehr erfahren, wenn die PISA-Gruppe ihre Aufgabenkategorisierungen vollständig veröffentlichen würde: Sie suggeriert in ihrem Theoriekonstrukt, dass die Aufgaben in verschiedene Stufen der Schülernähe eingeteilt wären (OECD 2000, S.50). Es wäre interessant zu erfahren, ob ein Bauernhof dort einen höheren „Schülernäheindex" erhält als ein Stadthaus.

Hier siehst du ein Foto eines Bauernhauses ...

Die Aufmerksamkeit des Lesers wird auf das Foto gelenkt. Es hat damit nicht nur illustrativen Charakter, sondern direkten Bezug zum Text. Dieser Bezug wird im weiteren nicht eingelöst, das Bauern(?)haus selbst spielt im Folgenden keine Rolle mehr. Sowohl das Foto selbst als auch die Verbindung zum Text produzieren einen Anspruch, der nicht eingelöst wird. Das wäre z.B. möglich, indem die folgenden Geometrieaufgaben irgendeine inhaltliche Rückführung auf das Haus erhalten würden. Diese Logik der Nichtbedeutung des Gegenstandes (die sich bereits mehrfach in der Überschrift fand) findet sich weiter in der Formulierung *ein Foto eines*: Wohlgeformt wären *... das Foto eines ...* (Spezifizierung des Fotos, Entspezifizierung des Gegenstandes) oder *... ein Foto des ...* mit nachfolgender Spezifizierung des Hauses (Entspezifizierung des Fotos, Spezifizierung des Gegenstandes). Die gewählte Formulierung schafft eine doppelte Entspezifizierung, zugespitzt: Es ist völlig egal, welches Foto welches Hauses du hier siehst.

... mit pyramidenförmigem Dach.

Die klare Hinführung zur Pyramidenform schafft einen Fokus. Kontrastierend kann man nach *mit* jedes Sujet einführen. Man erkennt, dass der Fokus weg vom Haus hin zum Sujet gelenkt wird. Man kann auch noch Kombinationen mit den Kontrastvarianten zu ... *ein Foto eines* ... betrachten. Die sprachliche Verwerfung zwischen der manifesten Bedeutung des realen Gegenstandes „Haus“ und der latenten Vernichtung dieser Bedeutung wiederholt sich hier.
Am Ende des ersten Satzes kann man den objektiven Aufgabentext polemisch zusammenfassen: Wir tun mal so, als ob wir über Bauernhöfe reden. Du siehst hier irgendein Foto irgendeines Hauses. Du sollst dich auf die Pyramidenförmigkeit seines Daches konzentrieren.

Nachfolgend siehst du eine Skizze mit den entsprechenden Maßen, ...

Auffallend ist hier das *entsprechend*. In welchen Situationen sieht man *eine Skizze mit <u>entsprechenden</u> Maßen*? Ausschließlich in Situationen, in denen vorher klar ist, wozu man die <u>entsprechenden</u> Maße benötigt. Das kann zur praktischen Nutzung (Bau, Rechnung) sein, aber auch um darüber nachzudenken. Ein Beispiel: *Hier siehst das Foto eines Werkstücks. Nachfolgend siehst du eine Skizze mit den entsprechenden Maßen.* Es ist deutlich, dass zwischen den beiden Aussagen etwas fehlt: *Du sollst es bauen.* oder *Du sollst eine technische Zeichnung davon erstellen.* Wenn dieser Satz fehlt, dann ist genau die textlich vorliegende Kombination der Ersatz für diese Mitteilung.
Für unsere Aufgabe liegt hier also etwas Unausgesprochenes vor: *Du sollst mit den Maßen etwas machen.* Mit der bereits geleisteten Interpretation wird auch deutlich, warum das nicht explizit da steht: Die Mitteilung hätte keinen sprachlichen Anschluß an den bisherigen Text. Manifest steht ja nur da, dass man ein Foto eines Bauernhauses mit pyramidenförmigem Dach sieht. Die latente Mitteilung „Das Wichtige ist die Pyramidenförmigkeit, alles andere interessiert im Grunde nicht.“ liegt nicht offen.
Wohlgeformt wäre z.B. der Text: *In dieser Aufgabe geht es um ein pyramidenförmiges Dach. Nachfolgend siehst du eine Skizze mit den entsprechenden Maßen.* Das Problem mit diesem Text wäre nur, dass er nicht so tun würde, als ob er sich spannenderweise mit Bauernhöfen und mit Fotos realer Häuser befaßte. Mit diesem Text würde aber die Implikation einer praktischen, technischen Problemlösung transportiert, die mit einem mathematisch konsistent formulierten Problem zusammenlaufen könnte.

..., die eine Schülerin vom ***Dach*** *des Bauernhauses gezeichnet hat.*

Der Schüler wird erneut an der Nase herumgeführt, denn natürlich hat keine Schülerin die Skizze gezeichnet, sondern der professionelle Grafiker, den die PISA-Gruppe dafür bezahlt hat. Hier soll offenbar wiederum ein Schülerbezug konstruiert werden - der sich gleich selbst dementiert. Wahrscheinlich kann man die Aufgabe durch diesen „Schülerbezug“ auf der dritten PISA-Konstruktionsebene „Situa-

tions and contexts“ in die Kategorie „daily life“ einordnen. Auch diesbezüglich wäre die Veröffentlichung der Zuordnungen der Aufgaben zu den Kategorien interessant: Steigt der Schülernäheindex, wenn man behauptet, eine Schülerin hätte hier gezeichnet? Und steigt er noch mehr, wenn man wirklich eine Schülerin zeichnen läßt? Oder steigt dann nur noch der Wirklichkeitsindex?
Die Schlinge, die latent schon gelegt war, wird nun auch manifest langsam zugezogen: Wir sind ausgehend von den Bauernhöfen über das Bauernhaus beim Dach gelandet. Spätestens jetzt wird deutlich, dass das bisherige Manifeste nur schmückendes Beiwerk war. Jetzt kommt noch einmal der Dachboden, danach ist die Einkleidung vergessen, man benötigt sie nie wieder - das sieht man jetzt bereits ohne Analyse der latenten Textebene. Selbst in den Fragen wird die Einkleidung nur noch der Dekoration dienen. Die Illusion eines ernstzunehmenden realen Gegenstandes ist aber bereits hier zerstört.

Zeichnung/Räumlichkeit des Problems

Die „Skizze“, die keine Skizze ist, sondern eine Zeichnung, stellt ein räumliches Problem dar. Die Räumlichkeit schwächt alle Lösungswege: Der Weg übers Messen wird geschwächt, weil eine räumliche Darstellung Verkürzungen birgt. Es verlangt einen separaten Denk- und Selbstüberzeugungsschritt, um herauszufinden, ob das die Seitenverhältnisse verzerrt. Die Schwächung wird gemindert, weil bei allen vier Seiten wirklich die Hälftigkeit der „Balken“ gemessen werden kann.
Die Wege über den Strahlensatz und über die Mittellinie werden ausschließlich geschwächt, denn diese Wege verlangen ein Denken in der Ebene. Dazu muß man das Dreieck in die Ebene verlagern oder als Ebene denken, in der man sein Wissen anwenden kann.
Der Weg über die Intuition wird geschwächt, weil die Zeichnung viel mehr Elemente enthält, als wenn nur das ebene Gebilde ABEFT angebildet würde. Schon diese optischen Störsignale müssen gefiltert werden. Es hilft auch wenig, sich in das Dach hineinzudenken, um die Intuition zu stärken.
Eine sehr große Stärkung für die Intuition bringt hingegen die Tatsache, dass es sich um ein gleichseitiges Dreieck handelt. Für diese Dreiecksart ist die richtige Lösung ausgesprochen suggestiv.

Was bedeutet es, dass hier eine Zeichnung als Skizze bezeichnet wird? Die Umbenennung konstruiert im Vergleich zur tatsächlich vorliegenden Zeichnung Spontaneität, Entlastung von Aufwand, Genauigkeitsverringerung. Eine gewisse Lockerheit wird dadurch konstruiert, dass man eine Skizze direkt aus einer Idee bzw. von einem realen Gegenstand ausgehend erstellt, in diesem Fall direkt am Haus bzw. aus den Konstruktionsplänen. Man verringert den Zeichenaufwand, nicht aber den Meßaufwand: Will man mit der Skizze weiterarbeiten, so geben die Maße und nicht die Skizze vor, wie genau man dabei arbeiten kann.
Die Genauigkeitsverringerung und die erhöhte Rolle der Maße betrifft die Meßeigenschaft der Aufgabe. Die anderen Konstruktionen sind habitueller Art: Es wird Spontaneität und das Ausgehen vom realen Gegenstand konstruiert. Beides ist aber eben wirklich nur konstruiert: Weder wurde skizziert, noch ist diese Aufgabe entstanden, weil jemand eine Frage bezüglich des fotografierten Daches hatte. Beides wird wahrheitswidrig behauptet. Habituell bedeutet das, dass Spontaneität und die Verankerung des Problems im Realen so wichtig sind, dass sie eine Abkehr von der Wahrheit rechtfertigen.

Der Text ist bis hier bereits so stark in seinen Verwerfungen von behaupteter realitätsabgeleiteter, schülerorientierter Konstruktion einerseits und Testaufgabencharakter andererseits verstrickt, dass ich in der weiteren Interpretation diesen verworfenen Hintergrund ausblenden möchte[59].

Der nun folgende Textteil konstruiert auch manifest eine Mathematikaufgabe. Wenn er für sich interpretiert wird, ist damit die im ersten Aufgabenteil gezeigte Struktur falsifizierbar. Man kann dazu z.B. folgenden Vortext annehmen:

In dieser Aufgabe geht es um ein pyramidenförmiges Dach. Nachfolgend siehst du eine Skizze mit den zugehörigen Maßen.

Der Dachboden, in der Skizze ABCD, ist ein Quadrat.

Der Begriff Dachboden (Raum zwischen dem obersten Geschoß und dem Dach eines Gebäudes) wird hier regelabweichend verwendet. Für die Figur ABCD gibt es in der Umgangssprache lediglich die Bezeichnung „Boden des Dachbodens“, die hier wahrscheinlich vermieden werden sollte. Man könnte auch vom „Boden des Daches“ oder vom „Boden des Dachgeschosses“ sprechen. Alle vier Formulierungen enthalten Irritationspotential, es ist nicht vermeidbar, wenn man unbedingt bei diesem Gegenstand verbleiben will.

Weder der Dachboden noch der Boden des Dachbodens *ist* ein Quadrat. Kein Dachboden ist ein Quadrat. Und kein Boden eines Dachbodens ist ein Quadrat. Er kann höchstens quadratisch sein.

Der Unterschied zwischen beiden Formulierungen läßt uns erkennen, wie mathematische Fachsprache in die Umgangssprache eingeflossen ist: Wenn ein Kind sich im Verlauf der Sprachentwicklung das Wort „quadratisch“ aneignet, so hat es bereits ein „Vorkonzept“ von einem Quadrat entwickelt. Es weiß dann bereits - eher implizit als explizit -, dass Etwas, das man quadratisch nennt, vier Seiten hat, dass diese gleich lang sind und dass sie senkrecht aufeinander stehen. All diese Eigenschaften werden aber im Verlauf mathematischer „Weiter-Bildung“ erst expliziert und fließen in den Begriff des Quadrats. Man wird also - und Experimente sind hier leicht vorstellbar - Kinder finden, die das Wort *quadratisch* verstehen bzw. verwenden können, die eventuell sogar das Wort *Quadrat* verstehen bzw. verwenden können, ohne dabei über ein Konzept des Quadrates in mathematischen Begriffen zu verfügen. Ich nenne das „umgangssprachliches Konzept des Quadrats“. Daneben gibt es ein Konzept des Quadrats *in mathematischen Begriffen*, etwa: Ein Quadrat ist eine geometrische Figur, bestehend aus vier gleich langen Seiten, von denen die je zwei benachbarten senkrecht aufeinander stehen. „Mathematisierung“ des Quadratbegriffs bedeutet hier, die Alltagsvorstellungen in theoriegeleitete Begrifflichkeiten zu überführen und deren Besonderheiten (Idealisierung, Nichtempirie usw.) zu kennzeichnen.

Nun ist klar, dass es Zwischenstufen geben kann, auf denen der umgangssprachliche Begriff schon überwunden ist, ohne dass bereits ein explizit in mathematischen Begriffen arbeitendes Konzept vorläge. Auf dieser Zwischenstufe würde ein Kind zu einer quadratischen Gehwegplatte schon nicht mehr

[59] Wenn man das nicht tut, stellt man immer wieder Widersprüche des Texts zu den vorangegangenen manifesten Behauptungen fest. Das brächte hier aber keine neuen Erkenntnisse.

sagen, sie sei ein Quadrat. Es würde sagen, sie sei quadratisch.[60] Es könnte aber noch nicht begründen, warum es nicht vom Quadrat redet. Bereits auf dieser Stufe von Sprachentwicklung würde aber der Begriff *Quadrat* bereits in das Reich der Mathematik verweisen, wohingegen *quadratisch* ins Reale gehört. Der Aufgabentext bewegt sich **unterhalb** dieser Entwicklungsstufe und dieses Abstraktionsniveaus.

Diese Erkenntnis läßt sich in kontrastierenden Textvarianten vertiefen:
1. Der Boden des Dachbodens ist/sei quadratisch.
2. ABCD ist/sei ein Quadrat.
3. ABCD ist/sei quadratisch.
4. Der Boden des Dachbodens ist/sei ein Quadrat.
5. Der Boden des Dachbodens, in der Zeichnung ABCD, ist/sei ein Quadrat/quadratisch.

Auffallend ist die mangelnde Wohlgeformtheit von 3. Es ist eine besonders grobe Regelabweichung, wenn eine geometrische Figur als „quadratisch“ bezeichnet wird, nicht nur umgekehrt die Benennung eines realen Gegenstandes als Quadrat (Variante 4). Mit der vorliegenden Formulierung bewegt sich die Aufgabe auf einer Sprachebene, auf der der Unterschied zwischen Quadrat und quadratisch noch nicht realisiert ist. Der vorliegende Text stellt sozusagen eine Mesalliance der wohlgeformten Gelingensvarianten 1 und 2 dar. Hier wird also das Mathematische im Begriff „Quadrat“ zerstört, es wird aber auch das Reale zerstört, indem es in seiner Autonomie verleugnet und als Mathematisches bezeichnet wird. Eine für die Lösung hilfreiche Hinführung des Realen zum Mathematischen, also Hilfestellung durch Vorausführung von Teilen des Modellierungskreises, findet hingegen nicht statt. Auch eine Positionierung des Mathematischen zum Realen bzw. des Realen zum Mathematischen findet nicht statt.

Die konstruierten Varianten lassen auch eine Untersuchung der vorliegenden Modellierungsanforderung zu: Die höchste Modellierungsanforderung würde durch das Weglassen dieser Information entstehen. Der Schüler müßte dann einerseits selbständig seine Aufmerksamkeit auf die Fläche bzw. Figur richten und er müßte selbständiger ermitteln, um welche Fläche es überhaupt geht. 1 bis 5 vermitteln im Vergleich dazu eine zusätzliche Information, nämlich die Information des Quadratischseins. Die Mitteilung des quadratischen Charakter selbst ist für die Flächeninhaltsbestimmung eigentlich überflüssig, denn mitteilungsbedürftig wäre für ein reales Problem lediglich, wenn die Hauswände *nicht* rechteckig zueinander stünden, wenn der Boden also *nicht* quadratisch wäre. Die Angabe der Längengleichheit der Bodenseiten ist zusätzlich redundant.

Die vorgeschlagenen Formulierungen und mit ihnen der Aufgabentext erweisen sich somit als überflüssig. Sie antworten auf eine nicht gestellte und auch sachlich nicht naheliegende Frage: Stehen die Seiten des Vierecks senkrecht aufeinander? Mit der Antwort auf diese aus gutem Grunde nicht gestellte Frage wird die Realität nicht ernst genommen - denn die Untersuchung der Möglichkeit seiner Schiefheit ist absurd und konstruiert. Eine eventuelle Schiefe müßte explizit mitgeteilt werden.

[60] Diese Exemplifizierung der Zwischenstufe stellt nur eine These dar. Man mag das Bedürfnis verspüren, sie zu überprüfen. Zum Beispiel ist ebenso denkbar, dass der Unterschied zwischen dem Substantiv und dem Adjektiv nicht erst mit der genaueren Kenntnis des Objekts „Quadrat“ auftritt, sondern sich bereits vorher erschließt - dann bewegt sich der Aufgabentext ebenfalls vor dieser Stufe der Sprachentwicklung.

Auch der Modellierungsgedanke wird nicht ernst genommen - es wird zwar so getan, als ob hier eine Modellierungsbedingung mitgeteilt würde, aber es handelt sich nicht um die Herstellung einer Verbindung zwischen Realität und Modell, sondern um eine in beiden „Welten“ überflüssige Information. Eine mögliche Erklärung für diese überflüssige Mitteilung mag darin liegen, dass der Pyramidenbegriff in der Mathematik Körper mit beliebigen konvexen n-Ecken als Grundfläche einschließt. Das macht die Mitteilung aber noch abstruser, weil unter Behauptung mathematisch präziser Information lediglich unterstellt wird, der Schüler könnte ein „pyramidenförmiges Dach“ anders als quadratisch modellieren. In dieser ursprünglich unterrichtlich orientierten Aufgabe zeigt sich darin ein Habitus der Vermeidung von Provokationen[61]. In einer Testaufgabe führt dies zwar äußerlich zur Vermeidung nicht gewollter, also „falscher“ Modellierungen, erzeugt aber im Gegenzug die hier diskutierten Verwerfungen.

In den Varianten 1 bis 4 wurde keine Verbindung zwischen der gesuchten Realgröße und der Zeichnung hergestellt - diese Verbindung findet sich in Variante 5, welche sich an den vorliegenden Text anlehnt. Zum Herausarbeiten dieser Verbindung erweist sich die mit *ist ein Quadrat* festgestellte Zerstörung des Mathematischen als sachlich gänzlich überflüssig. Unter Reduktion auf diesen Zweck ergeben sich weitere Formulierungsalternativen:

6. Der Boden des Dachbodens ist/sei in der Zeichnung ABCD.
7. Das Quadrat ABCD in der Zeichnung ist/sei der Boden des Dachbodens.
8. Die Figur ABCD in der Zeichnung ist/sei der Boden des Dachbodens.
9. ABCD ist/sei der Boden des Dachbodens.
10. In der Zeichnung ist/sei ABCD der Boden des Dachbodens.

Jede der Alternativen richtet den Blick auf einen anderen Schwerpunkt der Betrachtung (und kann mit *quadratisch* ergänzt werden). Der Aufgabenkonstrukteur muß sich entscheiden. Was hier in jedem Falle konstruiert würde, ist eine Reduzierung der Modellierungsanforderung: Der Schüler muß den Übergang vom Realen zum Mathematischen nicht mehr leisten. Es zeigt sich, dass die Punktbezeichnungen in der Zeichnung diese Anforderungsreduzierung vorbereitet hat, hier wird sie vollendet, denn der Übergang wird vom Text vollständig geleistet.

Der vorliegende Aufgabentext vereint nun einerseits den Quadrathinweis - also den in sich ja nicht besonders nahe liegenden Ausschluß des Parallelogrammgedankens - und den Verzicht auf die Anforderung des Übergangs vom Realen zum Mathematischen. Damit verbleibt aber keine Modellierungsanforderung mehr, sie ist völlig unterlaufen. Jede nun folgende Aufgabenstellung lautet im Kern nur noch: Berechne den Flächeninhalt des Quadrats.

[61] In der Wahrnehmung der Schüler ist es oftmals willkürlich vom Lehrer bestimmt, welche Modellierung für ein Problem verwendet wird. Der Schüler kann nun mit dieser Willkür spielen, indem er andere Modellierungen vorschlägt. Dazu muß er in diesem Fall die offensichtlich gemeinte Alltagsdeutung von „pyramidenförmig“ verzerren und auf der mathematischen Deutung des Pyramidenbegriffs bestehen. Diesem notwendigen Infragestellen des Gegebenen kann man eigentlich nur mit professioneller Gelassenheit begegnen bzw. sie produktiv nutzen - indem man andere Modellierungen diskutiert. Man kann aber auch wie im Aufgabentext versuchen, solchen Provokationen vorzubauen. Damit produziert man dann Verwerfungen wie in dieser Aufgabe, die wiederum zu Provokationen Anlaß geben.

Der Text erweist sich unter dem kontrastierenden Blickwinkel als sprachlich schwierig. Die Reduzierung des mathematischen bzw. modellbildenden Anspruchs ist also mit einer künstlichen sprachlichen Verschwierigung verbunden. Die vorliegende Zerstörung des Mathematischen sowie der Modellierungsanforderung ist also nicht durch eine sprachliche Vereinfachung oder durch eine sachliche Vereindeutigung zu erklären.

Die Balken, die das Dach stützen, sind die Kanten eines Quaders (rechtwinkliges Prisma) EFGHKLMN.

Hier liegt zunächst eine Falschaussage vor. Die Balken sind nicht *die* Kanten dieses Quaders, sie sind Kanten dieses Quaders. Das Hinzufügen des *die* ist also eine Entexaktifizierung. Man kann weder behaupten, dass dies der Realsituation geschuldet wäre noch dass es eine sprachliche Vereinfachung darstellen würde.

Eine Konstruktion wie *... eines Quaders (rechtwinkliges Prisma) ...* schafft offensichtlich Irritationspotential: Wer nicht weiß, was ein Quader ist, weiß mit Sicherheit auch nicht, was ein rechtwinkliges Prisma ist, hingegen gibt die Tatsache des Angebens der Information „rechtwinkliges Prisma" eine zusätzliche Aufgabe: Finde heraus, ob diese Information relevant für das Finden der richtigen Antwort ist. Der Schüler weiß das ja zunächst nicht und kann es bei Unkenntnis des Begriffs Prisma nur schwer entscheiden - hier entsteht also Irritationspotential. Wenn er den Begriff kennt, so steht die Frage „Wozu wird das mitgeteilt?" Schließlich ist für einen Schüler, der den (ausgesprochen schwierigen) Begriff Prisma kennt, trivial, dass ein Quader rechtwinklig ist. Für ihn liegt hier also ohne zunächst sichtbaren Grund eine überflüssige Information vor. Selbst für den Schüler, der den Begriff kennt, entsteht also Irritationspotential. Sollte die Klammer dazu dienen, Redundanz zu schaffen und die Aufgabe damit zu erschweren, so muß man fragen, zu welchem Zweck das dient, denn die Redundanz liegt nicht im Problem begründet wie bei offenen Problemen. Sie ist künstlich hergestellt und spricht keine auf die Problemlösung bezogene Fähigkeit an, sondern nur die Fähigkeit, eine künstliche Irritation zu überwinden, also Testfähigkeit. Die künstliche Verschwierigung ist aber nicht nur ein Meßproblem, sondern auch Ausdruck eines Habitus.

Beim vorhergehenden Satz stellte sich heraus, dass er die Verbindung zwischen den Bezeichnungen ABCD in der Zeichnung und dem Boden des Dachbodens im realen Dach herstellen sollte und dass der Verweis auf den quadratischen Charakter für die Problemlösung eigentlich überflüssig war. Was soll nun dieser Satz leisten? Wofür benötigt man die Information?

Für das innermathematische Arbeiten benötigt man auch diese Information nicht. Hier stellt sie eine erhebliche Erhöhung der Schwierigkeit durch viele überflüssige Informationen her. Man benötigt nicht mal die Information, dass auch EFGH rechtwinklig ist, wenn man EF bestimmen will. Diese Tatsache stärkt aber den intuitiven Lösungsweg, da ein „symmetrisches" Denken nahegelegt wird, es wird die Gleichwertigkeit aller vier Seiten der Pyramide expliziert - im Rückblick zeigt sich damit auch ein möglicher latenter Grund für die eigentlich überflüssige Angabe des quadratischen Charakters von ABCD.

Man könnte dem entgegenhalten, dass auch in einem „echten“ Modellierungsvorgang überflüssige Informationen erhoben und diskutiert werden und dass genau dieser Vorgang des Fehllaufens zum Modellieren gehört. Dieser notwendige Vorgang ist aber etwas völlig anderes als die quasi kalkulierte, künstliche Fehlleitung des Schülers durch überflüssige Informationen. Die Aufgabe für den Schüler besteht dann ja nicht mehr darin, in einer authentischen realitätsbezogenen Situation mathematisch zu argumentieren und dabei Fehlwege zu erkennen und zu überwinden. Die Aufgabe für den Schüler besteht darin, künstlich vorgegebene, nicht in der Sache liegende Fallen zu umschiffen. Die Fähigkeit, diese Anforderung zu bewältigen, verliert ihren Sinn mit dem Verlassen der Testsituation.
Manifest wird in diesem Satz eine Verbindung zwischen dem Dach bzw. seinen Stützbalken und dem mathematischen Modell gezogen. Die dabei vermittelte Information ist für die mathematische Seite der Problemlösung nicht nur überflüssig, sie ist auch nicht hilfreich. Es liegt keine echte Verbindung zwischen der Realität und dem Modell als Modell für eine Problemlösung vor. Es findet also gar kein Modellierungsvorgang statt, er wird nur behauptet. Es wird aber nicht nur die Modellierungsanforderung unnötig zerstört, sondern auch die Authentizität des Realen an sich: Balken mit einem Querschnitt von 15 cm mal 15 cm können niemals Kanten eines Quaders „sein“. Sie sind selbst Quader und man fragt sich sofort, welche Kante dieses Quaders eigentlich die Kante des Quaders EFGHKLMN sein soll. Daran schließt sich dann auch die Frage an, welche Länge des Balkens eigentlich gemeint ist - er muß ja in Wirklichkeit sehr kompliziert gesägt sein, um in die Dachkonstruktion eingepaßt werden zu können.

Oben wurde herausgearbeitet, dass die Behauptung *Der Dachboden ... ist ein Quadrat.* den Text auf eine Sprachstufe vor jedem mathematischen Verständnis des Begriffs Quadrat stellt, auf der der Begriff Quadrat aber noch gar nicht verfügbar ist. Wird auch mit der Behauptung *Die Balken, die das Dach stützen, sind die Kanten des Quaders EFGHKLMN.* das Mathematische mißachtet und noch vor einen mathematischen Begriff zurückgeworfen? Die sprachliche Konstruktion wirkt manifest gleichartig. Gleichwohl ruft man nicht wie beim Dachboden sofort aus: Balken können keine Quaderkanten sein. Der umgangssprachliche Begriff des Quaders scheint also vom mathematischen Begriff nicht so weit entfernt zu sein. Zur näheren Untersuchung kann man sich der vorliegenden Konstruktion gedankenexperimentell annähern:

1. Die Balken, die das Dach stützen, werden als Kanten des Quaders EFGHKLMN modelliert.
2. Die Stützbalkenkonstruktion EFGHKLMN wird als Quader modelliert.
3. Die Stützbalkenkonstruktion wird als Quader EFGHKLMN modelliert.
4. Die Stützbalken werden als Quader EFGHKLMN modelliert.
5. Die Stützbalken werden als Kanten eines Quaders/des Quaders EFGHKLMN modelliert.
6. Die Stützbalken bilden einen Quader/den Quader EFGHKLMN.
7. Die Stützbalken sind ein Quader/der Quader EFGHKLMN.
8. Die Stützbalken sind Kanten eines Quaders.
9. Die Stützbalken sind Kanten des Quaders EFGHKLMN.

Variante 1 nimmt die vorliegende Formulierung auf und konstruiert einen expliziten Modellierungsgedanken. Er wäre auch durch *wird/werde als Quader angesehen* gegeben. Die Hölzernheit jeder solchen Explizierung ist sichtbar. Mit Variante 1 wird zwar deutlich, dass im Aufgabentext keine solche Explizierung vorliegt, zur Untersuchung des Quaderbegriffs muß der Satz aber weiter kontrastiert werden.
2 und 3 arbeiten nicht mit den Balken, sondern mit einer Gesamtkonstruktion. Sie ermöglicht ein Spiel mit der Grenze zwischen Realmodell und mathematischem Modell: In 2 wird die Punktzuordnung vor der Modellierung vorgenommen, in 3 nach der Modellierung. Beides ist wohlgeformt möglich. Das zeigt, wie bereits das Einfließen von mathematischen Bezeichnungsweisen Realmodell und mathematisches Modell einander annähern kann. Da dieses Spiel aber auch für das Quadrat aus dem vorherigen Satz funktioniert, erzählt es nichts über die Nähe von mathematischem und umgangssprachlichem Quaderbegriff. Es zeigt uns die Wichtigkeit der Tatsache, dass in der Aufgabe von den Balken, nicht aber von dem aus den Balken zusammengesetzten Gesamtkonstrukt die Rede ist, denn mit den Stützbalken funktioniert diese sprachliche Konstruktion nicht. Das zeigt z.B. Variante 4. Sie ist nicht wohlgeformt, da die Balken nicht als Quader modelliert werden können. Die wohlgeformte Heilung liefert Variante 5, 6 ist eine wohlgeformte Alternativvariante. 7 ist wie 4 nicht wohlgeformt, da Balken noch nicht der Quader sind. Das heilen 8 (nimmt die Wohlgeformtheit von 5 auf, wobei der Modellierungsgedanke weg ist) und 9 (Spezifizierung von 8). Wir sind nun bei zwei vereinfachten Varianten des vorliegenden Aufgabentextes angekommen. Sie sind beide wohlgeformt.
Den Übergang vom Realen zum Mathematischen sehen wir nun im direkten Vergleich:

5. Die Stützbalken werden als Kanten eines Quaders modelliert. - Dieser Quader ist ein mathematischer Quader.
6. Die Stützbalken bilden einen Quader. - Dieser Quader ist ein umgangssprachlicher Quader.
8. Die Stützbalken sind Kanten eines Quaders. - Bei diesem Quader bleibt unklar, in welcher Welt er sich bewegt. Und die beiden obigen Quader zeigen auch, warum das so ist: Der mathematische Quader ist etwas zu modellierendes. Der umgangssprachliche Quader wird von realen Gegenständen gebildet. Wer den Begriff verwendet, weiß zumindest implizit, dass bei diesem Körper gegenüberliegende Flächen gleich große Rechtecke sind und dass die Balken senkrecht aufeinander stehen - und dass diese Eigenschaften empirisch sind, weshalb der Einwand „Das kann doch gar kein Quader sein!“ eben auch als schulmeisterlich und nicht wohlgeformt angesehen würde.
Ein umgangssprachlicher Quader hat nun aber keine *Kanten*, er wird auch nicht von Kanten gebildet. In dieser Kontrastierung erkennt man, dass man sich auf den mathematischen Quader bezieht, wenn man von Kanten spricht. Diese Abstraktionsleistung ist dadurch kaum sichtbar, dass der umgangssprachliche Begriff sich aus dem mathematischen ableitet und nah an ihm geblieben ist. Deshalb ist es auch möglich, Balken als Kanten zu betrachten. Es läßt eben nur im Unklaren, in welcher „Welt“ man sich gerade befindet. Diese Unklarheit wiederholt die Struktur, nach der hier gar keine Modellierungshandlung vorzunehmen ist. Realität und Modell sind bereits „ineinandergeschoben“. Es liegt aber keine sprachliche Zerstörung des Mathematischen vor wie beim Quadrat. Der Grund liegt darin, dass der empirische Quaderbegriff in der Umgangssprache vorhanden ist, wohingegen ein empirischer Quadrat-

begriff in der Umgangssprache eher durch „quadratisch“ ersetzt wird.
Führt man die eben vorgenommene Kontrastierung mit „dem Quader EFGHKLMN“ durch, zeigt sich eine Verschiebung der Akzentuierung ins Mathematische. Selbst der Alltagsquader übersteht diese Verschiebung ohne wirkliche Verwerfung.
Der vorliegende Aufgabentext wählt die sprachlich schwierigere, gleichwohl kaum verständlichere Variante *Die Balken, die das Dach stützen, sind ~~die~~ Kanten des Quaders EFGHKLMN.,* um die bis hierhin für die Aufgabenlösung im wesentlichen überflüssige Information zu geben.

In diese Konstruktion platzt nun die Klammer *(rechtwinkliges Prisma)*. Mit der Klammer ist der Quader nun eindeutig in der Mathematik angesiedelt. Damit ist auf die denkbar ungeschickteste Weise (siehe Diskussion der Klammer oben) ein Übergang von den realen Stützbalken zum mathematischen Quader geschaffen. Aber auch dieser „brachiale“ Übergang zum Mathematischen zerstört wieder den Modellierungsgedanken: Man kann den Satz vereinfacht so lesen: *Die Balken sind Kanten des rechtwinkligen Prismas EFGHKLMN.* Dieser Satz dementiert nun ganz offen eine Modellierungsnotwendigkeit. In dieser Konstruktion existieren quasi keine Balken mehr. Die Klammer reproduziert also die Logik der Zerstörung der Modellierungsanforderung unter dem äußeren Anschein fachlicher Präzisierung, die keine Präzisierung ist. Ein inhaltlicher Grund für diese sprachliche und inhaltliche Aufblähung und Irritationseinführung zeigt sich auch an dieser Stelle nicht. Es scheint sich um ein reines Habitusproblem zu handeln.

E ist die Mitte von $\overline{AT}$*, F ist die Mitte von* $\overline{BT}$*, G ist die Mitte von* $\overline{CT}$ *und H ist die Mitte von* $\overline{DT}$.

Die Wiederholung der immer gleichen Information wiederholt die bereits diskutierte Struktur der für die Lösung überflüssigen bzw. redundanten Information, die das intuitive Vorgehen stärkt.

Der Begriff *Mitte* bewegt sich innerhalb von Umgangssprache, ist hier gleichwohl in die geometrische Terminologie der Punkte und Strecken eingebunden. Das ist äußerlich zunächst irritierend, denn ausgesprochen steht dort ja: Der Punkt E ist die Mitte der Strecke AT. In konsequent geometrischer Terminologie würde dort Mittel*punkt* stehen. Trifft man auf Punktbezeichnungen wie E, insbesondere aber auf die Streckenterminologie $\overline{AT}$, so liegt ein Verweis auf geometrische Terminologie vor. Schauen wir uns das näher an:
Bei Dachbalken kann man von einer Mitte sprechen. Diese Mitte ist etwas anderes als ein mathematischer Mittelpunkt, sie ist eine *empirische Mitte*. Sie hat eine Breite, die von der Anforderung abhängt: Will man unter der Mitte eines Dachbalkens einen Stuhl positionieren, so kann diese Mitte durchaus 10 cm breit sein. Will man einen Nagel in die Mitte eines Balkens schlagen, so ist diese Mitte nur einen Millimeter breit. Bezeichnet man diese Mitte nun innerhalb der Bezeichnungsregeln für Punkte, so ist damit nicht der Dachbalken geleugnet, sondern es findet ein Modellierungsschritt statt - nämlich eine Mathematisierung. Gleiches gilt für die Streckenbezeichnungen. Das zeigt, dass dabei nicht einfach Fachsprache in die Umgangssprache fließt, sondern dass dieser Übergang auch eine gewisse (Vor-)Form von Fachlichkeit herstellt: Auch wenn dieser Punkt und diese Strecke noch empirisch gedacht werden, so liegen sie doch schon nahe an einem geometrischen Umgang mit dem Gegenstand.

Das Seltsame an der Verwendung von *Mitte* statt *Mittelpunkt* bleibt damit aber bestehen: Mit der Verwendung der Punktbezeichnungen E, A und T ist der Gedanke von Punkten ja bereits dreifach eingebracht. Die Verwendung von *Mitte* wird damit nicht zu einem Verzicht auf den Terminus *Mittelpunkt*, sondern sie ist ein Schritt zurück in die Realsituation - als ob jemandem plötzlich eingefallen ist, dass man hier doch nicht einfach vom Mittelpunkt sprechen kann. Hier wird verweigert, sich klar innerhalb des Modellierungsprozesses zu verorten. Sequentiell gedacht, geht hier jemand einen Schritt in die Welt der Punkte, geht dann wieder raus und dann „vielleicht" wieder rein. Damit liegt eine Wiederholung der bereits gezeigten Struktur der statischen „Verwerfung" von Realem und Modell vor. Auf eine Zueinanderführung bzw. Vermittlung zwischen Realem und Modell wird auch hier verzichtet.
Das Problem kann noch präziser gefaßt werden: Es besteht das Problem, zwischen dem Realen und dem Modell zu „vermitteln" – diesen Begriff benutzt die PISA-Gruppe (siehe Abschnitt 5.1., Stichwort Modellbildung). Diese Vermittlung soll durch die Teilprozesse mathematisieren → verarbeiten → interpretieren → validieren innerhalb des Modellierungsvorganges stattfinden. Man möchte sich bei Aufgaben nicht darauf beschränken, entweder das Reale oder das Modell vorzugeben: Würde man nur das Reale vorgeben, so würde die Aufgabe zumindest als Testaufgabe zu schwer bzw. zu aufwendig. Würde man nur das Modell vorgeben, so wäre keine Modellierungsanforderung gegeben. Man muß also bereits in der Aufgabenstellung zwischen dem Realen und dem Modell „vermitteln". Das kann beispielsweise geschehen, indem man Teile einer Mathematisierung angibt oder Verarbeitungs- oder Interpretationshilfen einbaut. Dabei muß einerseits sowohl das Reale als auch das Mathematische (bzw. das Modell) in seiner Autonomie und Authentizität respektiert und aufgenommen werden. Andererseits soll das Reale in das Modell fließen und umgekehrt soll das Modell das Reale widerspiegeln und die mit Hilfe des Modells gewonnenen Erkenntnisse sollen auf das Reale projizierbar sein.
Das in dieser Aufgabe auftretende Phänomen ist nun nicht einfach eine Nicht-Vermittlung. Es ist der gescheiterte Versuch einer Vermittlung. Das Scheitern besteht in einem unsystematischen und unvermittelten Ineinanderschieben der beiden Elemente, die vermittelt werden sollen - also in einer Verwerfung. Ursache ist die Nichtrespektierung und Nichtaufnahme der Autonomie und Authentizität sowohl des Realen als auch des Mathematischen.

Jede Kante der Pyramide in der Skizze misst 12 m.

Hier wiederholt sich die bereits rekonstruierte Struktur der für die Lösung überflüssigen bzw. redundanten Informationen, die das intuitive Vorgehen stärkt.

Manifest liegt eine Modellierungshilfe vor, wie die PISA-Gruppe sie wegen der sonst zu hohen Aufgabenschwierigkeit begründet hat. Man mag die Frage stellen, ob diese Information nicht an irgendeiner Stelle gegeben werden muß: Im Sinne der Mitteilung, dass die Seitenlänge 12 m ist, ist die Information redundant, weil die Seitenlänge in der Zeichnung verzeichnet ist. Durch die Mitteilung wird also ein weiterer Informationskanal eröffnet und damit die Bearbeitungswahrscheinlichkeit eventuell vergrößert. Gleichzeitig wird nochmals die Ernsthaftigkeit des Realproblems zerstört, denn eine Schülerin, die ein Dach skizziert, hätte keinen Grund, diesen Satz dazuzuschreiben.

Im Sinne einer Modellierung ist die Information nicht nötig, weil ein Modellierungsschritt in der Annahme der Regelmäßigkeit des Daches bestehen würde: Drei Seitenlängen sind in der Zeichnung bereits angegeben, und es liegt nicht nahe anzunehmen, dass die anderen Pyramidenkanten andere Längen haben. Wenn man etwas anderes annimmt, dann erhält man ein anderes Ergebnis. Das wäre in einer Klassenarbeit auch kein Problem, denn auch mit der eher abseitigen Modellierung eines schiefen Daches kann man arbeiten. Im vorliegenden Aufgabentext ist dieser Weg aber durch das Quadratische des Bodens verstellt.
An dieser Stelle des Textes geht es nur noch um eine Pyramide. Der Realschnörkel ist jetzt endgültig abgelegt, wir scheinen endgültig in der Mathematik angekommen zu sein. Der Text kann sich aber auch hier nicht vollständig aus dem Empirischen in die Mathematik verabschieden, denn die Kante ist nicht 12 m lang, sie *misst*.

1. Berechne den Flächeninhalt des Dachbodens ABCD.
Der Flächeninhalt des Dachbodens ABCD = ______ m^2.

Hier wiederholen sich die Struktur der Verwerfung von Modell und Realität und des Nichternstnehmens des Problems. Der Dachboden ist nur noch Attrappe. Die Verwerfung wird nochmals zugespitzt: Jetzt ist der Dachboden nicht mehr nur „*in der Skizze ABCD*“, jetzt *ist* der Dachboden selbst *ABCD*. Um Modellierung geht es aber wiederum nicht, es geht ausschließlich um Berechnung. **In der Aufgabenstellung steht also eine komplex verworfene Einkleidung gegen eine triviale Rechenoperation:** Zu *berechnen* ist lediglich zwölf mal zwölf.
Die weitere Reduzierung des inhaltlichen Anspruchs durch Vorgabe der Einheit stellt ebenfalls nur eine Wiederholung von bereits Vorgefundenem dar.
Die unzulässige Vermengung eines deutschen Satzes mit mathematischen Symbolen reproduziert die bereits herausgearbeitete Struktur der Zerstörung des Fachsprachlichen.

2. Berechne die Länge von $\overline{EF}$*, einer der waagerechten Kanten des Quaders.*
Die Länge von $\overline{EF}$ *=* ______ *m.*

Hier wird nicht einmal mehr der Anschein aufrechterhalten, es ginge um etwas Reales. Auf bizarre Weise wiederholt sich die auch hier überflüssige Bezugnahme auf den Quader. Zur Wiederholung der Verwerfungsstruktur kommt auch hier die Reduzierung des Aufgabenanspruchs auf das Berechnen: An *Rechnung* ist nur die Halbierung von Zwölf zu leisten. Hier werden gar nicht Lösungswege geschwächt – denn es gibt keinen Lösungsweg, der sich im Halbieren von Zwölf erschöpfen würde. Hier werden die wesentlichen Teile des Lösungsprozesses unterlaufen. Die Formulierung *Bestimme* ist eine einfache Alternative, deren Generierung offenbar habituell nicht möglich war.

Ich beende die Interpretation an dieser Stelle, weil sie empirisch bereits gesättigt ist. Offensichtlich ist es unfruchtbar, anhand der Aufgabenstellungen einen weiteren Falsifikationsversuch für die bisher herausgearbeiteten Strukturen vorzunehmen.

Fazit

Diese Aufgabe wird von der PISA-Gruppe als typisches Beispiel für den *Realistic Mathematics Education*-Ansatz bezeichnet:

> „Charakteristisch ist vor allem wieder, dass eine außermathematische Situation (Foto) sogleich durch eine schematische Zeichnung ergänzt wird, sodass beides, außer- und innermathematische Zusammenhänge, gleichzeitig angezeigt sind.
> Das Item „Bauernhöfe 1“ fragt nach dem Flächeninhalt des Dachbodens. Dies erfordert eine rechnerische Modellierung, die auf einen einfachen Standardalgorithmus (Berechnung des Flächeninhalts eines Quadrats) führt. Mit diesen Anforderungen liegt die Aufgabe noch auf Kompetenzstufe II. „Bauernhöfe 2“ ist ebenfalls eine rechnerische Modellierungsaufgabe. Die Berechnung der Länge des Balkens EF erfordert jedoch, zusätzliche schulische Kenntnisse heranzuziehen. Solches Wissen kann - dahingestellt, ob es bewusst angewandt wird oder nicht und ob es vom Bearbeiter so benannt wird oder nicht - Kenntnis der Strahlensätze oder Kenntnisse über die Mittellinie eines Dreiecks bedeuten. Allerdings ist auch denkbar, dass man die korrekte Antwort (6 m) intuitiv abschätzt.“ (PISA 2000, S.151)

Bereits die Generierung von Lösungswegen hat gezeigt, dass die Aufgabe mehr bzw. anderes mißt als die PISA-Gruppe hier darstellt. Sie mißt zunächst auch andere Kompetenzen. Zusätzlich mißt die Aufgabe in vielerlei Weise Testfähigkeiten wie Resistenz gegen Realitätsbehauptungen, gegen An-der-Nase-herumgeführt-werden (bzw. Verhöhnung) und gegen verbale Umstellungen des Problems, Irritationsresistenz, Durchhaltefähigkeit.

> „Ziel des PISA-Tests ist es ... zu prüfen, ob Schülerinnen und Schüler grundlegende mathematische Konzepte so verstanden haben, dass sie mit diesen Werkzeugen Problemsituationen aus unterschiedlichen Kontexten behandeln können.“ (ebd., S.143)

In der Aufgabe Bauernhöfe 1 ist der Flächeninhalt eines Quadrats in einem Kontext zu bestimmen. Von einem „grundlegenden mathematischen Konzept“ kann man hier sicher nur schwerlich sprechen. Für die Aufgabe Bauernhöfe 2 bleibt unklar, *welches* „grundlegende mathematische Konzept“ der Schüler hier „verstanden“ haben soll. Die Mittellinie des Dreiecks? Intuition als grundlegendes mathematisches Konzept? Messen? Strahlensatz? Auch hier würde man nicht von „grundlegenden mathematischen Konzepten“ sprechen, die der Test erfassen möchte. Der Text dementiert die Orientierung an Konzepten zusätzlich, indem er auf die Anforderung des *Berechnens* fokussiert, wobei lediglich zwölf mal zwölf und zwölf durch zwei zu rechnen ist.

1. Die reale Situation wird nicht ernst genommen. Sie hat keine Bedeutung für das mathematische Problem.

Dies zeigt sich erstmals in der Überschrift „Bauernhöfe“. Es geht in keiner Weise um Bauernhöfe und die Überschrift verspricht eine Reichhaltigkeit, die die Aufgabe nicht liefert.

Die Aufmerksamkeit des Lesers wird auf irgendein Foto irgendeines Bauernhauses gelenkt, welches dann aber auch nicht weiter interessiert. Zusätzlich wird die manifeste Bedeutung des Hauses sofort durch die Fokussierung auf die Pyramidenförmigkeit des Daches zerstört.

Die Struktur wiederholt sich in der Zerstörung der Bedeutung des Realen durch die Konstruktion mit *Der Dachboden ... ist ein Quadrat.* Mit *Jede Kante der Pyramide ...* ist das Reale endgültig versunken und muß in den Aufgabenstellungen nur noch als Dekoration herhalten. Ein Rückbezug auf das Dach findet nicht statt, die Aufgabenstellungen haben mit dem Dach auch gar nichts zu tun.
Diese Struktur führt direkt zu einer weiteren Aussage:

2. Der Schüler wird in seiner Rolle als Problemlöser nicht ernst genommen.

Der Schüler wird einerseits an der Nase herumgeführt, wenn er sich auf die Realsituation einläßt. Er wird dann vergeblich nach in der Überschrift versprochenen Inhalten suchen. Hier liegt auch ein gewisser Zynismus gegenüber jenem Schüler, welcher sich ernsthaft interessanten Themenstellungen zuwendet. Er soll sich ein angebliches Bauernhaus ansehen, ohne dass sich ein Grund dafür zeigen wird. Ihm wird erzählt, eine Schülerin - also quasi eine Kollegin - habe eine Skizze angefertigt. Es gibt aber keine Skizze, sondern eine Zeichnung, und die wurde nicht von einer Schülerin angefertigt. Der Schüler wird sprachlich auf ein Niveau zurückverwiesen, auf dem er kein mathematisches Konzept eines Quadrats hat und auf dem er nicht einmal unterscheiden könnte, ob eine Gehwegplatte ein Quadrat oder quadratisch ist. Zu guter Letzt werden ihm zwei in dieser Konstellation triviale mathematische Probleme in einer prätentiösen und sprachlich verschwierigenden Verpackung präsentiert.

3. Die Aufgabe ist sprachlich unnötig verschwierigt. Dies ist nicht der realen Situation oder dem mathematischen Inhalt geschuldet.
Eine Anforderung an den Schüler besteht damit darin, diese Verschwierigung zu überwinden, um zur - relativ dazu - einfachen mathematischen Aufgabe vorzudringen. Eine zweite Anforderung besteht darin, Irritationen durch sprachliche Regelverletzungen zu überwinden.
In jeder Kommunikationspraxis erfolgt eine gedankliche Heilung von Regelverletzungen („Das hat sie wohl **so** gemeint“). Sonst wäre Kommunikation gar nicht möglich.[62] Dieser gedankliche Heilungsaufwand ist bei schriftlichen Texten besonders hoch, weil sie in besonderer Weise dem Anspruch an Wohlgeformtheit unterliegen.[63] Bei schriftlichen Prüfungen ist dieser Anspruch nochmals verschärft: Je wertvoller die durch die Prüfung zu vergebende Zukunftschance, desto bedeutender wird die sprachliche Wohlgeformtheit der Aufgabenstellung (und der Lösung) - desto stärker aber auch der

[62] Dies merkt man besonders klar bei Transkriptionen von Unterrichtssituationen: Würden im Unterricht nicht ständig gedankliche Heilungen und Interpretationen passieren, würde niemand etwas verstehen. Beim ersten Lesen eines Unterrichtstranskripts findet man oft nicht heraus, was der Inhalt des Textes ist. Man muß sich dazu erst in die mündliche, immerfort Regelverletzungen heilende und halbe Sätze gedanklich vervollständigende mündliche Situation versetzen.
Trotz der Normalität des Heilens von Regelverletzungen erfordert die Heilung intellektuellen Aufwand. (Extrembeispiel: Nur Eltern verstehen anfangs ihre Kinder. Je ferner von Sprachregeln die Kinder sprechen, desto weniger Personen sind bereit bzw. in der Lage, die „Heilungsarbeit“ auf sich zu nehmen.)
[63] Das wird durch das Aufkommen privater E-Mails besonders gut illustriert: Je vertrauter wir mit dieser schriftlichen Form sind, umso leichter fällt es uns, ihre mangelnde Wohlgeformtheit gedanklich zu heilen und umso größere Nichtwohlgeformtheit wagen wir. Orthographische, grammatikalische oder Ausdrucksfehler werden quasi überlesen. Diese gegenseitige vergrößerte Heilungsbereitschaft begründet den Charakter dieser Form als zwischen mündlichen und schriftlichen Texten stehend.

intellektuelle Anspruch, um mangelnde Wohlgeformtheit zu heilen. Der Prüfling kann einerseits nicht einfach sagen: Ach, „die“ werden das schon so und so gemeint haben. Er muß andererseits immer in Erwägung ziehen, dass der Umgang mit der Abweichung von sprachlicher Wohlgeformtheit **zur Prüfungsanforderung gehört**.

Im Text liegen mehrere sprachliche Verschwierigungen vor:
Der Dachboden, in der Skizze ABCD, ist ein Quadrat. Eine Verschwierigung liegt in der Überflüssigkeit der Quadratinformation, eine zweite in der Satzkonstruktion.
Die Balken, die das Dach stützen, sind die Kanten eines Quaders (rechtwinkliges Prisma) EFGHKLMN. Diese Aussage ist für die Problemlösung überflüssig. Die Klammer erzeugt nicht nur Verschwierigung, sondern birgt großes Irritationspotential. Der Satz ist im Vergleich zu denkbaren Alternativen schwierig konstruiert.
E ist die Mitte von $\overline{AT}$*, F ist die Mitte von* $\overline{BT}$*, G ist die Mitte von* $\overline{CT}$ *und H ist die Mitte von* $\overline{DT}$.
Die für die Aufgabenlösung überflüssige Anhäufung der quasi immer gleichen Information stellt eine Verschwierigung dar. Hinzu kommt die Verschwierigung durch die Anforderung, viele überflüssige Objekte und ihren Zusammenhang zu erfassen.
Der gesamte erste Teil der Aufgabe, in dem der Schüler mehrfach an der Nase herumgeführt wird, ist überflüssig und stellt insofern eine Verschwierigung dar.
Alle aufgefundenen Verschwierigungen sind unnötig. Alternativen sind leicht zu finden. **Alle Verschwierigungen hängen aber mit der Konstruktion einer Modellierungsillusion zusammen.**

Sprachliche Regelverletzungen finden sich bei *ein Foto eines*, im Quadratbegriff von *Der Dachboden ... ist ein Quadrat*., in der Verwendung des Wortes *Dachboden*, in der Falschaussage, dass die Balken *die* Kanten des Quaders sind, in der Verwendung von *Mitte*, im Einbau des Gleichheitszeichens in die Antwortsätze.

4. Die Interpretation hat außerhalb der Fragestellung die Erkenntnis produziert, dass „mathematische Ausbildung“ bezogen auf die Begriffe Quadrat und Quader eher „Weiter-Bildung“ ist. Diese Begriffe sind aus dem Mathematischen in die Umgangssprache geflossen, sind dort aber weniger präzise und klar begrenzt gefaßt, als sie es mittlerweile in der Mathematik sind. Dabei bestehen Unterschiede: Es gibt ein umgangssprachliches, außermathematisches Konzept des Quaders, wohingegen das umgangssprachliche Konzept des Quadrats viel näher am mathematischen Konzept liegt. Dies korrespondiert damit, dass das Wort „quadratisch“ sprachlich und inhaltlich näher am mathematischen Quadrat liegt als das Wort quaderförmig am mathematischen Quader. Man kann sich das veranschaulichen:

mathematischer Begriff	Alltagsbegriff nah am mathematischen Begriff	Alltagsbegriff entfernter vom mathematischen Begriff
Quadrat	quadratisch	quadratförmig
Quader	quaderig	quaderförmig

Die Wörter „quadratförmig“ und „quaderig“ existieren nicht, hier fehlt also ein adäquates sprachliches Ausdrucksmittel. Damit wird aber deutlich, warum es so schwer war herauszufinden, auf welcher Modellierungsebene sich der Begriff „Quader“ hier bewegt: Der Begriff „quaderig“ hätte klar auf ein Realmodell verwiesen, kann bzw. muß aber sprachlich durch „Quader“ ersetzt werden. Damit ist die Grenze zum mathematischen Modell nicht mehr sichtbar, auf das der Begriff „Quader“ ja auch bereits verweist.

Wenn man nur die Aufgabe untersuchen will, ist dieser Interpretationsteil forschungsökonomisch unerfreulich, denn er bestätigt mit riesigem Aufwand nur eine Struktur, die sich ohnehin mehrfach zeigt:

5. Statt den Schüler im Übergang von der Realität zum mathematischen Modell zu begleiten, werden das Reale und das Mathematische in ihrer Autonomie und Authentizität nicht ernst genommen, sondern verworfen.

Das deutet sich in der Struktur des Nichternstnehmens und an-der-Nase-Herumführens im Vortext an, zeigt sich dann erstmals deutlich in der verworfenen Aussage *Der Dachboden ... ist ein Quadrat.* Das Reale wird mit dem Modell gleichgesetzt, ebenso bei den Balken, die Kanten eines Quaders *sind*. Mit der überflüssigen Quadrataussage wird eine Frage beantwortet, die weder in der Realität noch im Modell steht. Gleichzeitig wird dadurch manifest ein Modellierungsgedanke konstruiert, der sich aber als nicht existent erweist. Auch eine Modellierungsanforderung besteht gar nicht mehr, nur noch die Aufgabe, den Flächeninhalt eines Quadrats zu berechnen.

Manifest wird im Aufgabentext eine Verbindung zwischen dem Dach bzw. seinen Stützbalken und dem Modell gezogen. Die dabei vermittelte Information läuft aber ins Leere, weil sie für die Problemlösung überflüssig ist. Es liegt also keine Verbindung zwischen der Realität und dem Modell als Modell für eine Problemlösung vor. Es findet kein Modellierungsvorgang statt, er wird nur behauptet.

Der durch die Klammer implizierte Satz *Die Balken sind Kanten des rechtwinkligen Prismas EFGHKLMN.* dementiert (wenig wohlgeformt) eine Modellierungsanforderung, verbunden mit dem äußeren Anschein einer fachlichen Präzisierung, die keine Präzisierung ist.

Im nächsten Satz und in den Aufgabenstellungen wiederholt sich die Verwerfung von Realem und Modell. Sie findet sich also in jedem Satz des mathematischen Haupttextes.

6. Das Mathematische wird in dieser Aufgabe mißachtet.

Im vorigen Punkt wurde bereits gezeigt, dass das Mathematische in seiner Autonomie gegenüber dem Realen beschädigt wird. Es wurde auch bereits darauf hingewiesen, dass hier inhaltlich einfache Aufgaben stehen: Man muß den Flächeninhalt eines Quadrats bestimmen und man muß intuitiv die Länge der Mittellinie eines gleichseitigen Dreiecks bestimmen, kann hier aber auch andere, anspruchsvollere Lösungswege beschreiten. Die Einfachheit wird durch sprachliche Verschwierigungen, sprachliche Irritationen bzw. Verwerfungen und pseudomotivationale Aufblähungen zugedeckt. Durch die Zerstörung der Modellierungsanforderung wird auch mathematischer Anspruch zerstört. Ein Dachboden wird als Quadrat eingeführt, damit wird der mathematische Begriff des Quadrats unterminiert, der Begriff des Mittelpunkts hingegen wird verweigert. Die Struktur zieht sich bis in den letzten Satz durch: Nicht einmal im dort sprachlich konstruierten mathematischen Raum darf die Länge mathematisch sein, sie bleibt empirisch.

7. Redundante Informationen erwachsen nicht aus der Problemstellung, sondern werden künstlich erzeugt.

Dies geschieht in den Informationen zum Quadrat und zum Quader sowie im Satz über die Mitten. Bei der Quader-Information nähert sich die redundante Information dem Charakter einer potentiellen Falle, ansonsten ist mit ihnen Irritationspotential verbunden.

8. Die Problemstellung und der Aufgabentext fördern für Aufgabe 2 den Lösungsweg über Intuition.

In dieser Interpretation konnten nur wenige textliche Beeinflussungsfaktoren *für bestimmte Lösungswege* eruiert werden. Im Text findet sich drei Mal eine Förderung des intuitiven Lösens durch Symmetriekonstruktionen (quadratischer Charakter von ABCD, Rechtwinkligkeit von EFGH, Gleichseitigkeit der Pyramide). Alle anderen Hinweise finden sich im Problem selbst und in der Zeichnung.

Die räumliche Problemkonstellation schwächt alle Lösungswege. Andererseits haben wir es mit gleichseitigen Dreiecken zu tun. Die Problemstellung entfaltet bei gleichseitigen Dreiecken eine ausgesprochen starke suggestive Kraft, die intuitives Vorgehen stärkt.

9. Die Einstufungen der PISA-Gruppe bezüglich der „Typen mathematischen Arbeitens“ und bezüglich der inhaltlichen Füllung der „Kompetenzstufen“ werden nicht bestätigt.

Beide Aufgaben gehören zur Leitideen „Raum und Form“ und werden dem Typ 1B „rechnerische Modellierung“ zugerechnet: „Zur Lösung der Aufgabe ist eine Modellierung erforderlich. Diese ist jedoch unter Rückgriff auf einen einzigen Algorithmus, eine einzige Formel möglich. Es ist also die passende Formel, das passende Verfahren, die passende Prozedur aus dem vorhandenen Wissen auszuwählen und dann anzuwenden. Das zur Modellierung erforderliche Wissen stammt aus einem einzigen mathematischen Gebiet.“ (Neubrand u.a. 2001, S.52)

Die Interpretation hat gezeigt, dass hier keine Modellierungen vorzunehmen sind, weil jegliche Modellierungsanforderung bereits zerstört ist. Mathematisch besteht die Aufgabe darin, ebene Probleme aus einem zunächst räumlichen Problem herauszuschälen und diese ebenen Probleme dann zu lösen (Da „der Dachboden ein Quadrat ist“, ist diese Anforderung wiederum gebrochen). Die PISA-Gruppe hat allerdings ihren Modellierungsbegriff so weit gefaßt, dass sie selbst das Herausschälen eines ebenen Aspekts aus einem räumlichen Problem noch als Modellierung bezeichnen wird. Das zeigt aber nicht, dass hier modelliert werden muß, sondern dass der PISA-Modellierungsbegriff keine Trennschärfe besitzt.

Eine Schwierigkeit der Aufgabe besteht darin, sich durch viel redundante bzw. überflüssige Information und durch eine erstaunliche Anhäufung von Irritationen durchzuarbeiten. Dieser Aspekt spielt in den veröffentlichten Kategorieeinordnungen keine Rolle.

Wegen der zerstörten Modellierungsanforderung gehört die Aufgabe „Bauernhöfe 1“ zum „Typ mathematischen Arbeitens“ 1A „technische Aufgaben“: „Die Aufgabe erfordert nur technische Fertigkeiten und/oder den Abruf von Faktenwissen.“ (Neubrand u.a. 2001, S. 52) Die Einordnung in die internationale Competency Class „reproduction, definitions and computations“ (siehe OECD 2000, S.47) kann bestätigt werden.

Für die Aufgabe „Bauernhöfe 2“ kann man eine Einordnung nur schwerlich vornehmen. Der intuitive Weg findet keinen Platz im Kategoriensystem der PISA-Gruppe. Der Weg des Messens und der Weg über die Mittellinie des Dreiecks kann unter 1A oder 1B eingeordnet werden, je nachdem wie weit man dem PISA-Modellierungsbegriff folgt. Der Weg über den Strahlensatz könnte dementsprechend unter 1A eingeordnet werden, wenn man dem PISA-Modellierungsbegriff folgt unter 2B. Die Einordnung in die internationale Kategorie „connections and integration for problem solving“ kann nur für den Weg über den Strahlensatz bestätigt werden, diese Einordnung widerspricht er aber der deutschen Einordnung unter 1B, hier deuten die nationale und die internationale PISA-Gruppe die Kategorien bzw. die Aufgabe wohl verschieden.

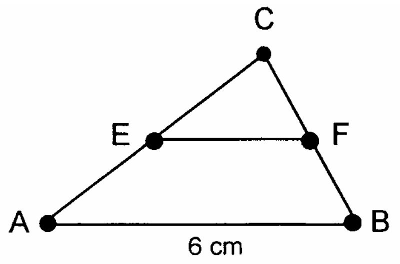
Die Aufgabe „Bauernhöfe 1“ wird bei einer internationalen Lösungshäufigkeit von 61 % (Deutschland 51 %) bei 492 Punkten und damit auf Kompetenzstufe II eingeordnet. Wenn man sich auf die Kompetenzstufenbeschreibung der PISA-Gruppe einläßt (zu Schwierigkeiten mit dieser Beschreibung vergleiche Meyerhöfer 2004), dann kann man dieser Einordnung nach der Interpretation folgen, und zwar unabhängig von der Zuordnung zum Typ mathematischen Arbeitens. Bei „Schwierigkeits“vergleichen mit nationalen Aufgaben ist allerdings zu beachten, dass die nationalen Aufgaben nur aufgrund der nationalen Lösungshäufigkeiten in die Skala eingeordnet werden, die internationalen Aufgaben hingegen aufgrund der internationalen Werte. Diese Aufgabe würde mit ihrem nationalen Lösungswert von 51% auf Stufe III eingeordnet werden, also plötzlich „schwerer“ sein, eine weitere Illustration der problematischen Konstruktion des Kompetenzstufenkonstrukts.
Die Aufgabe „Bauernhöfe 2“ wird bei einer internationalen Lösungshäufigkeit von 55 % (Deutschland 41 %) bei 524 Punkten und damit auf Kompetenzstufe III eingeordnet. Bereits bei Betrachtung der verschiedenen Lösungswege zeigt sich, dass diese Einordnung inhaltlich nicht gerechtfertigt werden kann, da die verschiedenen Lösungswege bereits auf verschiedene Kompetenzstufen verweisen. Die nochmalige Brechung der Kompetenzstufenzuordnung durch die sprachlichen und Verwerfungen kommt hinzu.

5.6. Die (nationale) PISA-Aufgabe „Dreieck“

Die PISA-Gruppe vergleicht die Aufgabe Bauernhöfe 2 mit der vom mathematischen Problem her ähnlichen Aufgabe Dreieck. Da dieser Vergleich die Schwäche des theoretischen Konzepts von PISA eindrucksvoll illustriert, soll die Aufgabe „Dreieck“ im weiteren objektiv-hermeneutisch interpretiert und mit der Aufgabe „Bauernhöfe 2“ verglichen werden. Innerhalb der kontrastiven Auswahl von

Aufgaben bietet sich an, diese Aufgabe zu interpretieren, weil sie - obgleich vom mathematischen Inhalt ähnlich - völlig anderen Charakters zu sein scheint: Es existiert keine Realitätsanbindung, die in der Aufgabe „Bauernhöfe“ so viele Verwerfungen produziert hatte. Hier scheint uns die pure Mathematik entgegenzutreten.

Die Seite $\overline{\mathrm{AB}}$ *des Dreiecks ABC ist 6 cm lang. Es werden die Mittelpunkte E und F der Seiten* $\overline{\mathrm{AC}}$ *und* $\overline{\mathrm{BC}}$ *eingezeichnet.*
Wie lang ist $\overline{\mathrm{EF}}$ *?*



(Bild nicht maßgenau)

Anmerkung: Es sind weder Multiple-Choice-Angebote noch Zeilen zur Beantwortung vorgegeben, aber es ist mehr als eine viertel DIN-A4-Seite Platz frei.

Lösungswege

- Messen: Dieser Weg ist mit einer hohen Zeitprämie versehen, führt also schnell zum Ziel. Er stellt aber keine mathematische Operation dar. Das Messenkönnen ist eine wichtige praktische Fähigkeit, bewegt sich aber doch eher außerhalb dessen, was man mit „mathematisch“ bezeichnet. Das Messen ist hier keine Annäherung an die geometrische Operation, sondern eine Abkehr von ihr. In jedem Fall ist das Ausmessen einer Strecke eine Fähigkeit, die man innerhalb des Kompetenzstufenmodells klar auf Stufe I („Rechnen auf Grundschulniveau“, siehe PISA 2000, S.160) verweisen würde.
- Intuition: Außer der Länge 6 cm ist kein weiteres Maß gegeben, trotzdem ist die Bestimmung eines weiteren Maßes verlangt. Es liegt - da die Mittelpunkte ja auf die Seitenhälften verweisen - nahe anzunehmen, dass EF halb so lang wie die gegebenen 6 cm ist. Man kann das ausgehend von der Intuition auch nachmessen, aber das intuitive Lösen stellt einen eigenen Weg der Lösungsgenerierung dar. Auch dieser Weg ist mit einer hohen Zeitprämie versehen, es läßt sich aber nicht abschätzen, um wieviel geringer diese Zeitprämie ausfällt als beim Messen. In das Kompetenzstufenmodell läßt sich dieser Weg überhaupt nicht einordnen.
- Lösen über den Satz von der Mittellinie des Dreiecks: Dazu muß man EF als Mittellinie des Dreiecks ABC erkennen und wissen, dass die Mittellinie eines Dreiecks halb so lang wie seine Basis ist. (Es ist ersichtlich, dass dieser Satz mit der eben beschriebenen Intuition zusammenhängt.) Dieser - geometrisch strenge - Weg ist mit einer höheren Zeitprämie als der intuitive Weg verbunden, weil die Irritation durch die mit Intuition verbundene Unsicherheit wegfällt: Wer den Satz aufrufen kann, löst die Aufgabe sehr schnell. Dieser Weg verlangt lokales mathematisches Formel- bzw. Satzwissen. Man würde ihn wohl auf Stufe II oder III einordnen. („Elementare Modellierungen“ oder „Modellieren und begriffliches Verknüpfen auf dem Niveau der Sekundarstufe I“)

- Lösen über den 2.Strahlensatz: Dieser Weg setzt das Wissen voraus, dass die Strecken CA und CB durch ihre Mittelpunkte halbiert sind. Außerdem muß erkannt werden, dass eine Konstruktion vorliegt, die mit Hilfe des Strahlensatzes bearbeitet werden kann. Das kann über Intuition oder über die Umkehrung des ersten Strahlensatzes erfolgen. Über eine allgemeine Betrachtung oder über das Einsetzen von Zahlen bzw. Variablen gelangt man mit Hilfe des zweiten Strahlensatzes zur Lösung. Dieser Weg ist der einzige Weg, der genuin mathematisches Denken verlangt. Er benötigt mit Abstand die meiste Zeit. Diesen Weg würde man wohl auf Stufe IV oder V ansiedeln. („Umfangreiche Modellierungen auf der Basis anspruchsvoller Begriffe“ oder „Komplexe Modellierungen und innermathematisches Argumentieren“)

Die Betrachtung der möglichen Lösungswege zeigt, wofür hier Zeitprämien vergeben werden: Die Wege des außerhalb mathematischen Denkens liegende Messens und des lokalen Formel- bzw. Satzwissens erhalten die höchste Zeitprämie. Der genuin mathematische Weg wird - von der Zeitprämie her gesehen - bestraft. Dies gibt einen Hinweis auf die hier vermittelte mathematische Kultur, zeigt aber auch, dass der hier zum Tragen kommende Aspekt von Testfähigkeit bedeutet, **nicht** mathematisch zu denken. Die Behauptung, man könne dieser Aufgabe einen inhaltlich interpretierbaren Schwierigkeitswert zuordnen, ist bereits hier nicht aufrecht zu erhalten.

Zeichnung

Die Zeichnung zeigt eine aus mehreren Strecken zusammengesetzte Figur. Auf dieser Figur sind fünf relativ große schwarze Kreise eingezeichnet, in deren Nähe jeweils einer der Großbuchstaben A, B, E, F und C steht. Dies kennzeichnet in einem geometrischen Kontext, dass es sich um Punkte mit den zugehörigen Bezeichnungen handelt. (Allerdings ist die Darstellung von Punkten durch große Kreise im Sinne der Geometrie befremdlich.) Die Figur ist lesbar

- als Strahlensatzfigur. Diese Lesart wird dadurch gestärkt, dass die Bezeichnungen AB und EF einen Fokus auf diese beiden Strecken stiften. Dies geschieht durch die Zusammengehörigkeit der Buchstaben im Alphabet und innerhalb geometrischer Bezeichnungskonventionen. Diese Lesart wird gleichzeitig durch die Zusammengehörigkeit von A, B und C geschwächt: In einer Strahlensatzfigur würde entweder von der Spitze ausgehend bezeichnet oder die Spitze mit einem „außenstehenden“ Buchstaben bezeichnet.
- als zwei ineinander gelegte Dreiecke - dafür spricht die Punktbezeichnung, denn die übliche Vergebung der Punktnamen deutet darauf hin, dass zunächst das Dreieck ABC gezeichnet wurde. Mit der Weiterführung mit E und F wird diese Konvention aber gebrochen, nach ihr würde mit D weitergearbeitet werden.
- als ein Dreieck ABC, in das eine Strecke EF einbeschrieben wurde.
- als Trapez, auf dessen obere Seite ein Dreieck aufgesetzt wurde bzw. umgekehrt. Auch diese Lesart würde eine Erklärung der abweichenden Bezeichnungen verlangen.

Unter der Strecke AB steht *6 cm*. Dies weist in der geometrischen Bezeichnungskonvention darauf hin, dass die Länge der Strecke AB 6 cm beträgt. In Klammern ist angegeben: *(Bild nicht maßgenau)*.

Dies ist wirklich der Fall, AB ist etwa 4,4 cm lang, eine Messung wird durch die Dicke der Punkte erschwert. Dies schwächt den Weg des Messens. Allerdings ist die Zeichnung maßstabsgetreu, man kann also auch mit Messen arbeiten.

Das Bild verweist insgesamt auf ein geometrisch konnotiertes Problem. Die Problemstellung erschließt sich nicht aus der Abbildung. Sie läßt mindestes vier geometrische inhaltliche Assoziationen zu, zwischen denen an dieser Stelle keine Entscheidung möglich ist. Der Weg über die Mittellinie des Dreiecks wird durch das Bild am meisten gestärkt.

Textanalyse

Die Seite $\overline{\mathrm{AB}}$ *des Dreiecks ABC ...*

Die Seite verweist auf eine fest bestimmte Seite (vergleiche mit *eine*). *Seite* stellt den Bezug zu einer Figur (*Kante* würde z.B. auf einen Körper verweisen) und einen Teil derselben, der fest begrenzt ist, her. $\overline{\mathrm{AB}}$ benennt die Begrenzung durch die Punkte A und B, führt Fachsprachlichkeit ein und expliziert durch die Überstreichung den Charakter der Seite als Strecke.
... des Dreiecks ABC spezifiziert die Figur, über die gesprochen wird. Einerseits geht es um ein *Dreieck*, andererseits um ein bestimmtes. Es wird nicht weiter charakterisiert, z.B. durch *gleichseitig*, *spitzwinklig* o.ä. Die Buchstabenfolge ABC bewegt sich innerhalb von geometrischer Bezeichnungskonventionalität. Hier wird der Fokus der Wahrnehmung auf dieses spezielle Dreieck gelenkt, aber der Text wird nicht mit der Zeichnung verzahnt. Zum Vergleich: In der für TIMSS herausgearbeiteten Logik würde diese Verzahnung als Beschleuniger der Texterfassung eingesetzt werden, dort könnte z.B. stehen: *In der Abbildung ist die Strecke* $\overline{\mathrm{AB}}$ *6 cm lang*. Diese Logik der Schnelligkeit wird hier nicht bedient.
Welche textliche Implikation hat dann aber der Einschub ... *des Dreiecks ABC* ...? Inhaltlich lenkt er weg vom Trapez ABFE hin auf das durch die Bezeichnungen fokussierte Dreieck ABC. Einerseits liegt hier eine Stärkung des Lösungswegs über die Mittellinie des Dreiecks vor. Andererseits wird der Lösungsweg über den Strahlensatz deutlich geschwächt, denn auch die Interpretation der Figur als Strahlensatzfigur wird durch das Fokussieren auf das Dreieck gestört.

... ist 6 cm lang.

Hier wiederholt sich mit *6 cm* die Struktur der Verwendung von Bezeichnungskonventionen. Ansonsten liegt hier zunächst eine falsche Aussage vor: $\overline{\mathrm{AB}}$ ist nicht 6 cm lang. $\overline{\mathrm{AB}}$ ist im Testheft nur etwa 4,4 cm lang, so dass man von einer Irritation zwischen manifestem Text und Bildwirklichkeit sprechen kann. Das hängt mit der seltsamen Verwerfung von mathematischer und empirischer Welt in diesem Satz zusammen: Der Text hat nicht in eine Gedankenwelt, sondern in die empirische Welt der Zeichnung eingeführt - sonst hätte im ersten Satz stehen müssen *Die Seite* $\overline{\mathrm{AB}}$ ***eines*** *Dreiecks ABC ...*. Gleichzeitig wird aber mit der äußerlichen Verwendung der geometrischen Bezeichnungskonventionen in *Die Seite* $\overline{\mathrm{AB}}$... in eine mathematische und damit nichtempirische Welt eingeführt. In der nichtempirischen Welt wäre es völlig egal, wie lang die skizzierte Strecke ist.

Um die Bedeutung des *ist ... lang* besser zu verstehen, kann man es mit anderen denkbaren Formulierungen kontrastieren:

- *... sei 6 cm lang*. Hier würde ein genuin mathematischer Gedanke eingebracht, nämlich der Konjunktiv der Annahme, dass AB 6 cm lang ist. Das *sei* würde aber auch elaborierte Fachsprachlichkeit einführen. Es würde klar in eine hypothetische Welt verwiesen.
- Nun kann man einwenden, dass das *ist* eine Art verstecktes *sei* ist, dass es ein konjunktives *ist* ist. Dagegen spricht die wenig elaborierte Fachsprachlichkeit der Kombination mit *lang*. Dies verdeutlicht ein Vergleich mit der Formulierung *... hat eine Länge von 6 cm.* Die *Länge* würde eher darauf verweisen, dass hier vom mathematischen Konzept der Länge einer Strecke die Rede ist. Im Vergleich dazu bewegt *lang* sich mehr auf der Ebene von Alltagssprachlichkeit. Die Kombination *... ist 6 cm lang*. kennzeichnet damit eine klare Entfernung vom Fachdenken hin zu einer empirischen, unscharfen Vorstellung von einer Strecke. Dieser Gedanke wird gestärkt, wenn man die Elaboriertheit noch weiter treibt. Dies geschieht in der Formulierung
- *... habe eine Länge von 6 cm.* Hier ist das Konzept der Länge explizit mit dem Konjunktiv der Länge verbunden.

Die vorliegende Formulierung ist also nicht nur weit von elaborierter Fachsprachlichkeit entfernt. Sie zerstört auch das mathematische Konzept der Strecke zugunsten einer empirischen Vorstellung von einer Strecke als etwas unscharf begrenztem. Gerade auch in Verbindung mit der Wahl der Einheit Zentimeter wird ein realer, handwerklicher Umgang mit der Strecke nahe gelegt. Das Konzept des Messens, aber auch die intuitive Herangehensweise werden hier deutlich gestärkt. Die Zerstörung des Mathematischen und die latente Entfernung von Fachsprachlichkeit steht in auffälligem Widerspruch zur manifesten Betonung von Bezeichnungskonventionalität und Fachsprachlichkeit.

Es werden die Mittelpunkte E und F der Seiten $\overline{\mathrm{AC}}$ *und* $\overline{\mathrm{BC}}$ *eingezeichnet.*

Es werden .. eingezeichnet. Das *werden* schafft hier eine empirische Entstehung der Zeichnung: Man hat förmlich die Abbildung im Telekolleg vor Augen: Bisher wurde das Dreieck ABC eingeblendet und die Seite AB dann rot eingefärbt und mit „6 cm“ beschriftet. Nun werden die Punkte E und F eingezeichnet, ein Bleistift schwebt ins Bild und macht zwei Striche. Mit dem eigentlichen Problem hat diese schaffende Dynamik nichts zu tun, z.B. wäre *E und F sind die Mittelpunkte ...* problemadäquater. Schon *gezeichnet* würde eher auf den geometrischen Gedanken beim Skizzieren verweisen. Durch die vorliegende Dynamik wird der Weg der Mittellinie des Dreiecks am stärksten gestärkt, weil das Einzeichnen der Mittelpunkte die Implikationen des Satzes deutlich hervorhebt. Der Weg des Strahlensatzes wird am meisten geschwächt, weil Strahlensatzfiguren statisch sind. Außerdem werden nicht die Mittelpunkte von Strecken, sondern von Dreiecksseiten eingezeichnet. Die weitere Lesartenbildung ist hier recht schwierig, deshalb schreibe ich die Geschichten auf, die mir zur Lesartenbildung dienen:

1. ein Stadtplan: *Es werden die Orte der Verbrechen eingezeichnet.*
2. Das Bild eines Würfels wurde konstruiert: *Es werden die Diagonalen eingezeichnet.*
3. ein großes Wandbild, die Helfer des Meisters vervollständigen die Vorzeichnung: *Es werden die Muster des Geschirrs eingezeichnet.*

Aus diesen Geschichten extrahiert man eine oder mehrere Lesarten, hier ergibt sich nur eine Lesart: Es wird eine körperliche Aktivität beschrieben, der Ausführende wird nicht in Betrachtung gestellt. Das Objekt ist aber bereits „vorgedacht“, es ist also gedanklich schon vorhanden, es wird nur noch physisch gemacht. Speziell für die Beispiele heißt das: 1. Die Orte der Verbrechen sind bereits erfaßt, 2. die Diagonalen sind in Gedanken bereits vorhanden und ihre Endpunkte sind physisch schon da, 3. die Muster des Geschirrs sind bereits entworfen bzw. vorgedacht.
Die Passivkonstruktion verstärkt die Trennung von Denken und Ausführen, indem es die Wichtigkeit dessen, der ausführt, reduziert. Kontrastierend wäre *Er zeichnete ... ein.* denkbar, diese Formulierung würde die Trennung verwischen.
Hier zeichnet also jemand, der nicht weiter zu beachten ist, die Mittelpunkte der beiden Seiten ein. Es war vorher klar, wo sie hinkommen. Sie *werden* nicht *konstruiert*, was sprachlich denkbar wäre, ein Konstruktionsprozeß findet also nicht statt. Die Mittelpunkte werden auch nicht einfach vorgegeben, z.B. durch *E und F sind/seien die Mittelpunkte der Seiten ...* Sie werden eingezeichnet, und es ist klar, wohin sie gezeichnet werden müssen. Gleichzeitig findet eine merkwürdige Bewegung weg von den Mittelpunkten statt, denn da steht nicht: *Die Mittelpunkte ... werden eingezeichnet.* Dies legte den Fokus auf die Mittelpunkte. Die Formulierung *Es werden ... eingezeichnet.* legt den Fokus auf den Vorgang des Einzeichnens.

Woher ist nun aber klar, wo die Mittelpunkte hinkommen, wohin sie gezeichnet werden müssen? Wenn Konstruktion und Vorgabe wegfallen, dann bleiben nur noch Messen bzw. Schätzen. Latent werden die Mittelpunkte hier also durch Messen bzw. Schätzen gefunden. Damit hat man einerseits diese Vorgehensweisen gestärkt. Damit hat man sich aber andererseits vom mathematischen Gedanken des Mittelpunkts abgewendet, denn der mathematische Mittelpunkt kann nicht durch Schätzen oder Messen gefunden werden. Der so gefundene Mittelpunkt verweist auf eine Kultur, die lediglich aus der mathematischen Kultur abgeleitet ist.[64]

... die Mittelpunkte E und F der Seiten $\overline{AC}$ *und* $\overline{BC}$ ...

Kontrastierende Formulierungen:

- *... der Mittelpunkt E der Seite* $\overline{AC}$ *und der Mittelpunkt F der Seite* $\overline{BC}$ ...
- *... die Mittelpunkte E der Seite* $\overline{AC}$ *und F der Seite* $\overline{BC}$*...*

Diese beiden Formulierungen würden die beiden Mittelpunkte stärker voneinander separieren. Die gewählte Formulierung separiert nicht, sie schafft damit eine Indifferenz zwischen den beiden Mittelpunkten. Diese Indifferenz verweist auf eine Gleichwertigkeit - und das heißt hier: Gleichwertigkeit zur Problemlösung.

Die gewählte Formulierung ist die kürzeste der alternativen Formulierungen. Sie ist gleichzeitig diejenige, die die Zuordnung des Mittelpunkts zur Seite am schwierigsten macht. Das bedient zwar inhaltlich korrekt die Logik der Indifferenz - es ist schließlich für jeden Weg der Problembewältigung egal,

[64] Natürlich kann man sich mit einer auch nur groben Skizze den Satz von der Mittellinie des Dreiecks erarbeiten. Man müßte dann im folgenden nichtempirisch, geometrisch argumentieren. Dabei würde aber nie auf den Akt des **Einzeichnens** der Mittelpunkte fokussiert.

welcher Mittelpunkt zu welcher Seite gehört. Es erschwert aber die sprachliche Wahrnehmung des Textinhalts zugunsten der Textkürze - die wiederum eigentlich die inhaltliche Wahrnehmung erleichtern sollte. Eine latente Erleichterung besteht also nur für den denjenigen, der sich hier bereits relativ sicher im Problem bewegt.

Die Mittelpunkte werden mit E und F bezeichnet. Mit der alphabetischen Reihenfolge bewegt sich der Text damit weiter innerhalb der Bezeichnungskonventionen. Gleichzeitig werden sie verletzt: D wird ausgelassen.

Exkurs: Die **geometrischen Bezeichnungskonventionen** für Punkte verschleiern ihre Regelhaftigkeit, denn Punktbezeichnungen sind scheinbar „dem Grunde nach“ beliebig. In Wirklichkeit liegen hier Regeln (also bedeutungsgenerierende Elemente) vor. Die Regel lautet, dass man in alphabetischer Reihenfolge vorgeht und bei A (Beginn des Alphabets) oder bei P (für Punkt) beginnt. Man beginnt mit der Bezeichnung, z.B. bei Figuren oder Körpern, links unten in der Darstellung und arbeitet sich dann entgegen dem Uhrzeigersinn vorwärts. Besondere Punktobjekte können abweichend von der Reihenfolge in Anlehnung an ihre Eigennamen bezeichnet werden, z.B. Z für Zentrum, M für Mittelpunkt, P für einen externen oder besonderen (bzw. im Sinne von o.B.d.A. allgemeinen) Punkt. Man kann auch mit Zahlindizes arbeiten, dann ist die Reihenfolge der natürlichen Zahlen einzuhalten. Die Bedeutungshaltigkeit der Bezeichnungen erkennt man einerseits an der besonderen Stellung bestimmter Buchstaben, andererseits an der Rolle von Regelverletzungen in Form von Lücken: Stößt man in einer mit A beginnenden Reihe von Punkten plötzlich außerhalb dieser Reihenfolge auf M, Z oder P, so wird sofort auf den besonderen Charakter dieser Punkte fokussiert. Eine Lücke in der Reihenfolge weist hingegen darauf hin, dass diese Lücke im weiteren Verlauf des konstruktiven Geschehens gefüllt wird oder dass zwei Objekte voneinander getrennt werden sollen.
Auch andere Regelabweichungen sind möglich. Man wird zur Übung des Satzes des Pythagoras ein Dreieck MNO betrachten, damit die Schüler den Satz außerhalb seiner formalen Darstellung benutzen. Hier liegt aber lediglich ein Beispiel für den Charakter von Regeln vor: Von Regeln kann abgewichen werden, aber die Abweichung ist erklärungsbedürftig. Die Nutzung von MNO wird didaktisch gerechtfertigt, und es wird auch kein Dreieck ZBL betrachtet, denn diese doppelt regelabweichende Bezeichnung ist durch den didaktischen Gedanken nicht gerechtfertigt. (An der Sinnhaftigkeit dieser Regeln und dieses Charakters von Abweichungen sind alle didaktischen Versuche gescheitert, eine Art permanente Regellosigkeit bei den Bezeichnungen durchzusetzen, um die Schüler zu flexiblerem Umgang mit Formeln zu zwingen. Außerdem ist die Regelvermittlung selbst Lerninhalt und kollidiert mit Regelverletzungen.)
Scheinbare Regelabweichungen sind möglich, wenn sich verschiedene Ebenen der Betrachtung eines geometrischen Objekts überlagern, wenn man zum Beispiel zum Führen eines Beweises Hilfslinien einzeichnet und die dadurch neu entstehenden Objekte bezeichnet. Aber gerade hier erweist sich die Sinnhaftigkeit der Regeln: Weil regelhaft vorgegangen wird, kann man die Genese des Beweisgedankens besser verfolgen.

Im Aufgabentext sind die Bezeichnungskonventionen mit ABC und mit der alphabetischen Reihenfolge von EF eingehalten, aber mit der Auslassung von D verletzt. Diese Regelabweichung produziert eine sprachliche Wirklichkeit: Sie schafft eine Lücke. Sie separiert das Dreieck ABC vom Gebilde EF. Die Verwendung von D und E statt E und F hätte ein zweites Dreieck CDE „geschaffen". Der vorliegende Text schafft statt dessen ein Dreieck, dem eine Strecke einbeschrieben ist. Damit ist die in der Bildinterpretation eingeführte entsprechende Lesart gestärkt. Damit erhält aber die Strecke EF auch eine Eigenständigkeit. Diese Eigenständigkeit macht diese Strecke beweglich - gerade im Zusammenhang damit, dass die Endpunkte nicht konstruiert sind. Diese Dynamik schwächt den Charakter der Figur als statische Strahlensatzfigur. Auch die Intuition wird durch die Beweglichkeit von EF eher geschwächt, schließlich eröffnet diese Beweglichkeit eine Vielzahl von Denkvarianten. Der Gedanke einer Parallelverschiebung nach unten wird durch diese Beweglichkeit hingegen gestärkt. Der Ansatz des Messens wird lediglich in dem Sinne gestärkt, dass man EF „nimmt" und versucht, zwei Mal in AB reinzulegen. Der Ansatz der Mittellinie des Dreiecks wird durch diesen Charakter von EF wohl am klarsten gestärkt.
Die Regelabweichung an sich stellt eine gewisse Entfernung zu mathematischer Kultur dar. Umso bemerkenswerter ist, dass sie den genuin mathematischen Weg zugunsten des formalen Weges schwächt.

Wie lang ist $\overline{\mathrm{EF}}$ *?*
Hier reproduziert sich die Fallstruktur, deshalb nur Stichworte: *Wie lang ist ...* statt *Bestimme ...* oder *Berechne ...* bedient alle Lösungsmöglichkeiten. Keine elaborierte Sprache (denkbar wäre *Welche Länge hat ..., ... ist die Strecke* $\overline{\mathrm{EF}}$ *?*), auch hier wieder Anschluß an eine Kultur, die keine mathematische ist, sondern nur aus ihr abgeleitet.

Fazit

1. Was wird gemessen?

Es liegt eine geometrische Aufgabenstellung vor.
Es bieten sich mehrere Lösungsmöglichkeiten für die Aufgabe an, von denen allerdings nur zwei - nämlich über den Strahlensatz und über die Mittellinie des Dreiecks - streng geometrische Lösungen sind. Die verschiedenen Lösungswege erfahren durch den Aufgabentext verschiedene Schwächungen und Stärkungen:

- Der Weg des Messens wird durch *... ist 6 cm lang ..., Es werden ... eingezeichnet., Wie lang ...* und z.T. durch die Abweichung von Bezeichnungskonventionen gestärkt. Schwächung erfährt dieser Weg durch die Tatsache, dass AB im Testheft nur etwa 4,4 cm lang ist.
- Der Weg der Intuition wird durch *... ist 6 cm lang* und *Es werden ... eingezeichnet.* gestärkt. Durch die Beweglichkeit von EF wird dieser Weg geschwächt.
- Der Weg über den Strahlensatz wird nicht direkt gestärkt, aber durch die latente Indifferenz der Mittelpunkte erhält er eine erleichternde Redundanz. Er wird geschwächt durch die Abweichung von Bezeichnungskonventionen und durch den expliziten Verweis auf das Dreieck ABC.

- Der Weg über die Mittellinie des Dreiecks wird durch den expliziten Verweis auf das Dreieck ABC, durch das Einzeichnen der Mittelpunkte von Dreiecksseiten (nicht von Strecken), durch die Abweichung von Bezeichnungskonventionen und durch die latente Indifferenz der Mittelpunkte gestärkt und nahegelegt. Es erfährt keine direkte Schwächung.

Man kann also konstatieren, dass der Weg über die Mittellinie des Dreiecks vom Text am stärksten gestützt wird und die implizite Lösungspräferenz der Aufgabenstellung darstellt. Die anderen Lösungswege erfahren aber keine Schwächung, die sie völlig ausschließen würde. Es ist also nicht zu entscheiden, welche der angesprochenen Fähigkeiten gemessen wird, eine Zuordnung der Aufgabe zu einem einzigen „Schwierigkeitswert“ im Sinne des Kompetenzstufenmodells ist nicht sinnvoll möglich.

Man kann zwar keine Schwergewichtigkeit der Stärkungen und Schwächungen angeben, aber der genuin mathematische Weg erfährt viele Schwächungen und keine direkte Stärkung, wohingegen der Weg des lokalen Formel- bzw. Satzwissens keine direkte Schwächung, wohl aber mehrere Stärkungen erfährt.

Die PISA-Gruppe ordnet diese Aufgabe in den Typ 1A „technische Aufgaben“: „Die Aufgabe erfordert nur technische Fertigkeiten und/oder den Abruf von Faktenwissen.“ (Neubrand u.a. 2001, S. 52) Diese Einordnung läßt sich nur für den Weg übers Messen und über die Mittellinie des Dreiecks bestätigen. Intuitives Vorgehen ist im Kategoriensystem der PISA-Gruppe nicht vorgesehen, und der Weg über den Strahlensatz würde in die Kategorie 2A eingeordnet werden müssen: „Der zur Lösung erforderliche Schritt ist überwiegend begrifflicher Art. Die Lösung der Aufgabe kann durch Anwendung eines einzigen begrifflichen Arguments, durch Herstellung eines begrifflichen Zusammenhangs erfolgen. Dabei genügt es auch, den erforderlichen Begriff als Wissensbestandteil abzurufen.“ (ebenda)

Für die Lösungshäufigkeit dieser Aufgabe liegt nur der nationale Wert von 33 % vor, das sind 618 Punkte und verweist auf Kompetenzstufe IV. Für technische Items schreibt die PISA-Gruppe den technischen Items auf Stufe I bis III folgende Eigenschaften zu: In den Aufgaben wird Faktenwissen bzw. Wissen über einfache Konventionen abgefragt, Umgehen mit elementaren mathematischen Begriffen und Verfahren der Sekundarstufe I wird verlangt, etwa Basiswissen über Funktionsvorschriften (z.B. Zahlenwerte einsetzen können). Auf Stufe IV kommt nun hinzu: Mathematische Grundtechniken, wie etwa Termumformungen, sind auszuführen.

Man erkennt, dass diese Zuschreibungen auch dann problematisch sind, wenn man lediglich die wirklich technischen Lösungswege über Messen und Mittellinie betrachtet: Das Messen gehört von seinem technischen Anspruch her auf Stufe I, der Weg über die Mittellinie des Dreiecks auf Stufe III. Die angebliche neue Qualität von Stufe IV ist es nicht, was hier die höhere Schwierigkeit erzeugt. Die Interpretation deutet eher darauf hin, dass die Schwierigkeit hier durch die Abseitigkeit des Formel- und Faktenwissens hergestellt wird, auf die die Aufgabe in besonders intensiver Weise zielt. Würde man hier Unterschiede in den Lösungshäufigkeiten verschiedener Populationen deuten wollen, so würde man eher auf verschiedene Häufigkeiten im Anwenden des intuitiven und messenden Vorgehen verweisen.

2. Aspekte von Testfähigkeit und Testpragmatik

Die Wege des außerhalb mathematischen Denkens liegenden Messens bzw. der Intuition und der Weg des lokalen Formel- bzw. Satzwissens erhalten die höchsten Zeitprämien. Der genuin mathematische Weg wird - von der Zeitprämie her gesehen - bestraft. Testpragmatisch verschafft mathematisches Vorgehen hier also Nachteile.

Der Weg des lokalen Formel- bzw. Satzwissens wird durch die Aufgabe am stärksten herausgefordert. Was kann nun der Schüler tun, der keine Kenntnis des lokalen Phänomens der Mittellinie des Dreiecks hat? Er kann einerseits den Weg über den Strahlensatz gehen - wenn er kann und wenn er die Zeitstrafe in Kauf nimmt. Er kann andererseits die nicht strengen Wege gehen. Sie verlangen aber entweder Unkenntnis von oder Unverfrorenheit gegenüber den Anforderungen geometrischen Arbeitens. Der Schüler, der nicht weiß, dass er hier gar nicht messen oder intuitiv „raten" darf bzw. soll, kann genauso erfolgreich sein wie derjenige, der das weiß, es aber trotzdem tut. Im Fall von Unverfrorenheit kann man von Testfähigkeit sprechen, im Fall von Unkenntnis wohl eher von Glück mit dieser Aufgabe. Das Gemeinsame der beiden Schüler wäre, dass sie über die Fähigkeit verfügen, einfach loszulegen. Der Unterschied ist, dass der zweite Schüler das Irritationspotential der Aufgabe explizit überwindet: Der Schüler gerät in einen Konflikt mit einem vorhandenen Geometrieverständnis.

3. Welches Bild von Mathematik wird hier transportiert?

Schon die Tatsache, dass der Lösungsweg über lokales Satz- und Formelwissen von der Aufgabe am klarsten herausgefordert wird, erzählt über einen bestimmten Umgang mit dem Mathematischen: Das Mathematische, aber auch das Intuitive wird damit tendenziell zugunsten des Technischen bzw. auswendig gelernten zerstört.

Diese Logik reproduziert sich im Text. Äußerlich arbeitet der Text mit Fachsprache und innerhalb von Bezeichnungskonventionen ($\overline{AB}$ - Begrenzung der Strecke durch Punkte und Streckencharakter; ABC und EF innerhalb geometrischer Bezeichnungskonventionen, 6 cm, Begriff Mittelpunkte). Er bricht aber ohne inhaltliche Rechtfertigung Bezeichnungskonventionen und verwendet bei genauerem Hinsehen eine wenig elaborierte Sprache an der Grenze zur Fachsprachlichkeit. Die mathematische Idee des Mittelpunktes wird sprachlich zerstört, ebenso die mathematische Idee der Strecke. Damit kann man etwas zugespitzt die Kultur beschreiben, in die hier eingeführt wird: Unter dem äußerlichen Anschein eines mathematisch aussehenden Problems ist es zwar auch möglich, ein mathematisches Problem zu lösen, aber latent wird der mathematische Charakter des Problems und der Objekte, die das Problem konstituieren, nicht konsistent aufrecht erhalten.

Die sich hier zeigenden Elemente eines mathematikdidaktischen Habitus zeigen eine neue Dimension: Bei der Aufgabe „Bauernhöfe" entstand die Zerstörung des Mathematischen wesentlich aus den Verwerfungen von Realem und mathematischem Modell. Bei „Dreieck" haben wir es mit einem rein mathematischen Problem zu tun. Umso erstaunlicher ist, dass auch hier das Mathematische (an und für sich oder zugunsten des Technischen) unterlaufen wird und dass sogar das äußerlich so vorstechende Fachsprachliche latent unterminiert wird.

5.7. Der Vergleich der Aufgaben „Dreieck“, „Bauernhöfe“ und „Pyramide“ im theoretischen Konzept von PISA

Im Vergleich zur Aufgabe „Dreieck“ konnten bei „Bauernhöfe“ nur wenige textliche Beeinflussungsfaktoren *für bestimmte Lösungswege* eruiert werden. Im Text findet sich lediglich drei Mal eine Förderung des intuitiven Lösens durch Symmetriekonstruktionen. Alle anderen Hinweise finden sich im Problem selbst und in der Zeichnung.

Das Problem bei „Bauernhöfe“ ist insofern eindeutig schwieriger als bei „Dreieck“, als hier eine räumliche Problemkonstellation vorliegt. Diese schwächt zunächst alle Lösungswege. Andererseits haben wir es bei „Bauernhöfe“ mit gleichseitigen Dreiecken zu tun. Die Problemstellung entfaltet bei gleichseitigen Dreiecken eine ausgesprochen starke suggestive Kraft, die intuitives Vorgehen stärkt.

In der Aufgabe „Dreieck“ wird latent permanent auf die Mittellinie des Dreiecks fokussiert. Der Schüler wird dort auf den Weg von lokalem Satz- und Formelwissen geführt, der bei Abwesenheit dieses Wissens dann zum Scheitern führt. In der Aufgabe „Bauernhöfe“ erfolgt diese Fokussierung nicht.

Es zeigen sich in den Interpretationen also zwei Gründe, warum „Bauernhöfe“ im Vergleich zu „Dreieck“ „einfacher“ sein könnte: Stärkung des intuitiven Vorgehens durch Text und Problem und Schwächung des potentiell in eine Sackgasse führenden Weges über lokales Formelwissen. Dem stehen nun zwei Argumente entgegen, die „Bauernhöfe“ gegenüber „Dreieck“ schwieriger machen: Das Problem ist in den Raum versetzt und die Aufgabe ist sprachlich enorm künstlich verschwierigt. Der Schüler muß sich durch Irritationen und Verschleierungen hindurcharbeiten. Den deutschen Schülern gelingt das erheblich schlechter (41 % Lösungshäufigkeit) als dem internationalen Durchschnitt (55 % Lösungshäufigkeit). Es gelingt ihnen aber immer noch erheblich besser als bei „Dreieck“ (33 % Lösungshäufigkeit). Diese Zahlen sind natürlich beliebig interpretierbar, denn wir wissen nach wie vor nicht, welche der vielen möglichen Gründe für die Ergebnisunterschiede wie stark gewirkt haben. Man könnte aus unserem Wissen ein Lob der Kraft der Intuition machen oder ein Argument gegen die Singularität von lokalem Formelwissen. Die Interpretation der PISA-Gruppe ist jedenfalls nicht überzeugend:

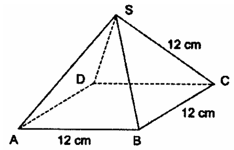
> „Bezeichnenderweise ist aber die Berechnung der Mittellinie des Dreiecks etwas schwerer als die analoge „Bauernhof“-Aufgabe. In der kontextualisierten Version scheint es den Schülerinnen und Schülern leichter zu fallen, auf geometrische Kenntnisse zurückzugreifen oder die Lösung intuitiv abzuschätzen.“ (ebd., S. 151)

Diese Aussage leuchtet wenig ein. Warum sollte es den Schülern in der kontextualisierten Version leichter fallen, auf geometrische Kenntnisse zurückzugreifen oder die Lösung intuitiv abzuschätzen? Es liegt viel näher anzunehmen – und würde mit bisherigen Forschungsergebnissen korrespondieren – dass es ihnen zunächst sogar schwerer fällt, da der zusätzliche Schritt vom Raum in die Ebene zu bewältigen ist. Hinzu kommt die mißlungene Kontextualisierung mit den angesprochenen Verwerfungen.

Die Rolle der Gleichseitigkeit für das intuitive Vorgehen kann man auch ohne Objektive Hermeneutik erkennen. Die Rolle des Textes wird durch Objektive Hermeneutik methodisch kontrolliert beschreibbar. Die oben gegebenen Deutungen der unterschiedlichen Lösungshäufigkeiten sind textlich und inhaltlich aus der Aufgabe heraus begründet, wohingegen die PISA-Deutung lediglich versucht, durch eine oberflächliche Betrachtung der Aufgabe die Lösungshäufigkeiten zu „erklären“. Die PISA-Gruppe zieht kein einziges Element ihres eklektischen Theoriegebäudes zur Deutung der Resultate heran. Dies verweist nicht nur auf den Charakter dieses Theoriemixes als theoretisches Mäntelchen, welches dem Test umgelegt wird. Es verweist auch auf eine wissenschaftliche Praxis, die ad-hoc-Deutungen mehr Gewicht beimißt als theoretisch eingebundenen Deutungen, welche mit den Elementen des Theoriegebäudes durchaus möglich wären.

Die PISA-Gruppe vergleicht die Aufgabe „Bauernhöfe 1“ mit der nationalen Aufgabe „Pyramide 1“, die sie als dekontextualisierte Version betrachtet:

Die Grundfläche einer Pyramide ist ein Quadrat. Jede Kante der skizzierten Pyramide misst 12 cm. Berechne den Flächeninhalt der Grundfläche ABCD.



> „Im Fall der Standardmodellierung (Berechnung der Grundfläche der Pyramide bzw. der Dachbodenfläche) zeigen beide Varianten praktisch denselben Schwierigkeitsgrad. Bezeichnenderweise ist aber die Berechnung der Mittellinie des Dreiecks etwas schwerer als die analoge „Bauernhof“-Aufgabe. In der kontextualisierten Version scheint es den Schülerinnen und Schülern leichter zu fallen, auf geometrische Kenntnisse zurückzugreifen oder die Lösung intuitiv abzuschätzen.“ (PISA 2000, S. 151)

Die Deutung des gleichen Schwierigkeitsgrades für „Bauernhöfe 1“ und „Pyramide 1“ scheint nun der objektiv-hermeneutische Interpretation von „Bauernhöfe“ zu widersprechen, denn die Irritationen, sprachlichen Verwerfungen und Verstellungen gelten auch für „Bauernhöfe 1“. Andererseits gilt der Vorteil der besonderen Suggestivität einer intuitiven Lösung für das Flächeninhaltsproblem nicht. „Bauernhöfe 1“ müßte also schwerer sein, wenn nicht in „Pyramide 1“ noch eine andere Schwierigkeit lauert. Mit dem bisherigen Wissen muß man nun keine neue objektiv-hermeneutische Interpretation leisten, es genügt eine abgekürzte Interpretation: Der erste Satz erzeugt zwar Redundanz zur Zeichnung, enthält aber keine Verwerfung, denn die Grundfläche einer Pyramide kann ein Quadrat sein. Der zweite Satz wurde bereits eigenständig interpretiert, das zusätzliche „skizziert“ erzeugt allerdings eine abweichende Schwierigkeit: Plötzlich wird die Option des Vorhandenseins zweier Pyramiden eröffnet, nämlich der Pyramide des ersten Satzes und der skizzierten Pyramide. Wohlgeformt könnte es zum Beispiel heißen: *Die Grundfläche der skizzierten Pyramide ist ein Quadrat. Jede Kante der Pyramide misst 12 cm.* Hier wird also Irritationspotential erzeugt, allerdings offenbar in weitaus geringerem Maße als bei „Bauernhöfe“.

Auch die Aufforderung wurde bereits interpretiert. Im Text steckt keine im Vergleich zu „Bauernhöfe“ zusätzliche Schwierigkeit, warum sind beide Aufgaben also gleich schwer?

Sie sind „gleich schwer“, weil die PISA-Gruppe ihr Berechnungsmodell für die „Schwierigkeits“werte nicht reflektiert. Die internationale Lösungshäufigkeit für „Bauernhöfe 1“ beträgt 61 %, die nationale 51 %. Die nationale Lösungshäufigkeit für „Pyramide 1“ beträgt 59 %, eine internationale Lösungshäufigkeit gibt es nicht, weil die Aufgabe nur in Deutschland gestellt wurde. Die fast gleichen

„Schwierigkeitswerte“ (Bauernhöfe 1: 492, Pyramide 1: 503) für die beiden Aufgaben ergeben sich daraus, dass für die internationale Aufgabe die internationale, für die nationale Aufgabe die nationale Lösungshäufigkeit zur Berechnung herangezogen wurde. Hier liegt ein Beispiel dafür vor, welche Fehldeutungen sich daraus ergeben: In einer vergleichbaren Population, nämlich bei den deutschen Schülern, wird nämlich „Bauernhöfe 1“ mit 51 % Lösungshäufigkeit seltener gelöst als „Pyramide 1“ mit 59 % Lösungshäufigkeit. Dieser schlichte Befund – der durch die Interpretationen eine Erklärung findet – wird durch die Behauptung von „Schwierigkeitswerten“ nicht nur zugeschüttet, sondern sogar geleugnet.

Zur Begründung des weiteren fallkontrastiven Vorgehens bei der Auswahl der zu interpretierenden Aufgaben möchte ich die für die Fragestellung der Arbeit zentralen Charakteristika der bisher interpretierten Aufgaben zusammenfassen:

Die Aufgabe „Bauernhöfe“ geht – äußerlich betrachtet – von einem realen Problem aus. Es zeigte sich eine Struktur der Zerstörung der Modellierungsanforderung und des Mathematischen bei gleichzeitiger intensiver Verwendung von Fachsprache. Es werden Zeitprämien für die Abkehr von mathematischem Denken und für das Zurückgreifen auf Intuition, auf Messen oder auf lokales Satz- und Formelwissen vergeben. Dieser testpragmatische Aspekt von Testfähigkeit ist gepaart mit hohem Irritationspotential. Bei „Bauernhöfe 1“ ist eindeutig zu benennen, welche mathematische Fähigkeit gemessen wird. Bei „Bauernhöfe 2“ ist das nicht möglich.
Die Aufgabe „Dreieck“ wurde interpretiert, weil sie das mathematische Problem von „Bauernhöfe 2“ aufnimmt, wobei auf einen äußerlich realen Kontext verzichtet wird. Auch hier findet sich eine Struktur der Zerstörung des Mathematischen bei gleichzeitiger intensiver Verwendung von Fachsprache. Zeitprämien werden vergeben für die Abkehr von mathematischem Denken und die Zuwendung zu lokalem Satz- und Formelwissen, zu Messen und Intuition. Es tritt weit weniger sprachliches Irritationspotential auf als bei „Bauernhöfe“, Testfähigkeit ist hier eher die Fähigkeit zum pragmatischen Umgehen mit den Zeitprämien. Auch bei „Dreieck“ ist nicht eindeutig zu benennen, welche mathematische Fähigkeit gemessen wird.
Die Aufgabe „Pyramide“ stellt das gleiche mathematische Problem wie „Bauernhöfe 1“, aber ohne äußerlich realen Kontext. Hier liegt eigentlich ein Kandidat für eine Aufgabe vor, bei der einerseits eindeutig benennbar ist, welche mathematische Fähigkeit gemessen wird, bei der andererseits keine sprachlichen Verwerfungen auftauchen, so dass kein Irritationspotential vorhanden ist. Die irritationserzeugende Struktur wirkt aber auch hier: Die Hinzusetzung des einzigen Wortes „skizzierten“ ruft eine sprachliche Verwerfung hervor.
Es ist also weiterhin eine Aufgabe zu finden, die irritationsfrei ist. In ihr sollte das Mathematische nicht beschädigt werden.[65] Der Blick für in den Aufgaben steckendes Irritationspotential ist durch die bisherigen Interpretationen bereits geschärft, das erleichtert die Suche nach einer äußerlich irritationsfreien Aufgabe:

[65] Eigentlich wäre nach einer Aufgabe zu suchen, bei der zusätzlich noch ein eindeutig benennbarer Meßprozeß stattfindet. Wir wissen aus „Bauernhöfe 1“, dass man mit anspruchsloseren Aufgaben eindeutig oder „scharf“ messen kann, d.h. relativ genau benennen kann, welche Fähigkeit man mißt. Bei den anspruchsvolleren PISA-Aufgaben zeigt sich aber, dass sie mehrdeutig messen, dass also eindeutiges Messen bei ihnen nicht auftritt.

5.8. Die (nationale) PISA-Aufgabe „Feriengeld“

Wenn Beate jeden Tag 10 DM ausgibt, reicht ihr Feriengeld für 9 Tage.
Wie lange reicht es, wenn Beate nur 6 DM täglich ausgibt?

- □ *5 Tage*
- □ *10 Tage*
- □ *15 Tage*
- □ *20 Tage*
- □ *25 Tage*

Lösungswege

- Arbeiten mit dem Gesamtgeldbetrag: Mit 10 $\frac{\text{DM}}{\text{Tag}}$ reicht das Geld 9 Tage, es sind also 90 DM da. Man kann nun sukzessive 6 DM für jeden Tag abziehen und sehen, wie weit man kommt (fehleranfällig und aufwändig). Man kann diesen Prozeß abkürzen, indem man überlegt, wie oft die 6 DM in die 90 DM „reinpassen“, wie oft dieses sukzessive Abziehen also erfolgt (Hilfe: Multiple-Choice-Angebote). Eventuell überblickt der Schüler dieses Aufteilungsproblem auch sofort, erkennt also sofort die Notwendigkeit einer Division bzw. der Lösung eines Problems wie „x mal 6 DM = 90 DM, wie groß ist x?“ Eine weitere Möglichkeit ist, dass der Schüler an dieser Stelle 90 durch sechs teilt, weil dies die einzige einfache Operation ist, die zu einem sinnvollen Resultat führt. Formalisierung dieses Wegs: $10\,\frac{\text{DM}}{\text{Tag}} \cdot 9 \text{ Tage} = 6\,\frac{\text{DM}}{\text{Tag}} \cdot x \text{ Tage}, \quad x = 15$

- Kapitänsweg: Auch die ohne jedes inhaltliche Verständnis arbeitende Operation 9+6 führt auf die richtige und in den Multiple-Choice-Angeboten vertretene 15. Im Sinne der Verarbeitung als Kapitänsaufgabe kann man unter Mitnutzung der 10 auch noch zu 5 Tagen und zu 25 Tagen gelangen, das verringert die Wahrscheinlichkeit, auf dem Kapitänswege zufällig zur richtigen Lösung zu gelangen.

- Bezug auf einen Tag: 10 $\frac{\text{DM}}{\text{Tag}}$ entspricht 9 Tagen, also entspricht 90 $\frac{\text{DM}}{\text{Tag}}$ einem Tag. Jetzt kann man entweder zur Betrachtungsweise mit dem Gesamtgeldbetrag übergehen oder den Vergleich zu den geforderten 6 $\frac{\text{DM}}{\text{Tag}}$ ziehen, z.B. durch eine Dreisatzbetrachtung: Um von 90 $\frac{\text{DM}}{\text{Tag}}$ auf 6 $\frac{\text{DM}}{\text{Tag}}$ zu kommen, muß man durch 15 teilen, also muß man den einen Tag mal fünfzehn nehmen. Die Notwendigkeit der Umkehrung von Division zu Multiplikation kennt man dabei oder erarbeitet bzw. errät sie aus Plausibilitätsgründen.

- Differenzbetrachtungen: Pro Tag soll Beate 4 DM weniger ausgeben. Sie hat also nach 9 Tagen 36 DM weniger ausgegeben. Sie hat also Geld für weitere 6 Tage, insgesamt also für 15 Tage.

  Probe: $15 \cdot 6 = 90 = 10 \cdot 9$

Mit Differenzbetrachtungen kann auch sukzessiv gearbeitet werden: Gibt Beate pro Tag eine Mark weniger aus, so sind nach 9 Tagen noch 9 Mark übrig, gibt sie pro Tag zwei Mark weniger aus usw.

- Abschätzen: Wenn täglich 10 Mark für 9 Tage reichen, dann reichen täglich 6 Mark für etwas weniger als doppelt so viele Tage: Bei täglich 5 Mark wären es genau doppelt so viele Tage. Für mehr als 9 Tage und weniger als 18 Tage gibt es die Multiple-Choice-Angebote 10 Tage und 15 Tage. Davon kommt nur das Angebot 15 Tage in Frage – dies kann intuitiv oder über verschiedene Plausibilitätsbetrachtungen bzw. über Rechnungen entschieden werden. Dieses Vorgehen ist im Lernprozeß ein ausgesprochen fruchtbarer Schritt zum flexiblen Umgang mit umgekehrt proportionalen Problemen. Dieses Vorgehen hat seine Grenzen aber dort, wo es um genaue Lösungen geht und keine Antwortalternativen vorliegen.
Der Weg des Abschätzens würde verbaut, wenn z.B. 14, 15 und 16 zur Auswahl stünden.

- diverse Kombinationen, z.B.: 1 Tag = $90 \frac{\text{DM}}{\text{Tag}}$, 2 Tage = $45 \frac{\text{DM}}{\text{Tag}}$, ..., 10 Tage = $9 \frac{\text{DM}}{\text{Tag}}$. Pro Tag sollen es aber 3 DM weniger sein, also 30 DM übrig, also 5 Tage mehr. An dieser Stelle kann auch Multiple Choice zur Hilfe genommen werden: 20 Tage = $4{,}50 \frac{\text{DM}}{\text{Tag}}$, das ist zu wenig, hingegen sind $9 \frac{\text{DM}}{\text{Tag}}$ zu viel, also bleibt nur das Lösungsangebot dazwischen: 15 Tage.
- formales Arbeiten mit umgekehrter Proportionalität: $10 \frac{\text{DM}}{\text{Tag}} \cdot \frac{6}{10} = 6 \frac{\text{DM}}{\text{Tag}}$, also 9 Tage $\cdot \frac{10}{6} = 15$ Tage.

Die Lösungsalternativen zeigen, dass selbst bei Inanspruchnahme der durch Multiple Choice gegebenen Lösungshilfen und bei reinen größenordnungsbezogenen Plausibilitätsbetrachtungen eine Anforderung erhalten bleibt: Die Vorstellung des „monotonen Fallens“, also die Vorstellung, dass mit sinkendem täglichem Betrag die Anzahl der Tage steigt.
Durch die Lösungsvielfalt wird diese eingekleidete Aufgabe zu einer offenen Aufgabe: Ausgehend vom aus dem Alltagsverstehen ableitbaren Erkennen des „monotonen Fallens in regelmäßigen Schritten“ kann unterrichtlich über das Aufzeigen mehrerer Lösungswege die Schärfe und kalkülmäßige Einfachheit der formalsten Wege - über die Formel zur umgekehrten Proportionalität bzw. über die Formalisierung des Arbeitens mit dem Gesamtgeldbetrag - aufgezeigt werden. Hier zeigt sich auch, dass die Offenheit einer Aufgabe (insbesondere im Unterricht) nur dann zur Geltung kommen kann, wenn den verschiedenen Wegen auch Raum zur Verfügung steht. Eine offene Aufgabe erzwingt noch keinen offenen Unterricht.
Gleichzeitig sorgt diese Offenheit im Test dafür, dass man hier sehr unterschiedliche Fähigkeiten gleichzeitig mißt und nicht weiß, welche der Fähigkeiten man konkret mißt. Man weiß immerhin, dass man die Fähigkeit des Umgehens mit dem „monotonen Fallen“ mißt. Allerdings eröffnet Multiple Choice nicht nur die Option des Abschätzens, sondern sogar die Möglichkeit des Lösens ohne jedes inhaltliche Verstehen. Für einen eindeutigeren Meßprozeß müßte also das Zahlenmaterial verändert oder eine offene Antwort abgefordert werden.
Die Sach- und die manifeste Sinnstruktur der Aufgabe scheinen keine spezifische Fragestellung für eine objektiv-hermeneutische Interpretation zu ergeben. Insbesondere ist kein Irritationspotential erkennbar, so dass die Aufgabe als Kandidat für eine irritationsfreie Aufgabe gelten kann. Die Interpretation widmet sich also nur den sonstigen Fragen und dient ansonsten dazu, diese äußerliche Verwerfungslosigkeit zu untersuchen.

Textanalyse

Wenn Beate jeden Tag 10 DM ausgibt, ...

Manifest handelt es sich um eine Bedingungsangabe, die eine durch *dann* eingeleitete Folgerung erfordert.
Zu dieser Sequenz lassen sich verschiedene Arten von Geschichten erzählen:
a) Abschätzungsproblem: *Wenn Beate jeden Tag mit der Straßenbahn fährt, dann rentiert sich ein Monatsticket.* Hier liegt ein Gedankenexperiment vor. Es ist nicht davon auszugehen, dass Beate wirklich jeden Tag mit der Straßenbahn fährt, man kann sogar davon ausgehen, dass sie es nicht tut.
b) Betrachtung von Folgen: *Wenn Beate jeden Tag ein Ei ißt, wird sie wieder gesund/dick/herzkrank usw. Wenn Beate jeden Tag fünf Mark spart, hat sie am Jahresende Geld für die Party.* Auch hier liegt ein Gedankenexperiment vor.
Der Charakter als Gedankenexperiment wird auch kontrastiv deutlich, wenn man mit Geschichten arbeitet, die wirklich tägliche Beschäftigungen aufnehmen: *Wenn Beate jeden Tag in den Spiegel sieht ..., Wenn Beate jeden Tag atmet ...* Die letzte Konstruktion kann nicht wohlgeformt weitergeführt werden (es sei denn ironisch), die erste Konstruktion führt wieder auf ein Gedankenexperiment: *..., dann kann sie das Auf und Ab des Lebens beobachten, dann vergißt sie das Kämmen nicht* usw.

Im ersten Teil wird also eindeutig und verwerfungsfrei in die Bedingung eines **Gedankenexperiment**s eingeführt. Die Bedingung besteht in der täglichen Periodizität eines zu spezifizierenden Handelns. Hier hat bereits eine Modellierung im Sinne einer Durchschnittsbildung stattgefunden, denn Beate gibt ja nicht jeden Tag exakt 10 DM aus.

... 10 DM ausgibt, ...

Die Bedingung wird spezifiziert. Es werden eine Zahl und eine Einheit eingeführt. Zahlenschreibweise und Einheit bewegen sich innerhalb alltäglicher Textlichkeit, durch ihr Auftauchen als Bedingung innerhalb eines Gedankenexperiments ist hier aber schon der Charakter als Rechenaufgabe eingeführt. Die Neutralität von *ausgibt* (vergleiche mit verschwendet, rausschmeißt, verteilt, verschenkt) stärkt diesen Charakter.
Was kann nun folgen? Entweder eine Fragestellung: *Wenn Beate jeden Tag 10 DM ausgibt, wie lange reichen dann 20 Mark?* Oder eine Schlußfolgerung bzw. eine weitere Spezifizierung der gedankenexperimentellen Problemlage.

..., reicht ihr Feriengeld für ...

Die Problemsituation wird weiter spezifiziert: Es geht um die Reichweite der Finanzen und es geht um Feriengeld - eine ungewöhnliche, innerhalb der (bis hier durch die Tagesperiodizität und die Größenordnung 10 DM spezifizierten) Situation gleichwohl eindeutige Bezeichnung für Taschengeld in Ferien. Es handelt sich um eine lebenspraktisch authentische Formulierung im Sinne von: Wie lange reicht unser Urlaubsgeld?

... reicht für ... ist wiederum neutral und zeichnet die denkbaren Weiterführungen bereits vor:
... sie und ihre Schwester.
... sieben Tage.
... ein Drittel der Urlaubszeit.
Alle denkbaren Weiterführungen zeigen an, dass nun eine Schlußfolgerung gesetzt wird, mit der dann lediglich eine Feststellung im Raum steht. Die Ausgestaltung im Text ist

... 9 Tage.

Auch hier werden eine Zahl und eine Einheit innerhalb alltäglicher Textlichkeit verwendet. Der bereits eingeführte Charakter als Rechenaufgabe wird fortgeführt. Die Grundsituation ist jetzt eingeführt. Da bereits relativ klar ist, dass eine Rechenaufgabe vorliegt, sind nur eine weitere Spezifizierung des Experiments oder eine Fragestellung möglich.

Wie lange reicht es,...

Wie lange ... führt die Fragestellung ein. Die Einheit wird nicht vorgegeben, d.h. dass der Schüler die Aufgabe hat, die Einheit der Lösung selbst zu finden.
... reicht es, ... Die Wiederholung des *reicht* knüpft an den ersten Satz an und verbindet die Fragestellung sprachlich mit der eingeführten Situation. Dies vermindert die Aufgabenschwierigkeit im Vergleich z.B. zu *Wie lange kommt sie mit ... aus.* Gleichzeitig heißt es aber nicht *Wie lange reicht ihr Feriengeld ...* In dieser Konstruktion wäre der Bezug auf die eingeführte Situation doppelt gegeben, das würde die sprachliche Anforderung weiter verringern, da die inhaltliche Erfassung des *es* entfiele. Die Interpretationen der TIMSS-Aufgaben haben immer wieder Sprachmuster gezeigt, in denen diese sprachlichen Doppelungen auftraten, die dort als Beschleuniger interpretiert wurden. Hier wird auf eine solche Beschleunigung des Lesens und Erfassens der Aufgabe durch Formulierungswiederholung verzichtet.
An dieser Stelle ist zwar die Fragestellung eingeführt, aber noch nicht die Bedingung, unter der die Fragestellung beantwortet werden soll.

..., wenn Beate nur 6 DM täglich ausgibt?

Die Bedingung wird eingeführt. Dabei wird die Ausgangssituation aufgenommen, kontrastiv wäre *..., falls sie nur ...* denkbar. Diese Aufnahme stellt eine Parallelität der beiden Situationen her, die durch die *6 DM* und durch *ausgibt* verstärkt wird. Diese Parallelität mündet allerdings nicht in völlige Gleichheit, da *jeden Tag* durch *täglich* ersetzt ist. Dabei wurde auf die explizite Verwendung einer Einheit *pro Tag* verzichtet, was eine Arbeit mit dieser Einheit forciert hätte. Auch im Vergleich mit *jeden Tag* verlangt *täglich* eine etwas eigenständigere Handhabung des Einheitenproblems. Dies verstärkt die Logik des *Wie lange*, das ebenfalls auf eine Einheit verweist, aber diese Einheit nicht expliziert.

Das *nur*, welches auch weggelassen werden könnte, stellt einen echten Bruch der Parallelität des ersten und des zweiten Satzes dar. Es gibt einen expliziten Richtungshinweis, es verweist nämlich auf die Verringerung des täglichen Betrages. Dies stellt im Vergleich zum Weglassen eine Vereinfachung der Aufgabenerfassung dar und gibt einen Hinweis auf den Charakter des monotonen Fallens. Insbesondere wird der Fehler eines proportionalen Vorgehens geschwächt. Das *jeden Tag* und das *täglich* verweisen auf Periodizität.

Zusammenfassung

Insgesamt liegt ein in sich stimmiger Text ohne Irritationspotential vor. Es ist also gelungen, den Fall einer irritationsfreien Aufgabenstellung zu finden, bei der keine Verwerfungen zwischen latentem und manifestem Text auftreten. Die Aufgabe ist gleichzeitig ein gegenüber den Aufgaben „Bauerhöfe“ und „Dreieck“ kontrastierender Fall eines mathematikdidaktischen Habitus und des Umgangs mit Mathematik: Es wird auf Prätentionen mathematischer Fachsprachlichkeit verzichtet, gleichzeitig treten keine Zerstörungen des Mathematischen auf. Das inhaltliche Problem wird ernst genommen. Dabei wird nicht suggeriert, dass hier ein reales Problem zu lösen sei, bezüglich der Realität des Problems verhält sich der Text neutral. Dies ist problemadäquat, denn es handelt sich um eine eingekleidete Aufgabe. Es wird auf die Prätention eines realen Anspruchs verzichtet, gleichwohl kann diese Aufgabe im Sinne einer Modellrechnung reales Problemlösen unterstützen. In diesem Sinne liegt – wenn man die Multiple-Choice-Konstruktion ausnimmt – ein Gelingensfall des Umgangs mit Realität vor: Sowohl das Mathematische als auch das Reale werden in ihrer Autonomie ernstgenommen. Sie gehen eine Synthese ein. – Es wäre auch wenig sinnvoll, umgekehrte Proportionalität in einer Aufgabe bearbeiten zu wollen, die kein in die Realität orientiertes Problem bearbeitet, sondern sich ausschließlich im mathematischen Raum bewegt: Dort ist umgekehrte Proportionalität nur auf einer so abstrakten Ebene diskutierbar, dass es für die angesprochene Altersgruppe inadäquat wäre.

Durch Multiple Choice wird die inhaltliche Ernsthaftigkeit an einer Stelle allerdings zerstört, denn man kann durch reines Operieren mit den gegebenen Zahlen zufällig zum richtigen Ergebnis gelangen.

Die Lösungsalternativen für diese Aufgabe zeigen, dass auch bei Inanspruchnahme der durch Multiple-Choice gegebenen Lösungshilfen und bei reinen größenordnungsbezogenen Plausibilitätsbetrachtungen eine Anforderung erhalten bleibt: Die Vorstellung des „monotonen Fallens“, also die Vorstellung, dass mit sinkendem täglichem Verbrauchsbetrag die Anzahl der Tage steigt, für die Geld zur Verfügung steht.

Wenn man bei dieser Aufgabe überhaupt vom Mitmessen von Testfähigkeit sprechen wollte (weil die vielen Lösungsmöglichkeiten über unterrichtliche Standardlösungen hinausweisen), so ist sie die Fähigkeit, sich in dynamischer Weise in das Problem einzuarbeiten und es unter Nutzung aller verfügbaren Daten zu lösen. Hier laufen also nicht nur Testfähigkeit und mathematische Leistungsfähigkeit zusammen, sondern Testfähigkeit läuft sogar mit mathematisch verstehendem Denken zusammen. Selbst beim Abschätzen ist das noch der Fall. Die Möglichkeit des einfachen Verarbeitens der gegebenen Zahlen (9+6) erfordert hingegen Blindheit oder Unverfrorenheit gegenüber den Anforderungen der Aufgabe.

Im ersten Satz wird ein Gedankenexperiment eingeführt. Seine Bedingung ist die tägliche Periodizität des Ausgebens von 10 DM. Zahlenschreibweise und Einheit bewegen sich innerhalb alltäglicher Textlichkeit, durch ihr Auftauchen als Bedingung innerhalb eines Gedankenexperiments ist hier aber bereits der Charakter als Rechenaufgabe als möglich eingeführt, der durch die Neutralität von *ausgibt* und *reicht* verstärkt wird. Dieser Charakter wird im zweiten Satz durch die von alltäglicher Textlichkeit abweichende Ziffernschreibung der 6 aufgenommen und durch die Einheit DM manifestiert. Dieser Charakterisierung als Rechenaufgabe korrespondiert ein verminderter Anspruch im Vergleich z.B. mit *zehn Mark;* diese Formulierung würde eine Übersetzungsnotwendigkeit beinhalten. Gleichzeitig wird durch *jeden Tag* und *täglich* (im Vergleich zu *pro Tag*) und *wie lange* die Schwierigkeit der Einheitenproblematik offen gehalten.

Der erste und der zweite Satz verlaufen sprachlich zwar eher parallel, was die Erfassung der Aufgabenstellung erleichtert; gleichzeitig laufen sie nicht völlig parallel, was den sprachlichen und (über die Einheitenproblematik) den inhaltlichen Anspruch erhöht. Es findet keine besondere Texterfassungsbeschleunigung statt. Das relative Zusammenlaufen beider Sätze wird durch das nur im zweiten Satz gebrochen. Es gibt einen expliziten Richtungshinweis, es verweist nämlich auf die Verringerung des täglichen Betrages. Dies stellt im Vergleich zum denkbaren Weglassen des nur eine Vereinfachung der Aufgabenerfassung dar und gibt ein Hinweis auf den monoton fallenden Charakter des Problems.

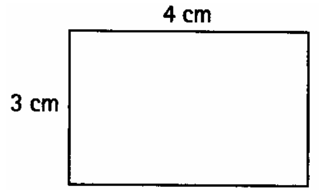
5.9. Weitere Fallauswahl

Innerhalb des fallkontrastiven Vorgehens ist es nun gelungen, eine textlich verwerfungsfreie Aufgabe zu finden. Das mathematische Problem wird in dieser Aufgabe ernst genommen und der mathematische Charakter wird nicht zerstört. Zeitprämien werden dafür vergeben, dass man sich an das Verfahren, dessen Kenntnis gemessen werden soll, annähert. Zwar ist diese Aufgabe „zufällig“ auch durch einfaches Operieren mit den gegebenen Zahlen „richtig“ zu lösen, aber es wurde deutlich, dass dies durch den Verzicht auf Multiple Choice und durch Variation des Zahlenmaterials prinzipiell zu beseitigen ist.

Im Sinne von fallkontrastivem Vorgehen fehlen für eine empirische Sättigung nun noch zwei Aspekte: Es wäre zunächst schön, eine **Aufgabe** zu finden, **für die zusätzlich auch noch eindeutig benennbar ist, welche Fähigkeit sie mißt**. – Schließlich geht es hier um ein Instrument zum Messen mathematischer Leistung. Es muß also ein Meßprozeß benennbar sein, es muß etwas zu leisten sein und das zu Leistende muß mathematisch sein. Wir haben aber bisher noch keine Aufgabe oder Aufgabenkombination gefunden, die „scharf“ mißt, gleichzeitig eine komplexe Fähigkeit mißt und auch noch verwerfungsfrei bezüglich des Mathematischen ist.

In den veröffentlichten PISA-Aufgaben geht eine hohe Meßschärfe aber immer mit geringer mathematischer Reichhaltigkeit einher. Das zeigte sich an der Aufgabe „**Bauernhöfe 1**“. Bei ihr läßt sich aus der richtigen Lösung ablesen, dass der Schüler entweder die technische Anforderung der Multiplikation bzw. des Quadrierens für eine Flächeninhaltsberechnung bewältigt oder das Konzept des Flächeninhalts im Sinne des Auszählens realisiert hat. Die Eindeutigkeit des Meßvorgangs wird zwar durch zwar durch die textlichen Verwerfungen zerstört, aber die Aufgabe ist ohne textliche Verwerfungen konstruierbar.

Ein weiteres Beispiel ist die nationale Aufgabe „**Rechteck“**, die zumindest im Haupttext verwerfungsfrei ist:

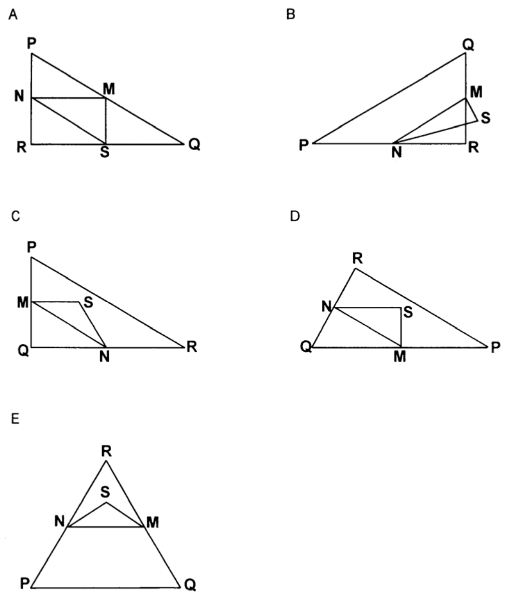


Ein Rechteck ist 4 cm lang und 3 cm breit.
Wie groß ist sein Flächeninhalt?
□ *12 cm*^{2} □ *7 cm* □ *7 cm*^{2} □ *12 cm* □ *14 cm*

(Zeichnung nicht maßgenau!)

Diese nationale Aufgabe (Lösungshäufigkeit 85 %) ist wenig reichhaltig, ein solcher Fall wurde aber bereits mit „Bauernhöfe 1“ untersucht.

Ein interessantes Übergangsphänomen stellt die internationale Aufgabe „**Dreiecke**“ dar (Lösungshäufigkeit international 59 %, Deutschland 65 %):



Kreise die Figur ein, die zur folgenden Beschreibung passt.
Das Dreieck PQR hat einen rechten Winkel in R. Die Strecke $\overline{RQ}$ *ist kürzer als die Strecke* $\overline{PR}$. *M ist Mittelpunkt der Strecke* $\overline{PQ}$ *und N ist Mittelpunkt der Strecke* $\overline{QR}$. *S ist ein Punkt im Innern des Dreiecks. Die Strecke* $\overline{MN}$ *ist länger als die Strecke* $\overline{MS}$.

Hier ist zwar klar, welche einfachen Begriffskenntnisse gemessen werden. Wenn der Schüler scheitert bleibt aber unklar, ob er nicht über die Begriffskenntnisse verfügt oder ob er an der Wiederholung und Kombination der immer gleichen Anforderung scheitert. Die Aufgabe ist also insofern ein Übergangsphänomen, als im positiven Sinne eindeutig klar ist, was gemessen wird, im negativen Sinne aber unklar bleibt, welche Fähigkeit im Falle des Scheiterns *nicht* vorliegt.

Insgesamt gelingt es nicht, für das weitere fallkontrastive Vorgehen eine veröffentlichte PISA-Aufgabe zu finden, die reichhaltig ist, bei der aber gleichzeitig ein eindeutig zu benennender, also „scharfer“ Meßprozeß stattfindet.

Für die empirische Sättigung der kontrastiven Fallauswahl wäre es im weiteren schön, im Sinne des Schließens des Argumentationskreises eine Aufgabe zu finden, die zusätzlich wieder einen realen Kontext in den Blick nimmt, und zwar einen echten realen Kontext. Die Anforderungen an eine solche Aufgabe wurden bereits formuliert: Entweder es wird ausschließlich ein reales Problem in den Blick genommen und eine Fragestellung dazu formuliert. Oder es werden zusätzlich Teile der zu leistenden Modellierung bzw. andere Teile der Lösung angegeben. Das Reale und das Mathematische sollen dabei vermittelt werden. Dabei muß einerseits beides in seiner Autonomie und Authentizität respektiert und aufgenommen werden. Andererseits soll das Reale in das Modell fließen und umgekehrt soll das Modell das Reale widerspiegeln und die mit Hilfe des Modells gewonnenen Erkenntnisse sollen auf das Reale projizierbar sein. Diese Synthese wäre der Gelingensfall, nach dem wir suchen. Dabei ist der erste Schritt, in den Aufgabenstellungen nach einem echten realen Problem zu suchen. Wenn wir ein solches finden, ist die Aufgabenstellung daraufhin zu untersuchen, ob sie wenigstens äußerlich verwerfungsfrei ist. So eine Aufgabe könnte weitere Erkenntnisse über Gelingensfälle bringen oder weitere Elemente des Scheiterns zum Vorschein bringen.

Zusätzlich wäre für die weitere Fallauswahl wünschenswert, zumindest unter den internationalen Aufgaben eine zu finden, die wirklich Mathematical Literacy mißt: „Ziel des PISA-Tests ist es ... zu prüfen, ob Schülerinnen und Schüler grundlegende mathematische Konzepte so verstanden haben, dass sie mit diesen Werkzeugen Problemsituationen aus unterschiedlichen Kontexten behandeln können." (PISA 2000, S.143)

SPAREN (national)

Karina hat 1000 DM in ihrem Ferienjob verdient. Ihre Mutter empfiehlt ihr, das Geld zunächst bei einer Bank für 2 Jahre festzulegen (Zinseszins!) Dafür hat sie zwei Angebote:
a) „Plus"-Sparen: Im ersten Jahr 3 % Zinsen, im zweiten Jahr dann 5 % Zinsen.
b) „Extra"-Sparen: Im ersten und zweiten Jahr jeweils 4 % Zinsen.

Karina meint: „Beide Angebote sind gleich gut." Was meinst du dazu?
Begründe deine Antwort!

Mit dem durch die bisherigen Interpretationen geschärften Blick erkennt man eine Reihe textlicher und inhaltlicher Verwerfungen:
- Warum sollte Karinas Mutter empfehlen, das Geld „zunächst" und gerade für 2 Jahre festzulegen? Das entspricht keinem realen Umgang mit im Ferienjob verdientem Geld. Was hat die Mutter in dieser Aufgabe überhaupt zu suchen?
Warum steht in Klammern Zinseszins? Ruft die Mutter das aus oder ist es die Anmerkung des Lehrers? Auch hier lauert eine Zerstörung des Realen bzw. der Modellierungsanforderung.

- Bei a) beträgt der Zinseszins 5 % von 30 DM, also 1,50 DM, bei b) 4 % von 40 DM, also 1,60 DM.

Liegt hier der wesentliche Vorteil des Anlegens über 2 Jahre? Oder soll man die Differenz von 10 Pfennigen einfach nur nicht vergessen? Das Ausrechnen der Zinsen ohne Zinseszins ergibt für a) 30 DM + 50 DM, für b) 40 DM + 40 DM. Was wollen die Tester jetzt hören: Sind 10 Pfennig Unterschied noch „gleich gut" oder nicht? Die Frage ist ja wegen der unbekannten sonstigen Bedingungen offensichtlich nicht beantwortbar: Selbst bei einer Anlage von 100000 DM wäre der Unterschied ja nur 10 DM, also durch jede Kontoführungsgebühr bzw. andere Nebenkosten, Fahrtkosten zur Bank, selbst durch Mitnehmen von Werbegeschenken ausgeglichen.

Das Problem für den Getesteten ist immer, erfolgreich zu erahnen, was die Tester hören wollen. Hier bleibt das unklar. Es könnte sogar sein, dass man auf eine sinnvolle Argumentation hin einen Punkt bekommt, egal wie man sich entscheidet. Ein Element von Testfähigkeit besteht hier darin, trotz der in ihrer Bedeutung für die Antworterwartung nicht einschätzbaren Zehn-Pfennig-Differenz *irgendetwas* hinzuschreiben. Da das Reale hier ohnehin nicht ernst genommen wird, könnte man den Punkt erhalten, wenn man irgendetwas über Zinseszins hinschreibt – womöglich sogar, ohne ihn berechnet zu haben.

- Ist zur Zehn-Pfennig-Differenz etwas zu „meinen“? Im Sinne mathematischer Bildung kann man hier eine Abwägung der „Bedeutsamkeit“ der Differenz vornehmen. Sachlich ist das eher lächerlich, Bildung kann in dieser Situation auch bedeuten, eine Debatte über ein Zehntel Promille zurückzuweisen. Ein autonomes Subjekt, das dies tun würde, kann sich aber der Folgen seines Tuns nicht sicher sein, weder in der potentiell repressiven Situation von Unterricht noch in der bewertenden Situation eines Tests: In der Aufgabenstellung manifestiert sich ein Habitus des „den Pfennig ehren“. Die Zurückweisung einer Positionierung zu zehn Pfennigen beinhaltet die Zurückweisung dieses Habitus. Im „normalen“ Leben ist so eine Zurückweisung gestattet und üblich, im Unterricht ist sie gefährlich, im Test ist sie „tödlich“, wenn man den Punkt nicht erhält. Die Frage ist aber nicht nur, ob ein Schüler den Punkt erhält, der die Aufgabe von ihrem mathematischen Anspruch her erfüllt und hinschreibt: „Ich weise eine schriftliche Auseinandersetzung mit einer Differenz von zehn Pfennigen zurück.“ Die Frage ist auch, wie die Bepunktung für einen Schüler aussieht, der die zehn Pfennige errechnet hat, aber nicht weiß, was man dazu als „Meinung“ schreiben sollte.

Die Aufgabe fällt also bereits an dieser Stelle als Beispiel für eine irritationsfreie, das reale Problem ernst nehmende und meßscharfe Aufgabe weg. Gleichwohl läßt sich diese Aufgabe in der schlichten Form einer eingekleideten Aufgabe produktiv umdeuten:

Karina hat 1000 DM und möchte sie für 2 Jahre anlegen. Sie hat zwei Angebote:
a) ...   b) ...
Vergleiche beide Angebote. Berücksichtige den Zinseszins.

FAHRRADUNFÄLLE (national)

Eine Zeitung meldet:

70 % aller mit dem Fahrrad verunglückten Kinder sind Jungen. Jungen auf dem Rad sind also stärker gefährdet als Mädchen.

Die Zeitungsmeldung beruht auf folgender Tabelle, in der die 10 000 Schülerinnen und Schüler einer Region, die mit dem Fahrrad zur Schule fahren, nach Geschlecht und Unfallbeteiligung aufgeführt sind.

	Verunglückt	*Nicht verunglückt*	*Insgesamt Jungen/Mädchen*
Jungen	*70*	*8 400*	*8 470*
Mädchen	*30*	*1 500*	*1 530*
Kinder insgesamt	*100*	*9 900*	*10 000*

Beurteile die Zeitungsmeldung mittels der Tabelle:

(1) Die Zeitungsmeldung, dass 70 % aller mit dem Fahrrad verunglückten Kinder Jungen sind, ist

☐ *richtig.*
☐ *falsch.*
☐ *nicht anhand der Tabelle zu beantworten.*

(2) Begründe: Die Zeitungsmeldung, dass Jungen stärker gefährdet als Mädchen sind, ist

☐ *richtig, weil ...*

☐ *falsch, weil ...*

Auch diese Aufgabe fällt als kontrastives Beispiel aus: Auch hier wird die vermeintliche Realsituation nicht ernst genommen: Offensichtlich sind weder die Zeitungsmeldung noch die Daten der Realität entnommen, sondern künstlich für die Aufgabe erstellt worden. Es ist dabei nicht verzeichnet, ob wenigstens die Tendenz bzw. die Größenordnung stimmen. Das Problem ist allerdings nicht, dass die Daten nicht echt sind, sondern dass gleichzeitig behauptet wird, dass sie echt seien: Es ist ja völlig legitim und innerhalb der Logik von Aufgabenerstellung auch authentisch, eine vorhandene „echte" Tabelle zu glätten oder Daten zu erfinden, mit denen es sich leicht rechnen läßt. Hier wird aber so getan, als ob die Wirklichkeit unverfälscht in die Aufgabe geflossen sei. Es ist auch völlig legitim und authentisch, die Beurteilung von künstlich konstruierten Positionen zu verlangen. Die Verwerfung entsteht durch die offensichtlich erfundene und noch dazu schlecht erfundene Zeitungs„meldung", von der auch noch behauptet wird, sie beruhe auf den (in Wirklichkeit ebenfalls erfundenen oder geglätteten) Daten. Es wird also kein authentisches reales Problem betrachtet, aber genau das wird behauptet. Und es wird keine Aufgabe konstruiert, die sich authentisch in der Logik der Konstruktion von Aufgaben bewegt. Es handelt sich hier wiederum um eine Verwerfung zwischen Realität und Modell: Weder gibt es ein authentisch eingebrachtes Reales noch gibt es ein authentisch eingebrachtes Modelliertes. Das Modellierte wird als Reales behauptet und durch eine schlecht konstruierte Zeitungsmeldung (in der die Verwerfung sich reproduziert) auch noch als nicht ernst zu nehmendes herabgewürdigt. Diese mangelnde Ernsthaftigkeit wiederholt sich in (2): Bei der Behauptung, dass Jungen stärker gefährdet sind als Mädchen, handelt nicht um eine Meldung, sondern um eine Deutung einer Meldung. Nur der

erste Satz des „Zeitungstextes“ ist eine Meldung. Wenn man die Deutung als Meldung behandelte, würde man die Richtigkeit nicht in der hier verlangten Weise diskutieren können. Die Differenz zwischen einem statistischem Datum und seiner Deutung wird hier einfach eingeebnet.
Die Form der Antwortfelder in (2) läßt zusätzlich unklar, was man tun soll: Soll man sich für eine Aussage entscheiden und eine entsprechende Argumentation führen, oder soll man Argumente dafür und dagegen versammeln? Es bleibt auch unklar, wofür man prämiert wird: für eine relative oder für eine absolute Betrachtung oder für beides zusammen?
Neben dem Nichternstnehmen der nur als real behaupteten Situation finden sich hier also schon äußerlich sichtbare textliche und inhaltliche Verwerfungen. Deshalb kommt auch diese Aufgabe als kontrastierender Fall nicht in Frage.

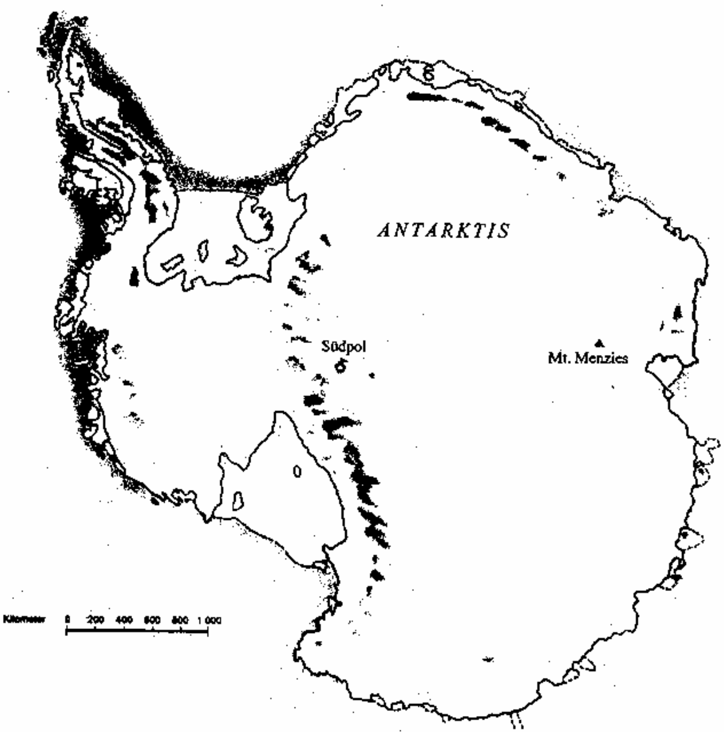
FLÄCHE EINES KONTINENTS (international)

Hier siehst du eine Karte der Antarktis.

1. Wie groß ist die Entfernung zwischen dem Südpol und dem Mt.Menzies? (Benutze den Maßstab der Karte für deine Schätzung.)
A Die Entfernung beträgt zwischen 1600 km und 1799 km.
B Die Entfernung beträgt zwischen 1800 km und 1999 km.
C Die Entfernung beträgt zwischen 2000 km und 2099 km.
D Sie kann nicht bestimmt werden.

2. Schätze die Fläche der Antarktis, indem du den Maßstab der Karte benutzt.

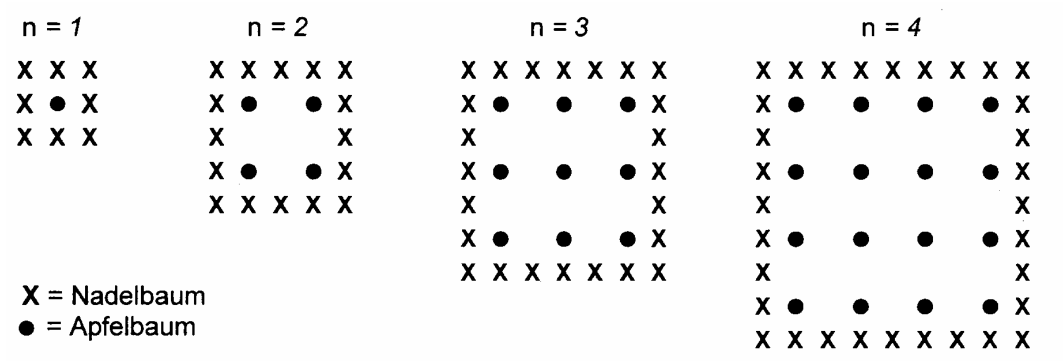
Schreibe deine Rechnung auf und erkläre, wie du zu deiner Schätzung gekommen bist. (Du kannst in der Karte zeichnen, wenn dir das bei deiner Schätzung hilft.)



Der erste Teil dieser Aufgabe ist nicht lösbar, wie man sofort bemerkt, wenn man sie zu lösen versucht. Man kann dort keine Entscheidung zwischen A und B treffen. Da unklar ist, welchen Einfluß diese Nichtlösbarkeit auf die Bearbeitung des zweiten Teils der Aufgabe hat, kann auch diese Aufgabe nicht als äußerlich irritationsfrei angenommen werden.
(Der erste Teil der Aufgabe wurde von der PISA-Gruppe aus der Wertung genommen, der zweite Teil wurde in die Wertung aufgenommen.)

ÄPFEL (international)

Ein Bauer pflanzt Apfelbäume an, die er in einem quadratischen Muster anordnet. Um diese Bäume vor dem Wind zu schützen, pflanzt er Nadelbäume um den Obstgarten herum.
Im folgenden Diagramm siehst du das Muster, nach dem Apfelbäume und Nadelbäume für eine beliebige Anzahl (n) von Apfelbaumreihen gepflanzt werden:



Äpfel 1:

Vervollständige die Tabelle:

n	Anzahl Apfelbäume	Anzahl Nadelbäume
1	**1**	**8**
2	**4**	
3		
4		
5		

Äpfel 2

Es gibt zwei Formeln, die man verwenden kann, um die Anzahl der Apfelbäume und die Anzahl der Nadelbäume für das oben beschriebene Muster zu berechnen:
Anzahl der Apfelbäume = n^2
Anzahl der Nadelbäume = $8n$
wobei n die Anzahl der Apfelbaumreihen bezeichnet.
Es gibt einen Wert für n, bei dem die Anzahl der Apfelbäume gleich groß ist wie die Anzahl der Nadelbäume. Bestimme diesen Wert und gib an, wie du ihn berechnet hast.

Äpfel 3

Angenommen, der Bauer möchte einen viel größeren Obstgarten mit vielen Reihen von Bäumen anlegen. Was wird schneller zunehmen, wenn der Bauer den Obstgarten vergrößert: die Anzahl der Apfelbäume oder die Anzahl der Nadelbäume? Erkläre, wie du zu deiner Antwort gekommen bist.

Auch in dieser Aufgabe wird die reale Situation in gewisser Weise nicht ernst genommen: Weder geht es um Äpfel, noch um einen algebraisierenden Bauern noch um Bäume. Allerdings wird der mathematische Inhalt so nah an der Baum-Konstruktion entlanggeführt, dass die Aufgabe sich als eingekleidete Aufgabe im oben beschriebenen Sinne erweist.

Als Testaufgabe ist sie allerdings mit Verwerfungen belastet: Eine wesentliche Lösungshilfe für Aufgabe 1 wird in Aufgabe 2 nachgereicht. Dadurch wird ein testfähiger Schüler mit einer erheblichen Zeitersparnis prämiert.
In Frage 2 wird behauptet, es gäbe einen Wert für n, bei dem die Anzahl der Apfelbäume gleich groß ist wie die Anzahl der Nadelbäume. Der Wert soll bestimmt werden und angegeben werden, wie man ihn berechnet hat. Es sind aber zwei Werte, nämlich 0 und 8. Für diese Aufgabe sind die Lösungskodierungen der PISA-Gruppe veröffentlicht. Aus ihnen geht hervor, dass ein Schüler, der n = 8 angibt, den Lösungspunkt erhält, auch wenn er keine Begründung bzw. Berechnung angibt - wenn er also die Aufgabenstellung nicht erfüllt. Ein Schüler, der nur n = 0 angibt, erhält hingegen keinen Punkt, selbst wenn er seine Antwort begründet und es bei diesem Wert beläßt, weil ja schließlich im Aufgabentext nur ein Wert gefordert ist.
In Frage 3 bleibt völlig unklar, auf welcher Argumentationsebene der Getestete arbeiten soll. Dass eine quadratische Funktion schneller wächst als eine lineare, kann man quasi empirisch zur Kenntnis nehmen. *Erklären* kann man es nur in einer Kommunikationssituation, in der dem Sprecher klar werden kann, ab welcher Stelle das Gegenüber das Gesagte für einleuchtend genug hält, um die Erklärung als abgeschlossen zu betrachten. Da in einem Test das Gegenüber unbekannt ist, bedarf es einer klaren Benennung der gewünschten Argumentationsebene. Testfähigkeit heißt hier zu wissen, dass bereits schlichte, vielleicht von einem selbst nur halb verstandene Erklärungen die Gnade der Kodierer finden. Insgesamt handelt es sich um eine Aufgabe, die im Unterricht eine hohe Produktivität (zum Begriff der produktiven Aufgabe vgl. Jahnke u.a. 2000) entfalten kann, als Testaufgabe aber nicht geeignet ist.

Als Aufgaben mit ernst zu nehmendem Modellierungsanspruch verbleiben aus dem Pool der veröffentlichten Aufgaben die nationale Aufgabe „Miete“ und die internationalen Aufgaben „Behälter“ und „Geschwindigkeit eines Rennwagens“. Mit den letzten beiden Aufgaben (siehe Anhang) treffen wir erstmals auf zwei äußerlich unverworfene Beispiele, die dem Ziel des PISA-Tests zu entsprechen scheinen „... zu prüfen, ob Schülerinnen und Schüler grundlegende mathematische Konzepte so verstanden haben, dass sie mit diesen Werkzeugen Problemsituationen aus unterschiedlichen Kontexten behandeln können.“ Es handelt sich in beiden Fällen um das Konzept der Funktion. Die Aufgaben sind produktive eingekleidete Aufgaben im obigen Sinne. Wir sind aber auf der Suche nach einer Aufgabe mit einem ernstzunehmenden realen Problem und einem sich daraus ergebenden Modellierungsanspruch. Das trifft eher auf die Aufgabe „Miete“ zu, die somit der beste Kandidat für die gesuchte Aufgabe ist.

MIETE (national)

In einer Großstadt kostete 1985 eine 70 m^2-Wohnung 1 000 DM Miete pro Monat. Seit 1985 stieg der Mietpreis alle 5 Jahre um 20 %.

Welche Monatsmiete musste dann 1995 für diese Wohnung gezahlt werden?
Schreibe auf, wie du rechnest.

Es geht nach wie vor darum, eine äußerlich verwerfungsfreie Aufgabe zu finden, die ein reales Problem bearbeitet. Die Aufgabe ist äußerlich verwerfungsfrei. Handelt es sich aber um ein reales Problem?

In einer Großstadt kostete 1985 eine 70 m^2-Wohnung 1 000 DM Miete pro Monat.

Auf der Suche nach Geschichten, in denen dieser Satz wohlgeformt fallen kann, zeigt sich, dass das Problem nicht real ist:

Vorstellbar ist ein ähnlicher Satz in einer Modellrechnung zur Verdeutlichung von Preissteigerungen, etwa in einer Zeitung oder in einer politischen Argumentation. Dann könnte ein Satz etwa so lauten: *In Großstädten kostete 1985 eine durchschnittliche 70 m^2-Wohnung 1 000 DM Miete pro Monat.* Varianten wären Spezifizierungen, z.B. *in deutschen Großstädten* oder in *Großstädten ab 100 000 Einwohner*. Auch das *durchschnittlich* könnte wegfallen, weil die durchschnittliche Betrachtungsweise durch *in Großstädten* bereits relativ deutlich gekennzeichnet ist.

Eine weitere Variante dieser Geschichte wäre die Angabe einer bestimmten Großstadt, für die diese Modellrechnung durchgeführt wird: *In Berlin kostete 1985 eine 70 m^2-Wohnung 1 000 DM Miete pro Monat.*

Vorstellbar ist ein ähnlicher Satz aber auch in einer Geschichte bezüglich einer bestimmten Wohnung: *1985 kostete die 70 m^2-Wohnung meiner Tante 1 000 DM Miete pro Monat.*

Diese Geschichten und Texte sind in dieser Aufgabe nicht verwirklicht. Hier ist von *einer Großstadt* die Rede. In der außerschulischen „Alltagswelt“ gibt es aber keine Situation, in der er wohlgeformt gesprochen werden könnte. Es handelt sich also um keinen Text aus dieser Welt. Als Aufgabentext handelt es sich um den Text einer eingekleideten Aufgabe. Er erscheint als Abkürzung eines Textes, der das Konstruierte bzw. Vormodellierte der Situation expliziter kennzeichnet, z.B. *Angenommen, in einer Großstadt kostete 1985 eine 70 m^2-Wohnung 1 000 DM Miete pro Monat.* Die von nicht benötigter Information gereinigte Variante wäre: *Angenommen, eine Wohnung kostete 1985 1 000 DM Miete.*

Die Annahme der Nichtrealität des Problems wird durch die Zahlenangabe 1 000 DM gestärkt. Sie soll im nächsten Satz aber nochmals überprüft werden:

Seit 1985 stieg der Mietpreis alle 5 Jahre um 20 %.

Auch dieser Satz ist in den beiden obigen Geschichten nicht wohlgeformt. Dort würde es wohlgeformt heißen: *Seitdem stieg der Mietpreis alle 5 Jahre um 20 %.* oder *Der Mietpreis stieg alle 5 Jahre um 20 %.*

Die wiederholte Angabe der Jahreszahl 1985 zerstört die Wohlgeformtheit – in der „realen“ Welt. In der Welt der eingekleideten Aufgaben ist diese Formulierung erklärbar, denn sie setzt den Anfangspunkt der verlangten Rechnung nochmals eindeutig fest.

Die Deutung, dass es sich nicht um ein reales Problem handelt, wird weiterhin dadurch gestützt, dass keine Betrachtung des Problems erfolgt, ob es sich um die Netto- oder um die Bruttomiete handelt. Diese Information ist in der „realen“ Welt wichtig, weil die Betrachtung der Entwicklung eines Mietpreises sich ja auf die wirkliche Belastung des Mieters bzw. auf die wirklichen Einkünfte des Vermieters bezieht.

Es handelt sich bei dieser Aufgabe also nicht um eine Aufgabe, die sich einem realen Problem stellt, sondern um eine eingekleidete Aufgabe. Eingekleidete Aufgaben wurden aber bereits ausführlich interpretiert, sowohl in verworfenen als auch in unverworfenen Varianten. Das Bedürfnis, im Korpus der veröffentlichten PISA-Aufgaben noch eine Aufgabe zu finden, die wirklich ein reales Problem behandelt, kann also nicht bedient werden. Es liegen ausschließlich technische und eingekleidete Aufgaben vor. Dies ist erstaunlich angesichts des Vorwurfs, den die PISA-Gruppe gegen eingekleidete Aufgaben erhebt und des Anspruchs, den sie selbst für ihre Aufgaben formuliert:

> „In Schulbüchern und im Schulunterricht findet man allerdings oft die so genannten eingekleideten Aufgaben, die den Mathematisierungsprozess praktisch ausblenden oder weitgehend trivialisieren, weil sie den Eindruck erwecken, genau eine Weise der Mathematisierung sei „richtig“. Dann wird also der für den Erwerb von Mathematical Literacy zentrale, ja charakteristische Vorgang des Mathematisierens abgeschnitten und die Aufgabe erscheint unmittelbar auf Modellebene.“ (PISA 2000, S. 145)

Wir haben gesehen, dass die PISA-Gruppe nur technische und eingekleidete Aufgaben veröffentlicht hat. Wir haben ebenfalls gesehen, dass auch in PISA-Aufgaben der „zentrale, ja charakteristische Vorgang des Mathematisierens abgeschnitten“ wird „und die Aufgabe unmittelbar auf Modellebene [erscheint]“. Dieses Abschneiden erfolgt aber in unterschiedlich starkem Maße. Besonders auffällig ist dieses Abschneiden dort, wo eine Mathematisierungsnotwendigkeit suggeriert wird, wie in der sich besonders realitätsnah gebenden Aufgabe „Bauernhöfe“. Hier ist Bestandteil der Testanforderung, herauszufinden, dass keine Mathematisierung nötig ist.

Wir haben aber ebenfalls gesehen, dass es nicht das Ausmaß an Mathematisierungsnotwendigkeit ist, die eine Aufgabe produktiv macht. Bereits der Vorwurf an die eingekleideten Aufgaben geht also am Kern vorbei. In den hier untersuchten Aufgaben gelangt man sogar zum Eindruck, dass der Versuch größerer Realitätsnähe Verwerfungen hervorruft, wenn lediglich das Mathematische mit Realität aufgefüllt werden soll. Die angedeuteten Alternativen zeigen aber, dass die Verwerfungen vermeidbar wären, wenn man sich dem Realen und dem Mathematischen in ihrer je eigenen Autonomie und Authentizität zuwendet und sie „vermittelt“.

5.10. Fazit: Der PISA-Test

Die kontrastierende Aufgabenauswahl hat uns durch den Korpus der veröffentlichten PISA-Aufgaben geführt. Die Aufgabe „Bauernhöfe“ hat dabei am Ausgangspunkt vielfältige Probleme mit kontrastivem Potential aufgeworfen. Im folgenden sollen die Erkenntnisse aus der PISA-Analyse zusammengefaßt werden:

Meßschärfe

PISA soll mathematische Leistungsfähigkeit messen. Dieser Anspruch wird durch das Mitmessen von Testfähigkeit (siehe unten) und anderer latenter Elemente und durch mangelnde Meßschärfe bezüglich des mathematischen Inhalts gebrochen.
Lediglich bei Aufgaben mit geringem inhaltlichem Anspruch (Rechteck, Bauernhöfe 1) ist relativ klar eindeutig zu benennen, was die Aufgabe mißt. Sobald der Anspruch höher wird, ergibt sich eine größere Vielfalt möglicher Lösungswege. Man kann bei den PISA-Aufgaben nicht nachvollziehen, welche der Lösungswege oder welche Mischwege beschritten wurden. Da die verschiedenen Lösungswege unterschiedliche Fähigkeiten ansprechen, bleibt unklar, welche dieser Fähigkeiten gemessen werden.

Zuordnung zu den Kategorien des Testkatalogs, Kompetenzstufenmodell

Bezüglich der Zuordnung zu den Kategorien des Testkatalogs zeigte sich, dass die Vielfalt der Lösungswege dazu führen kann, dass keine Zuordnung vorgenommen werden kann. Dies gilt ebenfalls für das Kompetenzstufenmodell. Um die Stufen dieses Modells inhaltlich beschreiben zu können, muß man beschreiben können, welche Kompetenzen die Aufgaben auf den einzelnen Stufen messen. Wenn diese Aufgaben aber vielerlei messen, dann kann man nicht Teile dieses Vielerlei herausnehmen und zum Stufeninhalt erklären (siehe Meyerhöfer 2004). Das Zuordnungsproblem wird verschärft, weil es für in den Interpretationen herausgearbeitete Phänomene wie Intuition oder Testfähigkeit keinen Platz in den Kategoriensystemen der PISA-Gruppe gibt. Diese Kategoriensysteme beziehen sich ausschließlich auf jene Aufgabencharakteristika, die man sich wünscht, nicht auf jene, die da sind.

Testfähigkeit

Auch bei PISA wird eine breite Palette von Komponenten von Testfähigkeit mitgemessen. Es gelang, mit der Aufgabe „Feriengeld“ eine Aufgabe zu finden, die nicht nur frei von sprachlichen Verwerfungen ist, sondern bei der auch ansonsten textlich keine Elemente von Testfähigkeit auftreten. Erst die Multiple-Choice-Konstruktion ermöglicht hier, auch zufällig zum richtigen Ergebnis zu gelangen.
Das Problem der Testfähigkeit wird im Abschlußkapitel eingehender diskutiert.

Modellierungsanforderung und Realitätsorientierung

Es wurde zunächst herausgearbeitet, dass bei der Aufgabe „Bauernhöfe“, die die PISA-Gruppe als exemplarisch für den Realistic Mathematics Education-Ansatz bezeichnet, keine Modellierungsanforderung besteht. Das angebliche reale Problem wird nicht ernst genommen und spielt für die Lösung der Aufgabe keine Rolle.

Es ist im weiteren nicht gelungen, eine Aufgabe zu finden, bei der ein wirklich reales Problem zu bearbeiten ist. In den Begrifflichkeiten der mathematikdidaktischen Debatte um Modellierungsprobleme liegt keine einzige Modellierungsaufgabe vor. Die PISA-Gruppe hat sich allerdings gegen dieses Problem durch weite Begrifflichkeiten immunisiert: So will sie untersuchen „ob Schülerinnen und Schüler grundlegende mathematische Konzepte so verstanden haben, dass sie mit diesen Werkzeugen Problemsituationen aus unterschiedlichen Kontexten behandeln können“. Sowohl der Problem- als auch der Kontextbegriff sind weit genug, um Beliebiges zu erfassen. Die PISA-Gruppe dementiert außerdem selbst, dass Realistic Mathematics Education „unbedingt „realistisch“ im Sinne von handlungs- oder alltagsbezogen sein“ muß. Mit der Abwandlung des Modellierungskreislaufs wird die Selbstimmunisierung gegen irgendeinen spezifischen Anspruch vollendet, weil man fast jede Aufgabe als Modellierungsaufgabe ansieht.

Auch wenn kein reales Problem zu lösen ist, dient der Modellierungskreislauf in den Darstellungen der PISA-Gruppe dazu, die Mathematisierung von nur teilweise mathematisierten Problemstellungen zu beschreiben. Die PISA-Gruppe arbeitet in ihren veröffentlichten Arbeiten aber nicht mit diesem Instrument, es scheint nur theoretisierendes Beiwerk zu sein. Dem entsprechend gibt es auch in den Aufgaben keine Elemente, die Teile eines Modellierungsprozesses sinnvoll ersetzen oder erkennen ließen, dass eine Aufgabe entwickelt wurde, die ein Modellierungsproblem zum Ausgang hatte und davon ausgehend im Anspruch reduziert wurde, um als Testaufgabe zu funktionieren. Die Interpretationen zeigen hingegen, dass in Aufgaben wie „Bauernhöfe“ oder „Fahrradunfälle“ die Modellierungsanforderungen zerstört werden und dass Verwerfungen zwischen Realem und Mathematischem produziert werden. In der Aufgabe „Sparen“ ist dies dadurch zugespitzt, dass die Spezifik der Testsituation mißachtet wird.

Realistic Mathematics Education und Mathematical Literacy

Die PISA-Gruppe stellt ihren Test als Benchmarking-Instrument vor (PISA 2000, S. 19). Am Testwert entscheidet sich also die Qualität schulischen Handelns. Für den Mathematikunterricht wird dieses Benchmarking-Ziel auf das Verstehen und Anwenden mathematischer Konzepte reduziert: „Ziel des PISA-Tests ist es ... zu prüfen, ob Schülerinnen und Schüler grundlegende mathematische Konzepte so verstanden haben, dass sie mit diesen Werkzeugen Problemsituationen aus unterschiedlichen Kontexten behandeln können.“ (PISA 2000, S. 142) Die Konkretisierung erfolgt in Aufgaben, die vorrangig dem niederländischen Konzept der Realistic Mathematics Education folgen: „Diese Aufgaben gehen häufig von einer „realistischen“ Problemstellung aus, an der, entlang systematisch aufgebauter Teilprobleme, mathematische Begriffe entwickelt werden.“ (ebd., S. 143)

Die internationalen Aufgaben „Äpfel“ und „Behälter“ bergen das Potential für eine solche Begriffsentwicklung. Insofern sind sie produktive Aufgaben für den Unterricht. In einem Test werden aber keine Begriffe entwickelt, sondern es soll die Begriffskenntnis und das Begriffsverständnis gemessen werden. Das ist mit diesen Aufgaben aber nicht möglich, weil man die Aufgaben auch ohne diese Begriffe lösen kann. Die Stärke dieser Aufgaben, viele Anknüpfungspunkte für die Begriffsentwicklung zu liefern, wird zur Schwäche beim Messen.
Die anderen internationalen Aufgaben messen zwar mathematische Fähigkeiten, aber sie messen nicht das Verständnis „grundlegender mathematischer Konzepte“. Bei allen Aufgaben (außer „Kontinent“) bleibt unklar, worin die besondere Stärke des „realistic“-Gedankens bestehen soll: Das Verstehen wird beim „Äpfel“-Problem nicht dadurch befördert, dass es gerade Äpfel sind. Eine produktive Umdeutung der „Bauernhöfe“-Aufgabe könnte ebensogut auf den Bezug zum Bauernhaus verzichten. Die „Realitätsorientierung“ von „Rennwagen“ oder „Behälter“ ergibt sich aus der Sache, nicht aus dem Versuch des Aufsetzens eines realen Kontextes.

Zur Deutung der Testresultate durch die PISA-Gruppe

Die PISA-Gruppe hat die Veröffentlichung der Testresultate mit umfangreichen Deutungen begleitet (siehe z.B. PISA 2000, S. 161-187, Deutsches PISA-Konsortium 2002, S. 101-127).
Bereits in Kapitel 3 wurde diskutiert, mit welchen Unsicherheiten die Deutung von Testresultaten prinzipiell belastet ist. Im PISA-Kapitel wurde speziell der PISA-Test in seinen Meßeigenschaften untersucht. Bereits vor Beginn der Interpretationen wurde deutlich, dass die PISA-Gruppe kein Instrumentarium besitzt, um die latente Ebene des im Test Gemessenen zu erfassen. Die Interpretationen der Aufgaben „Bauernhöfe“ und „Dreieck“ zeigten bereits deutlich, zu welch umfangreichen Verwerfungen und Meßproblemen das führt. Da dieses Problem ein systematisches ist, bleibt es zufällig, ob es in einer Aufgabe solche Verwerfungen und Meßprobleme gibt oder nicht. Als Meßinstrument ist dieser Test daher nicht geeignet. Das bedeutet aber nicht, dass einzelne Aufgaben nicht auch für einen mathematischen Leistungstest oder im Unterricht produktiv verwendbar wären.

Die aufgezeigte Struktur der manifesten Orientierung auf Fachsprachlichkeit bei gleichzeitiger latenter Zerstörung des Mathematischen berührt die Eignung einer Aufgabe im Rahmen eines mathematischen Leistungstests hingegen nicht: Hier wird „lediglich“ ein spezieller professioneller Habitus rekonstruiert (siehe nächsten Abschnitt). Man kann sich entscheiden, ob man die Bedienung dieses Habitus mitmessen möchte oder nicht.
Selbst wenn man jede einzelne PISA-Aufgabe für einen brauchbaren Bestandteil eines mathematischen Leistungstests hält, sind die PISA-Resultate *inhaltlich* nicht zu deuten: Man kann die Lösungshäufigkeiten der Aufgaben und die ordinale Anordnung beliebiger Gruppen auf der Punkteskala registrieren. Eine (über Beliebiges hinausgehende) inhaltliche Deutung läßt dieser Datenvergleich aber nicht zu: Die Lösungshäufigkeit einer Aufgabe erzählt nicht mehr als die Häufigkeit des „richtigen“ Lösens. Immerhin läßt sich aber wenigstens bei verwerfungsfreien Aufgaben mit eindeutigem Lösungsweg aus der Lösungshäufigkeit eine inhaltliche Aussage ableiten. Gruppenvergleiche lassen sich hingegen

überhaupt nicht inhaltlich deuten, weil diese inhaltliche Deutung über ein theoretisches Modell wie das Kompetenzstufenmodell laufen muß. Die inhaltliche Füllung so eines Modells ist aber mit den PISA-Aufgaben wegen der auftretenden Lösungswegvielfalt nicht möglich. Der Erkenntnisfortschritt aus dem mathematischen Leistungstest von PISA beschränkt sich also auf die Möglichkeit der inhaltlichen Deutung der Lösungshäufigkeit weniger Aufgaben.

5.11. Diskussion des aus den PISA-Aufgaben rekonstruierten mathematikdidaktischen Habitus

Manifeste Orientierung auf Fachsprachlichkeit und latente Zerstörung des Mathematischen

Das Auftreten dieser Struktur in PISA-Aufgaben überrascht doppelt: Es überrascht zunächst, weil PISA von Fachdidaktikern entworfen wurde. – Wenn man diese Struktur in einer Unterrichtsstunde findet, so kann man sie mit mangelnden fachlichen Fähigkeiten und mit zu geringer Reflexion über das Wesen des Faches erklären. Für erfahrene Fachdidaktiker ist diese Erklärung aber wenig einleuchtend - es handelt sich um ein habituelles und nicht um ein Kompetenzproblem.
Das Auftreten dieser Struktur überrascht weiterhin, weil sie nicht nur bei Aufgaben auftritt, die eine nicht ernst gemeinte bzw. ernst genommene Realitätsorientierung haben. Bei diesen Aufgaben – z.B. „Bauerhöfe“, „Sparen“ oder „Fahrradunfälle“ – wird dem eigentlich Gemeinten (also dem mathematischen Inhalt) eine scheinreale Hülle hinzugefügt. (Das schließt nicht aus, dass die Idee für die Aufgabe wirklich aus einem realen Kontext stammt.) Die damit verbundenen Verwerfungen kann man sich gut als Quelle der Zerstörung des Mathematischen vorstellen. Im Gegensatz dazu wird aber bei der Aufgabe „Dreieck“ das Mathematische sogar zerstört, ohne dass ein realer Kontext auf das Problem aufgesetzt wurde. Hier wird ein rein mathematisches Problem betrachtet, und trotzdem wird das Mathematische zerstört.
Auffallend ist, dass die Zerstörung des Mathematischen immer einhergeht mit manifester Orientierung auf Fachsprachlichkeit. Dies reproduziert ein Muster, das sich auch in Untersuchungen von Unterrichtsgesprächen findet: Es wird Wert auf korrekte Verwendung von Fachsprache gelegt, auch wenn dadurch der mathematische Gedankengang zerstört wird oder wenn dieser Gedankengang auch ohne Fachsprachlichkeit verständlich ist. In den Aufgaben ist es ja ebenfalls so, dass das Fachsprachliche oftmals ohne jede Notwendigkeit für den mathematischen Gedanken verwendet wird. Die hinter diesem Phänomen stehende Struktur bedarf näherer Untersuchung, um ihre Stellung in der Gesamtbetrachtung von Mathematikunterricht zu verstehen. Dies ist umso wichtiger, als es in einem normativen Sinne unstrittig sein dürfte, dass Fachsprache inhaltsorientiert und nicht als sprachlicher Schleier und somit als Distinktionselement gebraucht werden sollte. In der Aufgabe „Miete“ zeigt sich das Gegenmodell: Hier wird das Mathematische völlig ernst genommen, das Fachsprachliche hingegen ist auf das unbedingt Notwendige reduziert.

Mißlingen der „Vermittlung“ von Realem und Mathematischem

Diese Struktur ist nicht nur ein Meßproblem, als das sie oben bereits beschrieben wurde. Sie ist auch ein habituelles Problem. Es deutete sich schon in den theoretischen Grundlagen von PISA an:

- Als Charakteristikum für den *Realistic Mathematics Education*-Ansatz wird angegeben, „dass eine außermathematische Situation (Foto) sogleich durch eine schematische Zeichnung ergänzt wird, sodass beides, außer- und innermathematische Zusammenhänge, gleichzeitig angezeigt sind.“ (PISA 2000, S.151) Dieses gleichzeitige Anzeigen deutet auf ein Nebeneinander, nicht aber auf ein Vermitteltes. Die in den Aufgaben gefundenen Verwerfungen zeigen dieses nicht vermittelte Nebeneinander. Vom Schüler wird hingegen verlangt, dass er „in seinem Kopf“ ein Miteinander daraus formt – jenes Miteinander, das vom „Lehrenden“ hier nicht geleistet wird.

- Die PISA-Gruppe konzipiert Mathematical Literacy als Zuordnung eines Fähigkeitskonzept zum Freudenthalschen didaktischen Konzept der „Realistic Mathematics Education“. (siehe ebenda, S. 141-143) Freudenthals didaktischen Ansatz beschreibt sie mit folgenden Stichworten: Mathematik als System begrifflicher Werkzeuge, mit dem sich Schülerinnen und Schüler Phänomene ihrer natürlichen, technischen, geistigen und sozialen Umwelt erschließen können; Beziehungshaltigkeit; Orientierung an der Welt, aber kein instrumentelles Verständnis, sondern Ausbildung mentaler Modelle für mathematische Begriffe. (ebd., S. 142 f.) Als Ziel des PISA-Tests verbleibt allerdings nur noch „zu prüfen, ob Schülerinnen und Schüler grundlegende mathematische Konzepte so verstanden haben, dass sie mit diesen Werkzeugen Problemsituationen aus unterschiedlichen Kontexten behandeln können.“ (ebd., S.143) *Habituell* stehen sich hier intellektuelle Reichhaltigkeit und machbarkeitsorientierte Enge gegenüber. Dieser Unterschied ist einerseits dadurch verursacht, dass ein Test einen Zwang verursacht, Reichhaltigkeit abzuschneiden. Die „Konkretisierung“ des Freudenthalschen Ansatzes in der Nebeneinandersetzung von Realem und Mathematischem zeigt aber, dass der habituelle Unterschied bereits vor dem Problem der Testerstellung auftaucht: Die intellektuelle Abreicherung beginnt dort, wo das Problem kleingehackt wird, um es in das mathematikunterrichtlich Gewollte zu pressen und wo das Problem in seinem realen Charkter nicht ernst genommen wird. Das Interessante und Überraschende der Aufgabeninterpretationen ist, dass mit der Zerstörung des Realen auch das Mathematische zerstört wird. Erklärbar ist dies dadurch, dass *beides* in seiner jeweiligen Autonomie und Authentizität ernst genommen werden muß, damit eine Vermittlung zustande kommt.

Es könnte sein, dass das hier beschriebene Problem der Verwerfung von Realem und Mathematischem bereits bei Freudenthal mitschwingt, weil die Dominanz des Mathematischen zu Mißverständnissen führen kann. Entscheidend für den habituellen Bruch scheinen mir aber drei Aspekte zu sein:

1. Die Reduzierung der intellektuellen Reichhaltigkeit des Freudenthalschen Ansatzes bringt es mit sich, dass die Autonomie und Authentizität des Realen und des Mathematischen nicht mehr im Blick sind: Die intellektuelle Abreicherung von reichhaltigen Konzepten bringt es immer mit sich, dass Teile der beteiligten Gegenstandsbereiche aus dem Blick geraten.

2. Eng damit verbunden ist ein Ideologisierungsprozeß. Wenn die Verbindung des Mathematischen mit dem Realen zur Ideologie verkommt, dann geht es nur noch um die Bedienung der Ideologie. Die Authentizität der Verbindung sowohl für das Reale als auch für das Mathematische spielen dann nur noch eine untergeordnete Rolle. Theoretischer Ausfluß solcher Ideologie sind Konzepte wie die PISA-Ebene 3 „Situations and Contexts".

3. Ein dritter Aspekt des habituellen Bruchs ist die Technokratisierung. Tests sind Höhepunkte von technokratischer Annäherung an die Welt. Auch vor dem Konzept der Mathematical Literacy gab es technokratische Versuche, mathematikdidaktische Ausdeutungen zu finden für den richtigen und wichtigen Gedanken, dass sich Inhalte uns am besten vermitteln, wenn wir sie in unserer Realität verorten können. Die positive Rezeption von PISA in weiten Teilen der Mathematikdidaktik zeigt, dass dieses technokratische Spitzenwerk ein verbreitetes Bedürfnis nach einer bestimmten Form der Bearbeitung mathematikdidaktischer Probleme befriedigt. Es wird nicht überraschen, wenn sich Verkürzungen der Art, wie sie in dieser Arbeit aufgezeigt wurden, auch in anderen technokratisch orientierten Arbeiten finden.

Die Illusion der Schülernähe als Verblendung

Das internationale PISA-Konzept vermischt Realitätsnähe und Schülernähe auf eigentümliche Weise:

„Students` mathematical insight and understanding need to be assessed in a variety of situations, partly to minimize the chance of students finding that tasks are not culturally relevant.
One can think of a situation as being at a certain distance from the student. The closest is private life (daily life), next is school life, work and sports, followed by the local community and society as encountered in daily life, and furthest away are scientific contexts. In this way one can define a more or less continuous scale of situations." (OECD, S. 50)
Diese Vermischung spiegelt sich in der Aufgabe „Bauerhöfe" am eindrucksvollsten, bei „Sparen" zeigt sie sich abgeschwächt. Die PISA-Gruppe scheint auch keine getrennten Realitäts- und Schülernähedimensionen für ihre Aufgabenkatalogerstellung zu nutzen, sondern nur die sechsstufige Schülernähedimension. Realität wird damit „echter", wenn sie näher am Schüler ist. Dies erscheint der PISA-Gruppe vielleicht praktikabler, weil Schülernähe leichter zu fassen scheint als Realitätsnähe. Schließlich sind selbst „scientific contexts" real. Allerdings muß man dann die der Population gemeinsame Schülernähe finden. Dies bietet nicht nur Ansatzpunkte für beliebige Unterstellungen, z.B. jene, dass „work and sports" näher am Schüler ist als „scientific contexts" – oder dass Bauernhöfe dem Schüler näher sind als Stadthäuser. Dieses Herangehen bildet auch eine gute Plattform für romantisierende Projektionen: Es ist kein „Versehen", dass über einer Aufgabe mit einem Hausdach die Überschrift „Bauernhöfe" steht.

Die Konstruktion von Schülernähe führt zu zwei Verwerfungen:

1. Die oben beschriebenen Verwerfungen des Realen entstehen *auch* wegen der Konstruktion der Illusion von Schülernähe. Sowohl das Bauernhöfekonstrukt als auch die angebliche Schülerinnenskizze

als auch das seltsame mütterliche Zinseszinskonstrukt konstruieren die Illusion von Schülernähe, verursachen aber gleichzeitig Verwerfungen des Realen. Das Gelingen von „Feriengeld“ im Vergleich zu „Sparen“ beruht nicht auf einer größeren oder geringeren Nähe zum Schüler. Es beruht darauf, dass Schülernähe gar nicht erst illusioniert wird.

2. Kein Test ist schülernah. Keine Leistungskontrolle ist schülernah. Die Konstruktion der Illusion von Schülernähe verdeckt die für Schule konstitutive Grundstruktur der Vergabe von Zukunftschancen, welche sich in der Grundstruktur des Testens wiederspiegelt (siehe Kapitel 3).
Die Ursachen dieser „Verdeckung“ lassen sich in zwei Polen beschreiben. Der eine Pol ist Zynismus: Der strukturelle Zynismus ist hier unübersehbar. - Der Schüler ist in einer Leistungssituation, das heißt dass über ihn geurteilt wird, wobei er verschiedenen Graden der Vergabe von Zukunftschancen ausgesetzt ist. Gleichzeitig wird ihm Nähe suggeriert. Der zweite Pol ist pädagogische Ideologie: Die zynische Grundstruktur wird durch eine ideologische Konstruktion zugedeckt. Dabei ist egal, ob man die zynische Grundstruktur nicht erkennt oder ob man glaubt, sie durch gesteigerte Zuwendung bzw. Nähe zum Schüler ausgleichen zu können.

Bezeichnend ist, dass eine *wirkliche Zuwendung zum Schüler unterbleibt*. Sie würde darin bestehen, dass man dem Schüler deutlich sagt, was man von ihm erwartet, welche Leistung er also zu erbringen hat, um seine Zukunftschancen zu maximieren. Sie würde darin bestehen, Fachsprache inhaltsorientiert und nicht als sprachlichen Schleier und somit als Distinktionselement zu gebrauchen. Sie würde darin bestehen, Modellierungshilfen zu geben statt das Modell mit dem Realen zu verwerfen.
Wenn Ideologie die Strukturiertheit einer gesellschaftlichen Praxis verschleiert, so bezeichnet man das als Verblendung. Die beschriebene Verblendung wirkt sowohl in Richtung der Schüler als auch in Richtung der Lehrer, Didaktiker und Sonstiger. Die Schüler thematisieren den Zynismus dieser Verwerfung nur selten, insbesondere in konflikthaltigen Situationen. Die Heuchelei von Schülernähe ist ihnen so vertraut, dass eine Thematisierung erhebliche (auch intellektuelle) Distanz vom alltäglichen Unterrichtsgeschehen erfordert. Aber selbst Erwachsene – die diese Distanz erreicht haben – sind oftmals nicht mehr in der Lage, die Heuchelei als solche zu benennen. Angesichts der PISA-Aufgaben haben viele Zeitungsleser sicherlich nur das seltsame Gefühl aus ihrer Schulzeit wiedererkannt. Am stärksten wirkt die Verblendung aber gewiß bei den beruflich pädagogisch Tätigen. Für sie gehört die pädagogische Norm der Schülernähe zum Standardrepertoire professioneller Selbstdefinition. Man könnte einfach hinzusetzen: Vermeide die Illusion der Schülernähe. Damit wäre das Problem der Verblendung zwar nicht erledigt, aber immerhin thematisiert. Es geht dabei nicht um eine Gegenüberstellung von Zuwendung zum Schüler und Abwendung vom Schüler. Gerade im Mathematikunterricht gibt es auch eine unselige Tradition der Ignoranz gegenüber dem Schüler. Diese Ignoranz ist manchmal gepaart mit einer geradezu vorbildlichen Zuwendung zur Sache. Vielleicht läßt sich eine ideologiefreie Zuwendung zum Schüler und zur Sache mit dem Begriff „Zuwendung zum Lernprozeß“ fassen.

Kalkülorientierung statt mathematischer Bildung

Die Kalkülorientierung des deutschen Mathematikunterrichts ist vielbeklagtes Thema der deutschen Mathematikdidaktik. Die nationale PISA-Gruppe gibt sie als einen Grund für eine nationale Zusatzuntersuchung an. Zugleich reproduziert sie dieses Defizit:
In ihren theoretischen Grundlegungen bezieht sich die nationale PISA-Gruppe auf das Konzept mathematischer Bildung von Winter (1995). Sie verweist hier insbesondere auf den Charakter von Mathematik als „eine deduktiv geordnete Welt eigener Art", gelöst von den phänomenologischen Verankerungen. Auffallend ist nun, dass dieser besondere Charakter und damit das Verhältnis zu Welten anderer Art in keiner einzigen veröffentlichten Aufgabe thematisiert ist. Deshalb ist hier auch nicht diskutierbar, welchen Charakters diese Welt für die PISA-Gruppe ist. (Den bisher diskutierten Punkten ist aber zu entnehmen, dass man sie zwar nicht ernst nimmt, sie aber mit Fachsprachlichkeit umstellt.)
Die nationale PISA-Gruppe veröffentlicht nur Aufgaben, die technisch orientiert sind. Zwar bewegen sich auch diese Aufgaben in einer Welt eigener Art, aber die ist schulmathematisch und nicht mathematisch, und sie ist nur partiell deduktiv geordnet. Vor allem reflektieren diese Aufgaben die Eigenheiten ihres Charakters nicht.
Als Begründung für diese technische Orientierung nennt die PISA-Gruppe den beklagten Charakter des deutschen Mathematikunterrichts, welcher sich im Test wiederfinden sollte. Dies ist nun ein typisches Reproduktionsmuster. Tragisch ist dieses Muster, weil PISA gleichzeitig beansprucht, der Transformator zu sein, und genau dies nicht sein kann: Mit dem nationalen Test werden die Schulen genau auf jene Fähigkeiten und Fertigkeiten hin getestet, die sie gar nicht vermitteln sollen, auf die sie sich aber entgegen den didaktischen und Lehrplanintentionen konzentrieren. Wir haben in den Interpretationen gesehen, dass mathematisch und nicht technisch denkende und handelnde Schüler benachteiligt sind. Aber diese testpragmatische Widersprüchlichkeit ist für sich genommen eine Kleinigkeit.
Problematisch wird der Reproduktionscharakter von PISA durch den Benchmarkcharakter (PISA 2000, S.19) des Tests: Wer bei PISA gut abschneidet, hat nicht nur im großen Maßstab das bessere Schulsystem (die Nichthaltbarkeit dieser Aussage wird meist noch erkannt), sondern macht den besseren Unterricht (die Nichthaltbarkeit dieser Aussage ist in Expertenkreisen unbestritten, weshalb man bei der Beurteilung von Schulen auf Vorher-Nachher-Tests setzt) und hat auf jeden Fall die besseren Fähigkeiten. Diese letzte Aussage ist unvermeidbar, denn zur Messung dieser Fähigkeiten gibt es ja den Test. Der Reproduktionscharakter von PISA bringt nun aber mit sich, dass auf allen diesen Ebenen jene Leistungen bewertet werden, auf die es gar nicht ankommt. Dieses Problem bleibt unberührt davon, ob vielleicht die eigentlich gewollte mathematische Bildung sich *auch*, vermittelt oder unvermittelt, in technischen Fähigkeiten zeigt. Wollte man dies auf einen Test bezogen diskutieren, so müßte man in einen Operationalisierungsvorgang treten. Das ist aber nicht geschehen.
Der Reproduktionscharakter von PISA steht in eigentümlichem Widerspruch mit dem durch die Benchmark-Charakterisierung erhobenen Anspruch, die Richtung für Veränderung zeigen zu wollen (siehe auch Neubrand u.a. 2001, S.46).

Abkehr von der Sache

Der bis hierher beschriebene, aus den PISA-Aufgaben rekonstruierte Habitus läßt sich als „Abkehr von der Sache“ zusammenfassen. Auf der allgemeinsten Ebene findet sich diese Abkehr in der Überbetonung des Technischen bzw. der Kalkülorientierung gegenüber mathematischer Bildung. Schließlich ist Kalkülausübung kein Zweck mathematischen Tuns. Auf einer weniger allgemeinen Ebene ist auch die Zerstörung des Realen wie des Mathematischen eine Abkehr sowohl vom Realen als auch vom Mathematischen – obwohl der Sinn von modellierendem Arbeiten im Mathematikunterricht darin besteht, sich dem Realen und dem Mathematischen zuzuwenden, um in beidem zu neuen Erkenntnissen zu gelangen – sich also der Sache zuzukehren. Auf der Ebene der Aufgabe ist es ebenfalls eine Abkehr von der Sache, wenn die Ernsthaftigkeit des Gegenstandes der Aufgabe dementiert wird.

Auf allen drei Ebenen sind die in Mathematikunterricht Involvierten die Abkehr von der Sache völlig gewohnt. Dies bietet erste Hinweise darauf, dass es sich hier nicht nur um einen mathematikdidaktischen, sondern einen auch in die Schule eingedrungenen, mithin pädagogischen Habitus handelt, der auch die gesellschaftliche Wahrnehmung von Mathematik bestimmt. Auf der Ebene der Kalkülorientierung ist der Habitus so tief verwurzelt, dass regelmäßig der mathematische Studienbeginn der Lehramtsstudenten von enormen Brüchen des Fachhabitus begleitet ist. Der Widerstand gegen die Inhalierung eines mathematischen Habitus ergibt sich nicht nur aus dessen eigenen Verwerfungen (z.B. Überbetonung der Begabung, mangelnde Reflexion über eigenes Tun), sondern vorrangig aus der mangelnden Passung zwischen eigener Kalkülorientierung und mathematischer Deduktionsorientierung und Theoriebildung. Die mit der Zerstörung des Mathematischen einhergehende Orientierung auf Fachsprachlichkeit mag in diesem Spannungsverhältnis ihren Nährboden haben.

Auf der Ebene der Zerstörung des Realen wie des Mathematischen ist die Gewohnheit ebenso offensichtlich. Das ist eng verzahnt mit der Dementierung der Ernsthaftigkeit des Gegenstandes. Niemanden wundert es, in einer Mathematikaufgabe auf Behauptungen über das Reale zu treffen, die nicht stimmen. Gelegentlich wird sogar versucht, Sperren gegen das unkontrollierte Eindringen der Wirklichkeit aufzubauen – insbesondere wenn zu befürchten ist, dass die Schüler sich die Behauptung der Realität einer Aufgabenstellung zu eigen machen und nicht gewollte Modellierungen oder Lösungswege vorschlagen (Beispiel in Meyerhöfer 2002). Nicht zuletzt leben die Versuche von Herget (siehe z.B. seine Rubrik in „mathematik lehren“) davon, im Kontrast zum hier beschriebenen Habitus das Reale in seiner Authentizität in den Unterricht zu holen. Die Gefahr des Abgleitens ins Schulmeisterliche ist dabei dauerhaft präsent.

Die objektiv-hermeneutische Feinanalyse ruft das eigene Verhaftetsein in den Habitus des Nichternstnehmens ins Bewußtsein: Man weiß beim Lesen der Bauernhöfe-Aufgabe, dass keine Schülerin die Zeichnung erstellt hat. Man würde es gar nicht als bemerkenswert erachten. Erst das Nachdenken darüber, warum man diese offensichtliche Unwahrheit als normal ansieht, eröffnet den Blick darauf, was hier geschieht. Und es eröffnet den Blick darauf, in welchen Habitus der Schüler sich einarbeiten muß, wenn er im Unterricht erfolgreich sein will.

Es wird unseren Blick auf Mathematikunterricht erweitern, wenn der hier rekonstruierte mathematikdidaktische Habitus - insbesondere insofern er auch ein Mathematiklehrerhabitus ist - näher beschrieben und in seiner Entwicklung untersucht wird. Der Scheinwerfer der Betrachtung muß sich dabei auf die Art und Weise richten, in der Mathematik und im Gegensatz dazu Schulmathematik dem Lehrerstudenten bzw. dem Lehrer entgegentreten. Er muß sich auf den Charakter des Realen richten, des Realen „an und für sich“ und des gesellschaftlich und schulisch vermittelten Realen. Er muß sich aber auch auf Grunddispositionen des Lehrerseins richten: Schule ist und bewegt sich in einer Realität außerhalb und innerhalb des hier gemeinten natürlichen und kulturellen Realen. Mathematikunterricht bewegt sich innerhalb und außerhalb von Mathematik, die sich wiederum in spezifischer Weise zur Realität positioniert.

Ein weiterer fundamentaler Aspekt einer Habitusuntersuchung wird der entfremdete Charakter des Lehrberufs sein. – In der Unterminierung der Authentizität und Autonomie sowohl des Realen als auch des Mathematischen steckt ja ein Problem der Entfremdung von beidem; hier scheint die Reproduktion der beruflichen Entfremdung des Lehrers auf, von der sich auch die Didaktik nicht emanzipiert hat.

SCHLUSS

Testfähigkeit

In den Interpretationen wurde bei den meisten Aufgaben festgestellt, dass sie „Testfähigkeit mitmessen“. Der Begriff ist zunächst selbsterklärend und konnte deshalb problemlos verwendet werden. In diesem Abschnitt soll er aber präziser gefaßt werden.
In den Interpretationen wurde der Begriff Testfähigkeit verwendet, wenn inhaltliche Dimensionen mitgemessen wurden, die man nicht unter den Begriff „mathematische Leistungsfähigkeit“ fassen würde. Insbesondere wenn es sich dabei um Dimensionen handelt, die nur deshalb auftauchen, weil es sich um einen Test handelt, scheint die Bezeichnung „Testfähigkeit“ sinnvoll zu sein.

Bedeutung von Testfähigkeit innerhalb der Diskussion um Tests

1. Der hier verwendete Begriff der Testfähigkeit bezieht sich nur auf standardisierte (mathematische) Leistungstests. Diese Tests beanspruchen, die Relativität von Leistungsbewertung in der Schule zu heilen, also ein weniger relatives oder nicht relatives Maß für (mathematische) Leistungsfähigkeit zu erstellen. Nur aus diesem Anspruch heraus ergibt sich überhaupt die Notwendigkeit, das Mitmessen von Testfähigkeit zu diskutieren: Wenn Fähigkeiten mitgemessen werden, die nicht die zu messenden mathematischen Fähigkeiten sind, dann sind diese Fähigkeiten zu benennen, und sie sind daraufhin zu untersuchen, ob sie erwünscht sind. Man kann sich z.B. darauf festlegen, dass das Mitmessen verbaler Fähigkeiten gerade erwünscht ist. Ich gehe bei den folgenden Betrachtungen davon aus, dass TIMSS und PISA ausschließlich mathematische Fähigkeiten messen sollen. Die zusätzlich mitgemessenen Fähigkeiten können – wenn man sie erkannt hat – zwar dem Meßkonstrukt zugeschlagen werden. Man verwickelt sich dann allerdings in Probleme der Fairneß und der Zielstellung des Testens: i) Es ist in sich problematisch, dass Testkompetenz als mathematische Kompetenz erscheint. ii) Je mehr nichtmathematische Fähigkeiten man bereit ist mitzumessen, umso breiter und präziser muß diskutiert werden, was man mißt und warum man es messen möchte. iii) Man steht außerdem in der Gefahr, sich in der Beliebigkeit des zu Messenden zu verlieren und das zu Messende nicht mehr aus einem erwünschten Leistungskonstrukt heraus zu erarbeiten, sondern alles zu messen, „was die Items so mitmessen“. Das war auch die Vorgehensweise bei TIMSS und PISA. Man ist damit von der Meßunschärfe einer herkömmlichen Klassenarbeit nicht weit entfernt, verliert also den wesentlichen Grund für standardisierte Leistungstests.[66] Man verschlechtert damit sogar die Position des Schülers, denn die Unwägbarkeiten

[66] An diesen Überlegungen ist zu erkennen, dass das Defizit von schulischer Leistungsbewertung nur scheinbar in mangelnder Standardisierung liegt. Ein Test, der alles mögliche mitmißt, kann trotzdem hochstandardisiert sein. Er hat aber fast die gleiche Meßunschärfe wie eine Klassenarbeit. Hier reproduziert sich ein Irrtum, der uns auch im Forschungsprozeß oft begegnet, nämlich der Glaube, dass hohe Standardisierung zu präziseren oder „besseren“, breiteren, tieferen oder wenigstens allgemeiner gültigen Erkenntnissen führen würde. Standardisierung führt aber zunächst nur dazu, dass alle Mitglieder einer Population bezüglich bestimmter Aspekte den gleichen Bedingungen unterworfen sind. Das bedeutet zwar, dass bestimmte Rahmenbedingungen (oder auch: bestimmte Dimensionen eines multidimensionalen Kausalkonstrukts) für alle Mitglieder gleich konstruiert sind. Das bedeutet aber noch lange nicht, dass damit die Geltungserzeugung präziser, besser, breiter, tiefer, eindeuti-

in den Aufgabenformulierungen einer Klassenarbeit kann er durch sein im Unterricht erworbenes implizites oder explizites Wissen über den Lehrer heilen, notfalls kann er sogar fragen. Die durch Testfähigkeit entstehenden Unwägbarkeiten sind auf diese Weise nicht zu bearbeiten.

Man sollte den Begriff Testfähigkeit abgrenzen von der Fähigkeit, bei einem Test im Sinne einer Klassenarbeit gut abzuschneiden: Auch bei dieser Fähigkeit geht es zwar z.B. darum, erfolgreich zu erschließen (und das heißt manchmal: erraten oder erahnen), was der Lehrer mit seiner Frage meint und in welcher Tiefe bzw. auf welcher Ebene die Aufgabe zu erfüllen ist. In der Klasse ist aber die Vermittlung dieser Fähigkeit expliziter Bestandteil des Unterrichtsprozesses. Unterricht ist per se eine nichtstandardisierte Angelegenheit. Somit ist er allen Vor- und Nachteilen der Nichtstandardisierung ausgesetzt. Das schlägt sich auch in unterrichtlichen Tests nieder. Die daraus resultierende Relativität von Zensierungen läßt zwei polarisierte Schlußfolgerungen zu: Einerseits kann man eine größere Standardisierung von Leistungsbewertung anstreben. Andererseits kann diese Relativität Anlaß sein, den Zensuren mit einer gewissen Gelassenheit zu begegnen, also u.a. ihre Rolle für die Vergabe von Zukunftschancen ebenso zu relativieren. Sie sollte jedenfalls Anlaß sein, sich der Vielschichtigkeit der Ursachen von Schulerfolg zu stellen – denn Zensuren sind wahrscheinlich das beste Maß für Schulerfolg[67]. Leistungen sind nur *eine* Ursache von Schulerfolg, und Schulerfolg wirkt vielfältig zurück auf Leistungen. Will man das Leistungsprinzip in der Schule stärker zur Geltung bringen, so muß die Kopplung von Schulerfolg an Leistung gesichert werden. Werden nun andererseits in standardisierten Tests irgendwelche anderen als die zu leistenden Fähigkeiten mitgemessen, bedeutet dies wiederum eine Abkehr vom Leistungsprinzip – nur dass jetzt *andere* Nicht-Leistungskriterien einfließen als in der Klasse.

2. Tests setzen Standards. Sie tun dies in umso größerem Maße, je relevanter sie für die Vergabe von Zukunftschancen sind. Sie tun dies aber auch dadurch, dass sie sich als wissenschaftliche Instrumente gerieren. Diese Standards schlagen bis in den Unterricht durch. Dadurch ist es problematisch, wenn in

ger oder wenigstens allgemeiner gültig ist. Am Problem der Geltungserzeugung geht die Standardisierung eher vorbei – wobei natürlich bestimmte Standardisierungselemente die Geltungserzeugung unterstützen können.
Ein beredtes Beispiel für dieses Problem sind die in dieser Arbeit untersuchten Tests: Man kann 180 000 Schüler hochstandardisiert untersuchen. Wenn dabei unklar bleibt, was eigentlich gemessen wird, bleibt die Testaussage begrenzt. Selbst der hohe voyeuristische Wert einer Länderrangskala ergibt sich nicht aus hoher Standardisierung, sondern lediglich aus der großen Anzahl der Beteiligten.

[67] Diese Behauptung bedürfte einer tieferen Argumentation, die hier nicht geleistet werden kann. Die Argumentationsrichtung wäre etwa die folgende: Wenn man ein Maß für Schulerfolg erstellen möchte, dann muß man Schulerfolg definieren und in ein Meßkonstrukt überführen. Der Versuch wäre mit Meßunschärfen und anderen Konstruktionsproblemen behaftet. Bereits die Adressierung von „Schulerfolg“ würde zu unüberwindlichen Schwierigkeiten führen: Verschiedene gesellschaftliche Gruppen haben verschiedene Ansprüche an „Schulerfolg“, die Vielfalt an schulischen Aufgaben müßte in eine gewichtete Form gebracht werden usw. Die Schulzensur ist der Versuch, eine solche Gesamtmessung vorzunehmen. Das „Meßkonstrukt“ ist in einem langen Prozeß entstanden, in dem innerschulische und außerschulische Interessen in das Konstrukt eingeflossen sind. Es ist kaum zu überschauen, welche impliziten und expliziten Elemente hier zusammenfließen. Es handelt sich aber um ein Konstrukt von erstaunlich hoher gesellschaftlicher Akzeptanz: Obwohl die Probleme der „Meßunschärfe“ von Zensuren hinlänglich bekannt sind, sind Zensuren nach wie vor die vorrangigen Instrumente der Vergabe von Zukunftschancen in nachschulischen Feldern.
Im Zusammenhang damit ist bemerkenswert, dass es keine Untersuchung über den Zusammenhang von Zensur und Testleistung bei PISA gibt, obwohl die Zensuren erhoben wurden. Man kann sich unschwer vorstellen, dass Tests schnell als überflüssig angesehen würden, wenn sich herausstellte, dass die ordinale Anordnung erhalten bleibt, und dass vorrangig die Tests problematisiert würden, wenn die ordinale Anordnung nicht erhalten bleibt.

Tests Aufgaben auftauchen, die zu lösen sind, ohne dass man die Fähigkeit, die getestet werden soll, wirklich besitzen muß: In A3 muß man die Rundungsregeln nicht beherrschen, um die Aufgabe lösen zu können. In A5 muß man nichts von Kongruenz wissen, um die Aufgabe lösen zu können. Für den Lehrer ist es schwierig, Elemente von Testfähigkeit in den Aufgaben zu erkennen und zu beheben, wenn er ausschließlich auf die mathematischen Fähigkeiten rekurrieren möchte. Der Lehrer arbeitet unter Handlungsdruck und setzt darauf, dass mathematische Leistungstests wirklich mathematische Leistung testen.

Man trifft in der Debatte um Testfähigkeit auf das Argument, dass diese Fähigkeiten durchaus als Bildungsziele taugen bzw. mit ihnen korrespondieren. Dieses Argument wird anhand der in den Interpretationen rekonstruierten Komponenten von Testfähigkeit diskutiert werden. Es stellt sich dabei im wesentlichen als wenig fundiert und zynisch heraus.

3. Testfairneß ist verletzt, wenn Teile der zu vermessenden Population Teile der gemessenen Fähigkeiten nicht oder in geringerem Maße als andere Teile der Population erwerben konnten. Dies kann z.B. der Fall sein, wenn Inhalte getestet werden, welche in einer der vermessenen Schularten gar nicht unterrichtet wurden. Dies kann auch der Fall sein, wenn mit einem Realitätskontext gearbeitet wird, welcher einer Gruppe völlig unbekannt, einer anderen hingegen vertraut ist. Ideale Testfairneß kann es nicht geben und man muß sich dem stellen.
In Bezug auf Testfähigkeit liegt eine Verletzung von Testfairneß dann vor, wenn Tests Testfähigkeiten mitmessen und gleichzeitig Teile der zu vermessenden Population mehr Gelegenheit als andere Teile dieser Population hatten, diese Testfähigkeiten zu erlangen. So gab es eine kurze, aber intensive Debatte über Testfähigkeit, als die ersten TIMSS-Ergebnisse 1997 in Deutschland veröffentlicht wurden. Insbesondere wurde darauf verwiesen, dass die USA- und die asiatischen Nationalmannschaften viel besser auf den Test vorbereitet gewesen seien, weil in diesen Ländern eine ausgeprägte „Kultur“ des Testens zur Vergabe von Zukunftschancen herrscht. Man ging also davon aus, dass die asiatischen und die USA-Teile der vermessenen Population mehr Gelegenheit als die deutschen Teilnehmer hatten, Testfähigkeiten zu erlangen. Zwei polarisierte Schlußfolgerungen wurden daraus gezogen: Einerseits die Schlußfolgerung, die mangelnde Aussagekraft bei der Interpretation der Resultate zu berücksichtigen und vielleicht sogar auf solche Tests zu verzichten. Andererseits die Schlußfolgerung, die deutschen Teilnehmer ebenso intensiv in Testfähigkeiten einzuüben.

Erfolge von Testtraining

Mittlerweile tendiert diese Debatte in die Richtung des verstärkten Testens auch der deutschen Schüler. Allerdings wird das Problem der Testfähigkeit dabei kaum noch diskutiert. In der früheren Debatte fühlte sich die TIMSS-Gruppe noch genötigt zu behaupten, dass man solche Tests nicht trainieren kann (Baumert u.a. 2000, S.108). Das hieße nun allerdings zugespitzt, dass es Testfähigkeit im hier gemeinten Sinne nicht gibt, denn beim Testtraining geht es nicht um das Training der mathematischen Fähigkeiten, sondern um jene Fähigkeiten, die *neben* den mathematischen Fähigkeiten für den Testerfolg sorgen. Da in dieser Arbeit festgestellt wurde, dass und wie TIMSS und PISA Testfähigkeiten mitmessen, soll näher untersucht werden, wie die TIMSS-Gruppe zum Resultat gelangt, dass man

Tests nicht trainieren kann und es diese Testfähigkeiten also nicht gibt oder man sie vernachlässigen kann.

> „Auch Hagemeisters Einwand, daß in bestimmten Ländern (z.B. Japan) ein Testtraining zu besonders guten Ergebnissen geführt haben könnte, kann man kaum ernst nehmen. Ein *coaching* für spezifische der Struktur nach bekannte Tests, die regelmäßig wiederholt werden, hat leistungssteigernde Effekte, die jedoch sehr schnell eine Obergrenze erreichen. In den USA gibt es eine breite Forschungsliteratur zu den begrenzten Auswirkungen von Test-Coaching. ... In Deutschland liegen Coaching-Untersuchungen mit ähnlichen Resultaten vor, die im Rahmen des Zulassungstests für medizinische Studiengänge durchgeführt wurden (Klieme/Maichle, 1990). Bei repräsentativen Untersuchungen, die auf Zufallsstichproben beruhen, unbekannte Tests benutzen und keinerlei Folgen für die Untersuchungsteilnehmer haben, ist ein Coaching zu vernachlässigen.“ (Baumert u.a. 2000, S. 108)

Die „breite amerikanische Forschungsliteratur“ wird nicht angegeben. Bei der Diskussion um Testfähigkeit geht es um „spezifische der Struktur nach bekannte Tests, die regelmäßig wiederholt werden“. Gerade diese Wiederholung und das Vertrautmachen mit der Struktur ist das Ziel, wenn man fordert, die deutschen Schüler sollten durch verstärktes Testen besser auf internationale Vergleiche vorbereitet werden. Problematisch ist aber, in welcher Weise die TIMSS-Gruppe Forschungsergebnisse verbiegt, um zu begründen, dass Coaching zu vernachlässigen ist:

Klieme und Maichle haben ein Training für Teile der medizinischen Eingangstests entwickelt und durchgeführt. Sie wollten im wesentlichen herausfinden, ob bezahlte Vorbereitungskurse für diese Tests die Chancengleichheit der Kandidaten verletzen können. Im Kern betrifft die Fragestellung also die gleiche Chancengleichheitsdebatte bezüglich Tests wie heute.

Klieme und Maichle haben dazu zunächst die Strategien von guten und von schlechten Aufgabenlösern und die Defizite vorhandener kommerzieller Testtrainings untersucht. Daraus haben sie für zwei Untertests ein Training entwickelt, das meines Erachtens geradezu empfehlenswert ist, wenn man Tests trainieren möchte (Klieme/Maichle 1989). Es ging dabei ausschließlich um strategisches und Metawissen, nicht um die mathematischen oder naturwissenschaftlichen Inhalte, es war also keine „Nachhilfe“. Es gab ein Training zum Textverstehen und zum schlußfolgernden Denken mit 10 Personen und ein Training zum mathematisch-naturwissenschaftlichen Problemlösen mit 11 Personen. Problemlösen meint hier nicht Problemlösen im didaktischen Sinne, sondern es geht um Strategien zum Bearbeiten von Aufgaben, die die Tester als Problemlöseaufgaben bezeichnen.

Das Training (Klieme/Maichle 1990) dauerte sechs Zeitstunden, also acht Schulstunden. Die Zeit reichte nur, um die einzelnen Komponenten des Trainings[68] durchzuführen, sie reichte nicht mehr, um

[68] Beim Test „Medizinisch-naturwissenschaftliches Grundverständnis“ (MGNV): Klassifizieren von MNGV-Aufgaben hinsichtlich semantischer Merkmale des Aufgabentextes, adäquate externe Repräsentation der Textinformation, Herleitung von Behauptungen aus extern repräsentierten Textinformationen.
Beim Test „Quantitative und formale Probleme“ (QFP): idealtypischer Ablauf von Lösungsprozessen, Textanalyse, Repräsentation von Aufgaben, Aufbau eines Schemas für Aufgaben über Konzentrationen, Aufbau eines Schemas für Ursache-Wirkungs-Zusammenhänge mit proportionalen Gesetzmäßigkeiten, Vermittlung von Steuerungs- und Kontrolloperationen.

wirklich Testaufgaben zu trainieren, um also das Trainierte auch an den Testaufgaben anzuwenden. In den trainierten Komponenten zeigten sich tatsächlich Verbesserungen, es trat also ein positiver Trainingseffekt ein. Die Leistung im eigentlichen Test hatte sich allerdings nicht verbessert. Das überrascht nicht, denn die eigentlichen Testaufgaben wurden im Training ja gar nicht mehr behandelt. Es wäre verwunderlich, wenn nach sechs Stunden Training von Einzelkomponenten diese Einzelkomponenten sich von allein in eine Aufgabenlösefähigkeit transformiert hätten. Klieme und Maichle diskutieren das Ergebnis ihres Trainings dann auch recht vielschichtig. Sie verweisen auf die Kürze des Trainings, auf das Transferproblem, auf die problematische Motivationslage der Probanden. Zudem seien die Probanden Erwachsene, deren Problemlösestrategien sich schwieriger ändern ließen als bei Jüngeren. Es zeige sich auch, dass schwächere Problemlöser viel stärker von einem Training profitieren. Und je komplexer die zu bearbeitenden Probleme sind, desto schwieriger sei ein Training dafür (ebenda, S.301 ff.). Sie schließen allerdings in erstaunlicher Weise: „Auch die Resultate dieser spezifischen ... Fördermaßnahmen bestätigen letztlich die Aussage ..., daß komplexe Problemlöseleistungen im Sinne der Subtests MNGV und QFP nicht bzw. nur in relativ geringem Ausmaß trainierbar sind.“ (ebenda, S.307) Diese Schlußfolgerung steht in offensichtlichem Widerspruch zu den Ergebnissen der Untersuchung. Vielleicht erklärt sie sich aus der institutionellen Einbindung der Forscher heraus: Die Untersuchung sollte herausfinden, ob die Testfairneß durch bezahlte Vorbereitungskurse verletzt werden kann. Wäre die Antwort ein „Ja“ gewesen oder wäre auch nur das schwächere „Ja“ dieser Untersuchung herausgekommen, dann hätte der Test massiv verändert oder abgeschafft werden müssen.
Die TIMSS-Gruppe nimmt das verfälschte „Resultat“ auf, obwohl es sich nicht mal auf die Diskussion um Langzeiteffekte von Massentestungen bezieht. Die Studie von Klieme und Maichle wird offensichtlich lediglich vorgeschoben, um unerwünschte Nebeneffekte des Testens wegzudiskutieren. In der Diskussion um Testfähigkeit geht es um Langzeiteffekte bei kindlichen und jugendlichen Schülern, für die direkt oder indirekt Zukunftschancen an Testergebnisse gebunden werden. Man muß sich mit diesen Nebeneffekten, die zu Haupteffekten beim Lernen von Mathematik werden können, beschäftigen, um ihren Charakter und ihren Einfluß abschätzen zu können.

Perspektiven auf Testfähigkeit

In dieser Arbeit wird nicht der Trainingsprozeß betrachtet, sondern es wird untersucht, welche Itemeigenschaften dafür sorgen, dass neben mathematischen Fähigkeiten auch Testfähigkeiten gemessen werden. Ziel dieser Betrachtung ist es, den Beteiligten zu größerer Autonomie gegenüber dem Problem zu verhelfen.

Das Training von Testfähigkeit scheint eine Möglichkeit dazu zu sein, weil Testfähigkeit die Autonomie des Schülers gegenüber dem Testprozeß stärkt. Sie reproduziert aber auch die Autonomiezerstörung, indem sie den Schüler auf Fähigkeiten hin trainiert, die außerhalb von mathematischen Fähigkeiten liegen. Die autonomiezerstörende Grundstruktur von Tests ist nicht zu hintergehen. Sie kann nur durch distanzierte Reflexion gebrochen werden. Außerdem erfordert jedes Testtraining ein Zeitbudget, welches vielleicht für sinnvollere Lerninhalte einsetzbar ist.
Die Erweiterung von Autonomie kann ebenso auf seiten des Lehrers oder des bildungspolitischen

Raums stattfinden. Mit dem Wissen um Komponenten von Testfähigkeit kann man bewußter entscheiden, ob man bereit ist, Instrumentarien zur Vergabe von Zukunftschancen einzusetzen, welche Testfähigkeit mitmessen.
Erweiterung von Autonomie kann aber auch auf seiten der Testentwickler stattfinden. Mit dem Wissen um Komponenten von Testfähigkeit kann auch hier bewußter entschieden werden, was man alles (mit)messen möchte.

Testfähigkeit – eine erste Annäherung

Erinnern wir uns zunächst der Grundstruktur des Testens (siehe 3.2 und 3.6.): Tests werden erstellt, um Eigenschaften von Meßobjekten in einem Meßprozeß zu erfassen. Man hat also zunächst eine Vorstellung davon, was in unserem Fall „mathematische Leistungsfähigkeit" sein soll. Nun operationalisiert man diese Vorstellung, man schafft also Items, die in ihrem Zusammenwirken messen, inwieweit diese Fähigkeit vorhanden ist. Das so entstandene Konstrukt, die „operationalisierte mathematische Leistungsfähigkeit", soll natürlich möglichst identisch sein mit dem, was man sich vor der Operationalisierung unter mathematischer Leistungsfähigkeit vorgestellt hat.
Das operationalisierte Meßkonstrukt trifft – materialisiert in Form eines Testheftes - auf das Meßobjekt, also auf den Schüler. Für den Schüler ist egal, was mathematische Leistungsfähigkeit ist, ob sie richtig operationalisiert ist, ob die getesteten Fähigkeiten relevant sind usw. Für den Schüler ist nur eines wichtig: Er muß die Erwartung des Testers bedienen. Er muß sein Kreuz an der richtigen Stelle machen, er muß die richtige Zahl hinschreiben, er muß eine Antwort notieren, die der auswertende Kodierer mit einem Leistungspunkt belegen kann. Damit ist die Richtung für einen Begriff von Testfähigkeit festgelegt: Testfähigkeit ist die Fähigkeit der Optimierung des eigenen Punktwertes innerhalb des Testkonstrukts. Das heißt insbesondere, dass man erstens in der Lage ist, eine wirklich vorhandene mathematische Fähigkeit in einen Testpunkt umzusetzen, und dass man zweitens in der Lage ist, einen Testpunkt auch dann zu erreichen, wenn man *nicht* über die mathematische Fähigkeit verfügt. Das zeigt zunächst, dass für das Individuum Testfähigkeit umso wichtiger wird, je bedeutsamer der Test für die Vergabe von Zukunftschancen wird. Zur Testfähigkeit gehört aber auch die Fähigkeit, den Test sinnvoll zu verweigern. Wenn z.B. bei PISA die Schule und nicht das Individuum vermessen wird, dann sollte im Sinne der Schule ein „schlechter" Schüler den Test ebenso verweigern wie ein Schüler, der heute sein Leistungsoptimum nicht erreicht. Die Schule muß für eine solche Verweigerung dankbar sein und sie unterstützen.

Komponenten von Testfähigkeit

Nur kurz erwähnen möchte ich allgemeine Testbearbeitungsstrategien, die bereits andernorts ausgiebig dargestellt sind und zu deren Beschreibung ich hier nichts beitragen kann. Das sind Zeiteinteilungsstrategien, Fehlervermeidungsstrategien, Ratestrategien, Strategien zur Ausnutzung versteckter Lösungshinweise und formale Strategien zum deduktiven Erschließen der vermeintlich richtigen Antwort[69]. (siehe z.B. Klieme/Maichle 1989, S. 207)

[69] „- Zeiteinteilungsstrategien (z.B.: das Überspringen von schwierigen Aufgaben, das Markieren von ungelösten

Ebenfalls nur erwähnen möchte ich folgende Aspekte, die sich auch ohne eingehendere Aufgabeninterpretationen erschließen: Wenn man nichts weiß, muß man raten. Wenn man Multiple-Choice-Angebote mit nur einer richtigen Antwort abarbeiten muß, ist es besser, die Bearbeitung bei Erreichen der richtigen Antwort abzubrechen und die Aufgabe zu kennzeichnen. Erst wenn man später noch Zeit hat, sollte man zurückkehren und eine Fehlerkontrolle durchführen. Gleiches gilt für andere Unsicherheiten.
Wenn man sich durch viele überflüssige Informationen („Bauernhöfe") oder durch eine Anhäufung von Variationen der immer gleichen Wortgruppe (A5-Kongruenz, „Bauernhöfe") hindurcharbeiten muß oder wenn man eine Ansammlung von Aussagen abarbeiten muß („Dreiecke"), dann kann man von Konzentrations- bzw. Durchhaltefähigkeit sprechen. Diese Fähigkeit benötigt man zwar oft im Leben, aber es wäre sicherlich wünschenswert, wenn man aus der inneren Verfaßtheit und Ernsthaftigkeit eines Problems heraus durchhalten muß und nicht, weil eine Aufgabe schlecht gestellt ist bzw. weil sie Durchhaltefähigkeit *statt* der eigentlich zu testenden Fähigkeit mißt.

Ich möchte nun Komponenten von Testfähigkeit darstellen, die in den Aufgabeninterpretationen rekonstruiert wurden.

1. Die wohl am einfachsten zu erkennende und zu behebende Komponente von Testfähigkeit begegnete uns in der TIMSS-Aufgabe A1 (*schattieren*). Sie ist durch ungewöhnliche, schwierige, mehrdeutige, vielleicht auch falsch benutzte Wörter im Aufgabentext bedingt. Als Hauptursache für das Auftreten dieser Komponente sind Prätentionen, Übersetzungsfehler und auch mangelnde Sorgfalt bei der Durchsicht der Aufgaben zu nennen: Diese Fehler bewegen sich auf der manifesten Textebene und sind durch sorgfältige Durchsicht der Texte zu beheben. Dem Auftreten dieser Fehler tritt das bei TIMSS und PISA gewählte Verfahren der Beurteilung der Aufgaben durch eine Vielzahl von Experten schlüssig entgegen.

2. Als Verschärfung dieser ersten Komponente läßt es sich ansehen, wenn der Text in eine offen bizzare Form übergeht wie in A2, A5 und „Bauernhöfe": In A2 kann man sich nicht entscheiden, ob es um Gewicht oder um Masse geht, obwohl das für das Problem belanglos ist. In A5 wird der Schüler mit den Begriffen der Kongruenz und der Deckungsgleichheit konfrontiert, wobei Irritationspotential unabhängig davon erzeugt wird, welchen der Begriffe man kennt und ob man ihren Zusammenhang

Aufgaben oder solchen Items, bei denen man sich seiner Lösung nicht ganz sicher ist, das Markieren von Teillösungen, das Anlegen eines Arbeitsprotokolls, aus dem man ersehen kann, wie schnell man vorankommt, usw.),
- <u>Fehlervermeidungsstrategien</u> (sorgfältiges Lesen der Instruktion, Beachten der Aufgabenstellung, Überprüfen der Antwort usw.),
- <u>Ratestrategien</u>,
- <u>Strategien zur Ausnutzung versteckter Lösungshinweise</u> (das Beachten aller Merkmale, hinsichtlich derer sich die Antworten von den Distraktoren unterscheiden könnten – z.B. der Länge, der Position, des Stils der betreffenden Aussagen usw.)
- die <u>Beachtung sogenannter „specific determiners"</u> (gemeint sind Worte wie „immer", „niemals", „alle" usw., die nach Meinung der Veranstalter [von Testtrainings, W. M.] speziell die Distraktoren, also die Falschantworten, kennzeichnen),
- <u>formale Strategien zum deduktiven Erschließen der vermeintlich richtigen Antwort</u> (z.B. auf der Basis inhaltlicher oder formaler Abhängigkeiten zwischen den einzelnen Antwortmöglichkeiten)." (Klieme, Maichle 1989, S.207)

kennt. In „Bauernhöfe“ wird der Quader EFGHKLMN als rechtwinkliges Prisma erläutert. In allen drei Fällen stellte sich als Ursache ein habituelles Problem des „Schulmeisterlichen“ heraus. Hier wird probleminadäquat Wert auf Verwendung von Fachsprache gelegt – und das Fachliche unterminiert. Gleichzeitig wird ein Irritationsmoment geschaffen, denn der Schüler muß einen Umgang mit dieser sprachlichen Verwerfung finden. Konkret muß er entscheiden, ob die begriffliche Doppelung für die Lösung wichtig ist oder nicht. Ein Schüler, der das Schulmeisterliche des Textes erfassen und übergehen kann, erhält hier einen Zeitvorteil: Er spart die Zeit, die jemand benötigt, der erst über das Masse-Gewichts-Problem oder den Prismenbegriff oder über Kongruenz nachdenkt oder es gar tiefgründig in seine Überlegungen Einzug halten läßt.

Die Aufgabe für den Schüler besteht bei diesen ersten beiden Testfähigkeitskomponenten darin, die entstehende Klippe zu umschiffen. Das kann einerseits bedeuten, erfolgreich den Inhalt des „seltsamen“ Wortes zu erfassen – eine verbale Fähigkeit. Bei mehreren möglichen Bedeutungen ist die von den Testern intendierte Bedeutung zu erfassen. Habituell erfordert das, sich auf die von den Testern anvisierte Ebene der Problembearbeitung zu begeben, also nicht zu tiefgründig oder zu oberflächlich zu denken. Wer intellektuell zu weit nach unten oder oben denkt, ist einer erhöhten Gefahr des Scheiterns ausgesetzt. Testfähigkeit hat hier also eine inhaltliche und eine habituelle Dimension und kennzeichnet eine Tendenz zum Mittelmaß.
Die Klippe kann auch umschifft werden, indem man das Seltsame übergeht, und begriffliche oder inhaltliche Exaktheit zu vermeidet. - Es geht nicht darum, das Problem der Aufgabe vollständig zu verstehen, sondern es geht um die im Sinne des Tests richtige Lösung. An dieser Stelle wird deutlich, wie Tests die viel beklagte Resultatsorientierung (statt Inhaltsorientierung) der Schüler reproduzieren.
Das Umschiffen der Klippe muß nicht nur inhaltlich erfolgreich geschehen, sondern es muß auch unter möglichst geringem Zeitverlust geschehen, denn Zeit ist in einem Test eine kostbare Ressource. Testfähigkeit bedeutet dabei zu wissen, dass es auf das einzelne Wort nicht so sehr ankommt und dass man das Seltsame übergehen muß. Man nimmt es möglichst gar nicht zur Kenntnis oder erschließt aus dem Rest des Textes möglichst schnell, dass hier keine Falle lauert. Die Möglichkeit, dass es sich um ein wichtiges Wort handelt bzw. dass ein Begriff der fachlichen Präzision wegen eingeführt ist, ist die einzige große Gefahr für den Testfähigen. Beim Auftreten eines „seltsamen“ Wortes oder einer seltsamen Konstruktion in einem Test ist es aber ausgesprochen unwahrscheinlich, dass es sich um eine begriffliche Präzisierung handelt, die für die Erbringung der richtigen Antwort unbedingt verstanden werden muß. Testfähigkeit bedeutet hier, keine Zeit mit Nachdenken zu verbrauchen.
Die dargestellten Komponenten von Testfähigkeit können kaum als Bildungsziel deklariert werden: Zwar geht es beim Rezipieren von Texten immer auch darum, bei der schnellen Erfassung von Textinhalten die in den Texten auftretenden Prätentionen und Fehler zu „überlesen“, sich also an ihnen vorbei den Inhalt zu erschließen. Daraus läßt sich aber keine Rechtfertigung für das Mitmessens dieser Komponente von Testfähigkeit konstruieren, weil damit eine Tendenz zur Normalisierung von Fehlern verbunden ist: Der Schüler wird gezwungen, Defizite der Testerstellung zu übergehen und damit zu akzeptieren, statt sie zurückzuweisen – das kann er wegen der autonomiezerstörenden Grundstruktur bei Tests nicht ungestraft. Auch die zweite Komponente von Testfähigkeit kann nur unter großen Ver-

biegungen als Bildungsziel deklariert werden: Man müßte dazu voraussetzen, dass das Fachsprachliche einen Wert außerhalb des Fachlichen hat. Mir scheint das Fachsprachliche aber nur einen Wert zu haben, wenn dadurch Fachliches transportiert oder konstruiert wird. Die latente Unterminierung des Fachlichen durch das Fachsprachliche scheint mir als Bildungsziel wenig geeignet.

3. Eine weitere Komponente von Testfähigkeit tritt auf, wenn versucht wird, den Schüler künstlich schneller durch den Test zu schleusen. Dabei wird in einigen Fällen mit der manifesten Konstruktion einer Eindeutigkeit genau diese Eindeutigkeit latent zerstört. Dasselbe Prinzip tritt auf, wenn eine textliche Konstruktion, die die Texterfassung beschleunigen soll, Irritationspotential entfaltet, welches die Texterfassung verzögert.
Das erste Beispiel fand sich in der Konstruktion *Wie viele von den* ***kleinen*** *Quadraten ...* bei A1. Hier wird besonders auf die *kleinen* Quadrate verwiesen. Dieser Verweis vereindeutigt, denn es wird ausgeschlossen, dass sich der Schüler mit den aus den kleinen Quadraten zusammengesetzten „großen“ Quadraten auseinandersetzt. Der Verweis verwirrt aber auch, denn die Wahrscheinlichkeit, dass sich Schüler von sich aus mit den „großen“ Quadraten auseinandersetzen, ist ausgesprochen gering. Man wird also darauf gestoßen, besondere kleine und eventuell sogar große Quadrate zu suchen. Lediglich als Hilfe für den Schüler, der nicht weiß, was Quadrate sind, könnte man sich „kleine Quadrate“ vorstellen. Dieses Argument zerbricht aber daran, dass außer den Quadraten gar nichts da ist, womit man arbeiten kann. Die Vereindeutigung zerstört sich also selbst.
Testfähigkeit besteht hier darin, sich von solchen testvereindeutigenden und beschleunigenden Konstruktionen nicht irritieren zu lassen: Der testfähige Schüler ist also mit solchen Konstruktionen vertraut und weiß (implizit oder explizit), dass es lediglich um Vereindeutigung geht und dass über diese schlichte Funktion nicht hinausgedacht werden muß. Es geht darum, auf eine vielschichtige Auseinandersetzung mit der Aufgabe gerade zu verzichten, also *nicht* über die Rolle von großen und kleinen Quadraten und über die vielfältigen Möglichkeiten des Umgangs mit Mengen in der Zeichnung nachzudenken – wie es der explizite Verweis auf die *kleinen* Quadrate zunächst nahelegt. Wenn man auf vieldimensionales Nachdenken verzichtet und sich auf das Setzen des richtigen Kreuzes konzentriert, dann wird die Bearbeitung der Aufgabe durch *kleine* vielleicht sogar wirklich beschleunigt.

Das gleiche Prinzip wiederholt sich in A1 mit *... muß man ZUSÄTZLICH schattieren ...* Die Großschreibung ist zunächst eine Hilfe, da sie vor der Angabe der insgesamt zu schattierenden Quadrate warnt. Diese Hilfe ist nicht notwendig, weil die durch Multiple Choice angegebenen Lösungsvarianten dem Schüler seinen Irrtum signalisieren würden. Auch an dieser Stelle erfährt ein testfähiger Schüler einen Vorteil, weil er mit einer solchen testbeschleunigenden Konstruktion vertraut ist. Der testunerfahrene Schüler wird eher irritiert sein, weil in der Schriftsprache normaler Texte, auch bei schulischen Texten, Wörter in Großbuchstaben eine derart starke Exponierung erzeugen, dass ein Nachdenken über den Grund der Exponierung angezeigt ist. Testfähigkeit bedeutet hier, den Grund der Exponierung bereits zu kennen: Vermeidung naheliegender Fehler. Der Beschleunigungsvorteil durch diese Exponierung gilt natürlich nur für jenen, der der latenten Aufforderung widersteht, über den Grund der Exponierung nachzudenken. Auch hier bedeutet Testfähigkeit wieder, nicht über den Text nachzudenken, sondern dem Prinzip zu folgen, dass es um das Kreuz an der richtigen Stelle geht, nicht um in-

haltliche Auseinandersetzung.
Die Komponente der irritationshaltigen Beschleunigung findet sich auch in der Formulierung *Welche der Aussagen ... ist FALSCH?* von A5. Ursache ist hier eine Prätention: Eine mathematisch anspruchslose Fragestellung wird zunächst künstlich verkompliziert: Hier ist reine Konzentrations- und Fleißarbeit zu verrichten, deren Anspruch aus dem zu lösenden Problem heraus nicht zu begründen ist. Die künstliche Verkomplizierung durch die Umkehr des Anspruchs – man soll benennen, was falsch ist – motiviert wiederum die Hervorhebung durch Großbuchstaben: Das Ungewöhnliche muß hervorgehoben werden, um eine Verwechslung mit der erwartbaren Anforderung zu vermeiden. In der Variante „... ist richtig" käme der Gedanke, RICHTIG groß zu schreiben, nicht auf.
Ein weiteres Beispiel findet sich in der Aufgabe „Pyramide": *Die Grundfläche einer Pyramide ist ein Quadrat. Jede Kante der skizzierten Pyramide misst 12 cm.* Im zweiten Satz wird die Option des Vorhandenseins zweier Pyramiden eröffnet, nämlich der Pyramide des ersten Satzes und der skizzierten Pyramide. Manifest erfolgt durch die Einfügung des *skizzierten* eine Lesebeschleunigung durch die explizite Verknüpfung von Text und Bild. Latent wird eine Irritation erzeugt. Die Aufgabe an den Schüler lautet, sich von dieser Irritation nicht ergreifen zu lassen, also darüber hinweg zu lesen.

Auch die hier beschriebene Dimension von Testfähigkeit läßt sich als Bildungsziel diskutieren: Schließlich gibt es solche Brechungen zwischen dem textlich Gewollten und dem damit produzierten Irritierenden auch in den Texten, auf deren Rezeption Unterricht die Schüler vorbereitet. Man kann es zum Bildungsziel erklären, einen Umgang damit zu finden und zu lernen, diese Irritationen zu überwinden. Das Argument ist allerdings zynisch: Autonomievergrößerung würde bedeuten, das Auseinanderlaufen verschiedener Textebenen in irgendeiner Weise zu thematisieren. Dem Schüler würde dabei ermöglicht, Distanz zum Text und damit auch zum Prozeß der Leistungskontrolle zu erlangen. Er könnte sich damit intellektuell von schulischen und leistungsbewertenden Prozessen emanzipieren. In eienm Test ist er diesen Prozessen ausgeliefert, weil er den Punkt auch dann nicht bekommt, wenn er die Aufgabe intellektuell brilliant zurückweist. Mir scheint es weitaus einleuchtender, dass die Tester der Verpflichtung unterliegen, Irritationspotential zu vermeiden, indem sie gebrochene Vereindeutigungen und Beschleunigungen unterlassen. Dies ist aber offenbar nur möglich, wenn man diese Brüche überhaupt erkennt. Die erste Voraussetzung dafür ist ein Perpektivwechsel: Der Tester darf sich nicht nur darauf konzentrieren, was er hören will, sondern muß sich fragen, was der Text wirklich verlangt und ob das mit dem zusammenläuft, was er will. Die zweite Voraussetzung ist dann nur noch eine gewisse Textsensibilität. Objektive Hermeneutik bietet dieser Sensibilität ein Instrument methodischer Kontrolle.

4. Die Aufgabe „Äpfel" hatte sich als produktive unterrichtliche Aufgabe herausgestellt, die aber als Testaufgabe ungeeignet ist: Unter anderem wird in Äpfel 2 eine Formel für das Lösen von Äpfel 1 nachgereicht. Testfähigkeit besteht hier nicht vorrangig darin, erkennen zu können, in welcher Weise die Tester die Lösung oder Teile der Lösung bereits mitgeliefert haben. Schließlich hat es wenig Sinn, Aufgaben gezielt auf solche Möglichkeiten hin zu durchsuchen – dazu sind diese Möglichkeiten zu selten.
Testfähigkeit besteht vielmehr darin, den Gedanken der Fehlbarkeit der Tester zuzulassen und die Fehlung dann auch auszunutzen. Immerhin handelt es sich beim Test um ein Instrumentarium, das sich

auf die exponierte Zielgenauigkeit und Sorgfalt des Wissenschaftlichen beruft und bereits durch seinen Umfang und sein Auftreten signalisiert, dass hier viele Leute lange darüber nachgedacht haben, was sie an den Schüler herantragen. Es verlangt ein gewisses Maß an Autonomie, Gelassenheit, Abstand oder Unverfrorenheit, um den Gedanken zuzulassen, dass dieser riesige Apparat in Aufgabe 2 die Formel für die Lösung von Aufgabe 1 reinschreibt.

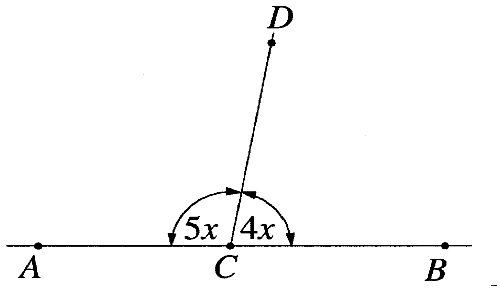
5. Eine Zuspitzung erfährt die eben beschriebene Dimension in der TIMSS-Aufgabe M7, in der die mathematische Anforderung unterminiert wird.

M7

AB ist in dieser Zeichnung eine Gerade.

Wieviel Grad mißt Winkel BCD?

A. 20
B. 40
C. 50
D. 80
E. 100



Hier soll der Schüler offenbar erkennen, dass 9x gleich 180 Grad ist, und daraus das Winkelmaß von 80 Grad für BCD bestimmen. Sehr viel effektiver ist es, mit Hilfe der Multiple-Choice-Angebote *abzuschätzen*, dass es nur 80 Grad sein können. Dass der Schüler *eigentlich* rechnen soll, erkennt man am ersten Satz und an der Tatsache, dass die außergewöhnlichen Angaben 5x und 4x angebracht sind. In einer Schätzaufgabe würde so etwas nicht vorkommen (Interpretation siehe Meyerhöfer 2001).
Ein Schüler, der das Problem rechnerisch nicht lösen kann, hat hier Glück, er kommt nämlich nicht in Gefahr, Zeit zu verschwenden. Für ihn besteht lediglich die Aufgabe, sich zu trauen, einfach das anzukreuzen, was er sieht. Das ist nicht trivial, denn mancher Schüler traut sich nicht, das Offensichtliche hinzuschreiben, wenn er merkt, dass er eigentlich rechnen soll. Ein Schüler, der das Problem rechnerisch lösen *kann*, kommt ebenfalls zum richtigen Ergebnis - wenn er nicht in die Fallen der Lösungsangebote A oder E fällt. Er verbraucht aber sehr viel von der kostbaren Ressource Zeit. Um diese Zeit einzusparen, benötigt er eine gewisse Unverfrorenheit gegenüber der rechnerischen Anforderung, gepaart mit einer gewissen Cleverneß im Erkennen der Möglichkeiten durch Multiple Choice. Für das Problem der Testfähigkeit ergibt sich damit eine weitere Komponente: Man muß sich trauen, einen nichtrechnerischen Weg zu gehen, auch wenn offenbar Rechnen verlangt ist, man muß also unverfroren gegen die Anforderung handeln, denn es geht nicht um den mathematischen Inhalt, sondern um das Kreuz an der richtigen Stelle. Die Aufgabe M7 eignet sich geradezu ideal dazu, Menschen zu identifizieren, die sich clever und effektiv der eigentlichen Anforderung stellen und dabei unverfroren gegen die manifest intendierte Aufgabe handeln. Im Vergleich dazu erscheint das Bedienen der Intention als braves Abarbeiten von fehlerbehafteten und probleminadäquaten mathematikunterrichtlichen Techniken.

Die gleiche Dimension von Testfähigkeit wird in den Aufgaben „Bauernhöfe“ und „Dreieck“ mitgemessen. Dort ist (in unterschiedlich starkem Maße) die Verwendung von lokalem Satz- und Formelwissen gefragt. Tendenziell schneller sind die Wege über Intuition bzw. Messen. Den höchsten Zeitverlust hat dort derjenige, der mathematisch denkt und handelt.
Reinhard Woschek hat im Rahmen seiner (noch nicht veröffentlichten) Dissertation untersucht, auf welch unterschiedliche Weisen deutsche und Schweizer Schüler TIMSS-Aufgaben lösen. Bei M7 stellte er fest, dass deutsche Schüler fast nur rechnen und auch oft damit scheitern. Schweizer Schüler hingegen schätzen fast nur. Es gibt natürlich auch in Deutschland Lehrer, die möchten, dass ihre Schüler an dieser Stelle schätzen - jedenfalls wenn nicht gerade das Aufstellen von und Umgehen mit Gleichungen angesagt ist. Testfähigkeit läuft hier zwar gegen die Aufgabenintention, hat aber durchaus einen Charakter, der Lehrintentionen entsprechen kann. - Man kann durchaus wollen, dass die Schüler mit gegebenen Problemen nicht stur rechnerisch umgehen, sondern sie der Situation adäquat möglichst effektiv lösen. Allerdings bleibt unklar, warum man gerade das künstliche und statische Instrument der Mutiple-Choice-Aufgabe wählen sollte, um sich einem dynamischen, problemadäquaten und effektiven Umgang mit mathematischen Problemen zu nähern, die noch dazu nach rechnerischer Bearbeitung verlangen.

6. Eine elementare, aber etwas trivial wirkende Komponente von Testfähigkeit läßt sich in die Aufforderung umschreiben: Egal wie wenig du weißt, schreibe immer irgend etwas hin. Die Multiple-Choice-Variante dieser Aufforderung heißt: Wenn du nichts weißt, dann kreuze irgend etwas an, und zwar möglichst das, was dir am meisten einleuchtet. Die Diskussion um das Raten bei Tests (siehe nächsten Abschnitt) kann man darauf zuspitzen, dass sich alle Populationsunterschiede in der Testleistung mit dem unterschiedlichen Grad der Verinnerlichung dieser Komponente von Testfähigkeit erklären lassen. Diese Behauptung ist allerdings ebensowenig überprüfbar wie die Behauptung, Raten spiele keine Rolle.
Ich möchte das Problem hier nur für offene Antwortformate diskutieren: In der Aufgabe „Äpfel 2“ heißt es: *Es gibt einen Wert für n, bei dem die Anzahl der Apfelbäume gleich groß ist wie die Anzahl der Nadelbäume. Bestimme diesen Wert und gib an, wie du ihn berechnet hast.*
Es sind aber zwei Werte, nämlich 0 und 8. Aus den Lösungskodierungen der PISA-Gruppe geht hervor, dass ein Schüler, der $n = 8$ angibt, den Lösungspunkt erhält, auch wenn er keine Begründung bzw. Berechnung angibt - wenn er also die Aufgabenstellung nicht erfüllt. Ein Schüler, der nur $n = 0$ angibt, erhält hingegen keinen Punkt, selbst wenn er seine Antwort begründet und es bei diesem Wert beläßt, weil schließlich im Aufgabentext nur *ein* Wert gefordert ist. Man kennt natürlich nie die Kodierungsanweisungen der Tester, wenn man getestet wird. Aber es wird deutlich, dass es nicht in jedem Fall darum geht, die Aufgabe wirklich zu erfüllen. Bereits das Hinschreiben einer Teillösung führt zum Punkt.
Es liegt nahe einzuwenden, dass die Lösung Null für den realen Kontext eher irrelevant ist. Das stimmt zwar inhaltlich, setzt aber eine Kernerfahrung mit Mathematikunterricht nicht außer Kraft: Dort geht es unsystematisch – das heißt, nicht durchschaubar aus der Sache heraus begründet, sondern aus dem Belieben des Lehrers heraus begründet – immer wieder um solche „Randbetrachtungen“. Für

den Schüler bleibt gerade in Tests undurchschaubar, in welchem Maße er „Randbetrachtungen“ mit zu leisten hat. Die Unsicherheit wird dadurch gestärkt, dass es offensichtlich um das Reale gar nicht geht. Diese Herabwürdigung des Realen läßt in der Aufgabe „Sparen“ den Eindruck entstehen, man könne irgend etwas über den Zinseszins hinschreiben – womöglich sogar, ohne ihn berechnet zu haben – und könnte trotzdem den Punkt erhalten.

Irgendetwas hinzuschreiben erweist sich auch als sinnvoll, wenn man sich die *Kodierungspraxis* vor Augen hält: Ein Kodierer muß in einem entfremdeten Arbeitsprozeß unter Zeitdruck eine große Menge an schlecht lesbaren Schülernotizen entziffern. Er muß versuchen, dem Geschriebenen einen Sinn abzuringen und diesen Sinn mit einer umfangreichen, die Wirklichkeit aber doch nur holzschnittartig erfassenden Bewertungsvorschrift in Einklang zu bringen. Er steht im ständigen Konflikt, dass einerseits die von ihm geleistete Geltungserzeugung dem Anspruch der Wissenschaftlichkeit ausgesetzt ist, dass ihm andererseits aber keine wissenschaftliche Methode der Geltungserzeugung zur Verfügung steht. (Der Konflikt existiert unabhängig vom Bewußtsein des Kodierers. Allerdings sind die Kodierer direkt mit dem Text konfrontiert und dürften am deutlichsten spüren, dass die Kategorisierungen weder das Latente noch die vielen verschiedenen Ausprägungen von Verstehen oder von Können zu greifen vermögen.) Ergebnis seines Tuns soll eine undifferenzierte Null-Eins-Entscheidung sein, und in gewisser Weise ist es auch egal, ob man sorgfältig oder gültig bepunktet oder nicht.
Die Kodierungspraxis birgt also eine sowohl im Kategorisierungsprinzip liegende als auch eine menschliche Komponente von Willkür - und diese Willkür ist ein wesentlicher Unterschied zur Klassenarbeit, nach der der Lehrer immer unter Rechtfertigungs- und damit unter Fairneßdruck steht. Es wird deutlich, dass man wenig Einfluß auf die „Gnadenstimmung“ des Kodierers und auf den Kategorienkatalog hat, dass es aber die Chance auf eine positive Bewertung erhöht, wenn man irgendetwas hinschreibt.
Diese Komponente von Testfähigkeit bewegt sich nah an Fähigkeiten, die in Klassenarbeiten benötigt werden, denn auch dort geht es darum, durch das Hinschreiben von Fragmenten „Punkte zu schinden“, selbst wenn man wenig weiß. Auch dieser Komponente von Testfähigkeit mag man deshalb „Bildungswert“ zuschreiben. Es ist aber ein rein innerschulischer Wert: Das Versammeln von Halbwissen oder Fragmenten von Wissen dient hier keiner Annäherung an Bildungsgut durch Versammeln des bereits Gewußten, durch seine Reflexion, Aufarbeitung und Erweiterung. Es dient lediglich dem Bedienen fremdgesetzter Anforderungen in einer asymmetrischen Konstellation, deren inhaltliche Füllung zunächst keinem Bildungsprozeß dient.

7. Wenn Tester sich dem Realen außerhalb der Mathematik zuwenden, öffnet sich ihnen ein mannigfaltiges Potential der Produktion von Verwerfungen, deren Bearbeitung Testfähigkeiten erfordern. Da gibt es in „Bauernhöfe“ Dachböden, die Quadrate sind, da gibt es Mitten von Strecken, da werden Modellierungsanforderungen behauptet und zerstört. In D2 bleibt offen, ob es sich um ein Küchenexperiment, ein Gedankenexperiment oder ein Laborexperiment handelt. Selbst in der Aufgabe „Dreieck“, die ein rein mathematisches Problem anspricht, gelingt eine Verwerfung. Dort wird ein empirisches Dreieck ABC eingeführt und behauptet: Die Seite $\overline{AB}$ des Dreiecks ABC ist 6 cm lang. Das ist sie zweifach nicht.

Im PISA-Kapitel wurde diskutiert, wie sich diese Verwerfungen vermeiden lassen: Grundbedingung ist, das Reale und das Mathematische in ihrer Autonomie und Authentizität zu respektieren. Damit ist die Grundrichtung der hier zu beschreibenden Testfähigkeit abgesteckt: Die Nichtrespektierung der Autonomie und Authentizität ist zu bearbeiten.

In A2 wird zwischen dem Realen, dem Mathematischen und dem Physikalischen hin- und herverworfen. Eine Möglichkeit des Scheiterns ergibt sich dort, wenn man das für einen Ziegelstein hält, was wie ein Ziegelstein aussieht, nämlich der „halbe" Ziegelstein. Der Fehler liegt nahe, weil Ziegelsteine mit quadratischem Querschnitt uns seltener begegnen und weil man sie nie längs teilt, wie das hier geschehen ist. Wir lernen für das Problem der Testfähigkeit: Man soll nicht dem Schein des Realen glauben. Man muß sich also in die spezifische Realität der Tester begeben.
In dieser Realität werden Ziegelsteine längs geteilt, Mutter ruft „Zinseszins!", Schülerinnen zeichnen Dächer von Bauernhöfen, die keine Bauernhöfe sind, und Carola holt Alice nie ein. Diese Welt ähnelt der Welt der Schulbücher und sicherlich auch der Welt von Mathematikunterricht. Insofern läuft die Fähigkeit, sich in die Realität der Tester zu begeben, wahrscheinlich mit der Fähigkeit zusammen, sich in die spezifische Realität von Mathematikunterricht zu begeben. Diese Komponente von Testfähigkeit hat also eine gewisse Ähnlichkeit mit einer Fähigkeitskomponente, die auch im Mathematikunterricht thematisch ist. Der Unterschied ist allerdings ein konstitutiver: Im Mathematikunterricht fordert der Bildungsgedanke, die Spezifität des Realen und des Mathematischen in den Blick zu nehmen. – Es geht hier gerade nicht um die unreflektierte Übernahme von Aufgabenmustern. Diese unreflektierte Übernahme würde man einem Mathematikunterricht zuschreiben, der seinen Bildungsauftrag nicht erfüllt. Deshalb ist es problematisch, dass in keiner einzigen der untersuchten TIMSS-Aufgaben und in keiner einzigen der veröffentlichten PISA-Aufgaben die Spezifität des Realen und des Mathematischen thematisch ist. Am Bildungsauftrag von Mathematikunterricht arbeiten diese beiden Tests diesbezüglich vorbei. Die hier mitgemessenen Komponenten von Testfähigkeit bedienen lediglich die Anpassung an Mathematikunterricht, der seinen Bildungsauftrag nicht erfüllt.
Es ist bedenklich, wie wenige Schulbuchaufgaben sich der Spezifität des Realen und des Mathematischen stellen, insofern ist der Bildungsauftrag von Mathematikunterricht zum Teil gegen die Praxis von Schulbuchaufgaben gerichtet. Das zerstört aber den Auftrag nicht, sondern behindert lediglich seine Umsetzung (und illustriert die mangelnde Verankerung des Bildungsauftrags im Feld). In Tests gibt es aber nichts als die Aufgaben selbst. Sie *konstituieren* das Ganze.

8. Eng verbunden mit der Forderung, sich dem Schein des Realen nicht hinzugeben ist die Forderung, das Problem nicht ernst zu nehmen. Wenn man in A2 das Problem ernst nimmt und durchrechnet, bekommt man heraus, dass der Ziegelstein 2,62 kg wiegt. Dafür gibt es kein Multiple-Choice-Angebot. Man würde demzufolge auf 3 kg runden und damit ein „falsches" Resultat erhalten. Hier liegt also ein Fall vor, in dem ein Schüler zum im Sinne der Tester falschen Resultat gelangen würde, obwohl er ein anspruchsvolles Problem löst und wahrscheinlich auch das kann, was die Tester zu messen glauben. Man könnte vereinfacht sagen: Ein Schüler, der „zu klug" ist, gelangt zum falschen Resultat. Defizitär formuliert: Dieser Schüler erkennt nicht, auf welcher Ebene er hier argumentieren soll.

Das Irritationsmoment liegt auch vor, wenn man lediglich über das Bild „stolpert", weil man die Problemstellung ernst nimmt. Hier besteht die zusätzliche Aufgabe zu erkennen, dass man das Problem nicht ernst nehmen darf, sondern eine schlichtere Überlegung anstellen soll. Ein Schüler mit Testfähigkeit, der exakte Überlegungen von vornherein ausspart, erfährt damit bei dieser Aufgabe einen zeitlichen Vorteil. Man könnte diese Komponente von Testfähigkeit also so formulieren: Du sollst nicht das Problem ernst nehmen und lösen, welches gestellt ist, sondern du sollst herausfinden, was die Tester wollen, dass du es hinschreibst bzw. ankreuzt. Aus den Erkenntnissen über das Testen kann man hinzufügen: Es ist wahrscheinlicher, dass du schlicht arbeiten sollst, als dass du kreativ oder intellektuell anspruchsvoll arbeiten sollst.
Wer in der Aufgabe „Bauernhöfe" das Problem ernst nimmt, der erfährt nicht mal, *welche* Länge er bestimmen soll, weil der zu berechnende Balken einen trapezförmigen Querschnitt haben muß. Glücklicherweise zerstört der Text die Modellierungsanforderung sehr gründlich, so dass man das Problem nicht ernst nehmen wird.

Empirisch zeigt sich, dass Testfähigkeit unvermeidbar dort auftritt,
- wo Multiple-Choice-Angebote das Raten ermöglichen,
- wo offene Antworten kategorial in Null-Eins-Entscheidungen kodiert werden.

Testfähigkeit tritt vermeidbar dort auf,
- wo ein Gegeneinanderlaufen von latenter und manifester Textebene zu Irritationspotential führt,
- wo der Inhalt nicht ernstgenommen wird, um den es zu gehen scheint. Dabei ist es zunächst nicht so, dass der Schüler den Inhalt nicht ernst nimmt, sondern der Aufgabenersteller, der den Text erstellt, nimmt den Inhalt nicht ernst.
- wo Mehrdeutigkeiten bzw. Unschärfen auftreten bezüglich dessen, was gemessen wird und was gemessen werden soll.

Dieses empirische Ergebnis verwundert nicht: Testfähigkeit spielt eben immer dort eine Rolle, wo die Aufgabe schlecht konstruiert ist, wo also latente und manifeste Textebene auseinanderlaufen, wo der mathematische Inhalt didaktisch verworfen wird und wo der Operationalisierungsprozeß nicht sorgfältig verlaufen ist, wo der eigene mathematikdidaktische Habitus nicht reflektiert wird. Diese Testfähigkeiten werden also im Sinne von Vermeidung bearbeitet, wenn die Tester das Zusammenlaufen von latenter und manifester Textebene bearbeiten, wenn sie didaktischen Illusionen und Verschleierungen selbst nicht aufsitzen und wenn sie sorgfältige Operationalisierungen ihrer Meßkonstrukte vornehmen.

Raten

Das Problem des Ratens scheint bei einer Multiple-Choice-Aufgabe ebenso konstitutiv wie kaum der Rede wert: Raten ist hier offensichtlich systematisch ermöglicht und läßt sich nicht verhindern.
Im vorigen Abschnitt wurde bereits deutlich, dass es Bestandteil von Testfähigkeit ist, raten zu „können“, es sich also „zu trauen“. Ebenso kann man es den Testfähigkeiten zuordnen, dass man in der Lage ist, Möglichkeiten zur Erhöhung der Ratewahrscheinlichkeit zu nutzen. Zum Beispiel erhöht man bei A4 die Ratewahrscheinlichkeit von 25 % auf 50 %, wenn man erkennt, dass Carola in der Konstruktion der Tester nicht schneller wird als Alice. Wenn man nun auch noch erkennt, dass Alice ihren Vorsprung ausbauen wird, dann verbleibt nur noch die „richtige“ Lösung.
An dem Beispiel wird deutlich, dass zur Erhöhung der Ratewahrscheinlichkeit ein teilweises Verständnis des Problem notwendig ist. Bei A4 ist es ein inhaltliches Vorverständnis des Proportionalitätsgedankens. Im Beispiel der PISA-Aufgabe „Rechteck“ erhöht bereits formales Wissen die Ratewahrscheinlichkeit auf 50 %.

Auch die intuitiven Lösungswege bei „Bauernhöfe 2“ und „Dreieck“ kann man in der Nähe des Ratens allozieren. Das hat mit der Parallelität von Intuition und Raten zu tun: Intuition ist Bestandteil von Problemlöseprozessen. Es handelt sich bei intuitivem Vorgehen um Hypothesenbildung. Das macht auch die Unsicherheit intuitiven Vorgehens aus, denn es ist eine Hypothesenprüfung notwendig, die außerhalb des Intuitiven stattfinden muß. Es gibt Probleme, bei denen die Hypothesenprüfung so einfach ist, dass man auch sie (fälschlich) als intuitiv wahrnimmt. Die Unsicherheit eines *rein* intuitiven Problemlösens besteht darin, dass keine Hypothesenprüfung stattgefunden hat.
Spricht man nun von jenem Aspekt von Raten, der parallel zur Intuition gelagert ist, so macht den Unterschied zwischen beiden nicht der *Charakter* des Ratens, sondern sein Gegenstand: Von Raten spricht man, wenn kein Problem zu lösen ist. Je deutlicher wird, dass kein Problem zu lösen ist, desto eher spricht man vom Raten statt von Intuition. In diesem Sinne ist es auch sinnvoll, dem Raten einen Bildungswert zuzusprechen – man meint dann gar nicht das Raten, sondern man meint, einen produktiven Umgang mit seiner Intuition zu finden.
Vom Raten ist insbesondere dann zu sprechen, wenn eine Lösung für die Aufgabe bereits vorliegt – wenn also auch in diesem Sinne kein Problem vorliegt. Schule ist ein exponierter Ort für das Stellen von Aufgaben oder Problemen, für die bereits eine Lösung vorliegt. Raten ist in diesen Fällen immer dann thematisch, wenn es sich nicht um eine Lern-, sondern um eine Leistungssituation handelt. (Man kann das Auftreten von Raten geradezu als Indikator dafür benutzen, ob es sich um eine Lern- oder eine Leistungssituation handelt. Zum Beispiel entfalten fragend-entwickelnde Lehrer-Schüler-Gespräche ihr produktives Potential, indem sie unter anderem auch die intuitiven Durchwebungen von Lernprozessen herausfordern; sie zerstören gleichzeitig produktives Potential, indem sie den Schüler zwingen zu erraten, was der Lehrer hören will.) Raten bedeutet dann, der bereits vorhandenen Lösung möglichst nah zu kommen. Das macht auch die Attraktivität des Ratens in der Multiple-Choice-Situation aus: Hier ist die Lösung nicht nur vorhanden, sondern sie ist sogar disjunkt und punktgenau vorgegeben. Man kann geradezu von einer Wahlverwandtschaft von Multiple-Choice und Raten spre-

chen. Es wird deutlich, dass mit Multiple-Choice-Aufgaben im Kern keine kognitive Problemlösepraxis angesprochen ist, sondern die Fähigkeit der Reproduktion bereits gelöster Probleme getestet wird. Im Lehrer-Schüler-Gespräch ist das Raten viel größeren Unsicherheiten ausgesetzt. Aber auch hier bedeutet Raten, sich an die *vorgegebene* Lösung heranzubewegen. Eine andere als die im Kopf des Lehrers vorgegebene Lösung stellt nämlich eine Störung dar.

Die Darstellungen machen deutlich, dass der Begriff der Ratewahrscheinlichkeit eine gewisse Mißverständlichkeit in sich birgt: Das ist die Wahrscheinlichkeit dafür, dass der Schüler rät. Aus der Sicht des Schülers ist das die Gewinnwahrscheinlichkeit, wobei so getan wird, als ob es sich um ein Lottospiel handelt. Die eben entwickelte Sicht auf das Raten verweist aber eher auf ein Börsenspiel. An der Börse geht es nicht nur um das Tippen einer Zahl, sondern um einen durchaus intellektuellen Umgang mit vielfältigen Randbedingungen, gepaart mit einer Menge an zufälligen Größen. - Allerdings ist dort das Ergebnis nicht bereits bekannt.

Der Umgang mit dem Raten bei PISA

Die PISA-Gruppe diskutiert den Fall, dass „Probanden bei fehlendem Wissen Alternativen rein zufällig ankreuzen. In diesem Fall geht die Lösungswahrscheinlichkeit einer solchen Aufgabe nicht gegen Null, sondern liegt bei m Antwortalternativen für leistungsschwache Probanden in der Nähe von $\frac{1}{m}$.“ (Knoche u.a. 2002, S. 167)

Diese Auffassung bewegt sich in der Sichtweise der klassischen Testtheorie, die das Problem des Ratens mit Ratekorrekturen („Zufallskorrekturen“) bearbeitet (vergleiche z.B. Lienert 1989, S.82 ff.) Allerdings sind diese Korrekturen mit inadäquaten Vorannahmen belastet, weil man für diese Korrekturen voraussetzen muß, wie oft ein Schüler geraten hat. Man kann z.B. nicht einfach voraussetzen, dass der Schüler den Test als Lotterie behandelt hat. Man kann aber ebensowenig voraussetzen, dass alle Falsch-Antworten durch Raten entstanden sind - genau das tut aber Lienert (1989, S.83). Die obigen Betrachtungen zum Charakter des Ratens zeigen gerade, dass Aussagen über Ratehäufigkeiten im Grundsatz inadäquat sind, weil das Raten mit dem Wissen untrennbar verwoben ist und deshalb oftmals gar nicht festzustellen ist, ob Raten oder Wissen zur Lösung geführt hat bzw. welche Kombination von beidem. Auch eine Selbstauskunft des Schülers unterliegt dieser Unschärfe.

Die PISA-Gruppe behauptet nun, das Problem des Ratens in den Griff zu bekommen, indem sie das Raten von vornherein verhindert:

> „Wird darauf geachtet, dass die *Distraktoren* bekannten Fehlertypen entsprechen oder wenigstens für Unkundige attraktiver als die richtige Antwort wirken, so lässt sich der Rateeffekt soweit abschwächen, dass die Schätzung der Schwierigkeitsparameter kaum noch verzerrt wird. *Sinnhafte* Distraktoren haben zudem noch den Vorteil, diagnostische Aussagen zu ermöglichen. Zur Verdeutlichung sollen zwei nationale technische Aufgaben dienen:

MULTIPLIKATION

Multipliziere aus und kreuze die richtige Antwort an: $(2x - 3y)^2 =$

- □ $4x^2 - 9x^2$
- □ $4x^2 + 6xy + 9y^2$
- □ $4x^2 - 6xy + 9y^2$
- □ $4x^2 - 12xy + 9y^2$
- □ $4x^2 - 12xy - 9y^2$

Hier sind vier der üblichen Fehlermuster unter den Distraktoren vertreten.
Die Lösungshäufigkeit lag an deutschen Gymnasien immerhin bei 66 %, über alle Schulformen dagegen nur bei 35%. Dass die Aufgabe nicht zum Raten animierte, zeigen die bevorzugte Fehllösung „ $4x^2 - 9y^2$ " und der sehr gute *Itemfit*." (Knoche u.a. 2002, S. 167)

Die Bevorzugung der Fehllösung „$4x^2 - 9y^2$ " zeigt keineswegs, dass diese Aufgabe nicht zum Raten animierte. Die Häufigkeit von Raten zeigt sich nicht in mehr oder weniger guter Gleichverteilung der Fehllösungen - das würde lediglich bei reinem Lotterieraten zutreffen. Wir müssen aber zunächst davon ausgehen, dass es bei einer Aufgabe Lotterieraten ebenso gibt wie inhaltlich oder intuitiv gestützes Raten. In dieser Aufgabe ist ebenso einleuchtend, dass das Rateprinzip wirkt, jene Lösung anzukreuzen, die von den anderen absticht. Inhaltlich hat jede Fehllösung ihre eigene Attraktivität. Auch die nicht vertretenen Fehllösung $4x^2 + 12xy + 9y^2$ und $4x^2 + 12xy - 9y^2$ wären nicht weniger „attraktiv" und sollten in einem diagnostischen Sinne vertreten sein (- das Format des Tests diktiert hier die Anzahl „erlaubter" Fehllösungen). Bei dieser Aufgabe bleibt unklar, welche Rolle das Raten spielen kann. Das Raten ist jedenfalls weder ausgeschlossen noch behindert. Das „Lotterieraten" ist von der Attraktivität der Distraktoren ohnehin nicht betroffen.

„RECHNUNG
Berechne und kreuze die richtige Lösung an!

$4 + 3 \cdot (2 + 1)$

- □ 11
- □ 13
- □ 14
- □ 15
- □ 21

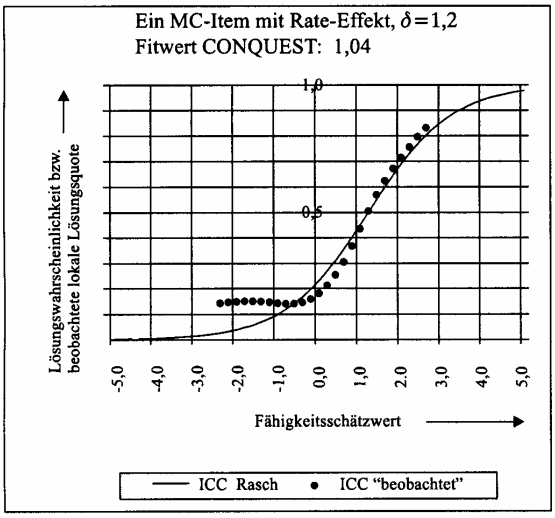
Hier gibt es für jede der Fehlantworten 11, 14, 15 und 21 eine Möglichkeit, sie durch falsches Umgehen mit der Beklammerung zu erhalten.
Mit einer Lösungsquote von rund 61 % war die Aufgabe „Rechung" so leicht, dass sie wohl ebenfalls nur sehr wenige Probanden zum Raten animierte. Dies lässt sich auch daraus schlie-

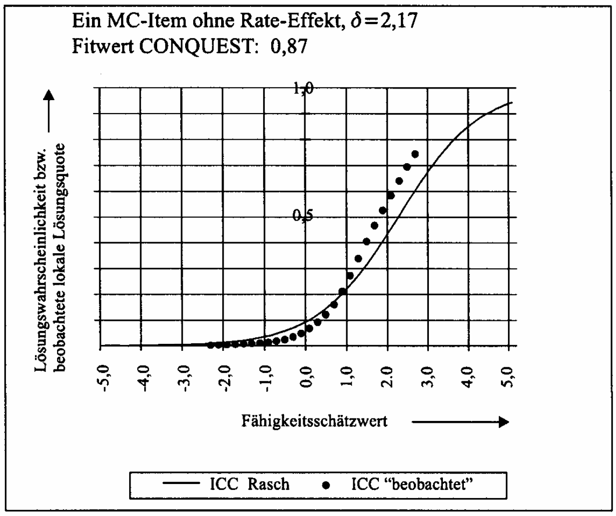
ßen, dass ihr *Itemfit* noch akzeptabel war." (ebenda)

Das Argument, dass die Aufgabe zu „leicht" ist, um zum Raten zu animieren, leuchtet wenig ein. Zunächst bleibt unklar, ab welcher Lösungshäufigkeit das Raten abnehmen soll. Zudem wird die inhaltliche Dimension des Ratens vernachlässigt: Die Aufgabe „Rechteck" hat eine viel größere Lösungshäufigkeit von 85 %. Nichtsdestotrotz animiert sie zum Raten. Wir können keine Aussage darüber treffen, wieviel von den 85 % durch Raten zustande kommt. Aber selbst die verbleibenden 15 % der Schüler hätten ihren Rateerfolg noch erhöhen können, wenn sie wenigstens die Multiplikation oder das Flächenmaß beachtet hätten. – Vielleicht sind aber diese 15 % auch genau jene Hälfte der Ratenden, die Pech hatten.

Itemfits

Die von der PISA-Gruppe angeführten Itemfits sind nicht geeignet, um das Problem des Ratens zu bearbeiten. Die folgenden Graphiken wurden mir freundlicherweise von Herrn Lind zur Verfügung gestellt.





Die durchgehende Linie zeigt die ideale Itemcharakteristik nach dem Rasch-Modell (siehe Knoche u.a. 2002, S. 164 ff.) Die Punkte sind aus den gemessenen Lösungshäufigkeiten entstanden. Der linken Grafik wird deshalb ein Rate-Effekt zugesprochen, weil die im Sinne des Tests schwachen Schüler „zu gut" sind, weil also die Meßwerte für die schwachen Schüler besser sind, als sie nach der idealen Rasch-Charakteristik sein dürften. Dahinter steckt ein Gedanke, der grundlegend für das Rasch-Modell ist: Die Aufgaben müssen so beschaffen sein, dass ein schwächerer Schüler die Aufgabe mit geringerer Wahrscheinlichkeit lösen kann als ein besserer, also erfolgreicherer Schüler. Es sind also z.B. Aufgaben verboten, die von schwachen Schülern häufig gelöst werden, von erfolgreicheren Schülern aber selten gelöst werden. Das Rasch-Modell läßt also nur Aufgaben zu, die der modellinternen Forderung nach Eindimensionalität gerecht werden. Diese Forderung ist in sich bereits problematisch,

weil das Testmodell eine Testcharakteristik erzwingt, die die Komplexität intellektueller Prozesse herabwürdigt.
Man mag das hinnehmen, weil man die Nutzung des Rasch-Modells für unerläßlich hält. Man kann aber nicht behaupten, dass die Bedienung der Modellogik das Problem des Ratens adäquat bearbeiten würde: Mit den Itemfits wird nicht untersucht, ob bei einer Aufgabe geraten wird. Es wird untersucht, ob Schüler mit einem immer geringeren Fähigkeitsschätzwert eine immer geringere Lösungswahrscheinlichkeit haben. Das kann mit Raten zusammenhängen, es kann aber auch mit einer geringen Schwierigkeit zusammenhängen („Bodenbildung") oder mit einem Aufgabenpotential, welches schwache Löser begünstigt. Wenn man davon ausgeht, dass Raten-Können oder andere Testfähigkeiten den Testerfolg relevant bestimmen, kann es sogar sein, dass diese Aufgaben gerade *nicht* diese Fähigkeiten messen. Nur eine Untersuchung der Aufgabe selbst gibt hier Aufschluß.
Die Argumentation der PISA-Gruppe geht außerdem von der falschen Annahme aus, dass schwache Schüler in stärkerem Maße raten als erfolgreichere Schüler. Ein schwacher Schüler ist aber zunächst ein Schüler, der die Testanforderungen in geringerem Maße bedient als ein besserer Schüler. Es gibt keinen Grund anzunehmen, dass er deshalb auch öfter rät. Ebensogut ist anzunehmen, dass der erfolgreiche Schüler öfter rät und gerade deshalb erfolgreicher ist.
Zusammenfassend läßt sich konstatieren, dass die PISA-Gruppe das Raten nicht ausgeschaltet oder behindert hat. Der technokratische Lösungsansatz über Itemfits hält einer näheren Betrachtung nicht stand.

Zur Eindeutigkeit des Meßprozesses

In den Aufgabeninterpretationen hat sich gezeigt, dass bei TIMSS und PISA die meisten Aufgaben „unscharf" messen, dass also nicht eindeutig benennbar ist, welche Fähigkeit mit einer Aufgabe gemessen wird. Eindeutige Meßprozesse wurden in einigen Aufgaben konstatiert, die wenig komplexe Fähigkeiten messen, z.B. „Bauernhöfe 1". Insgesamt entsteht eher der Eindruck einer „Unschärferelation": Je komplexer die zu messende Fähigkeit, desto schwieriger wird es, diese Fähigkeit scharf zu messen. Es bleibt in dieser Arbeit unklar, ob es *prinzipiell* möglich ist, Aufgaben so zu konstruieren, dass auch anspruchsvollere Fähigkeiten *eindeutig* oder *scharf* gemessen werden können. In den Interpretationen zeigten sich aber Problemkreise, welche die Eindeutigkeit des Messens verhindert oder gestört haben. Daraus lassen sich Hinweise ableiten, wie die Eindeutigkeit des Meßprozesses erhöht werden kann:

a) Das Problem der Meßunschärfe entsteht bei PISA unter anderem dadurch, dass nicht zwischen einer guten Testaufgabe und einer guten Aufgabe für den Unterricht unterschieden wird. Man könnte also versuchen, Items zu entwickeln, die mit herkömmlichen Mathematikaufgaben wenig Ähnlichkeit haben, die aber bestimmte höhere bzw. komplexe Fähigkeiten „scharf" messen. Beispiele für dieses Vorgehen, deren Qualität ich hier nicht diskutieren möchte, finden sich in Tests zum räumlichen Vorstellungsvermögen. Dieses Vermögen wird im Geometrieunterricht anhand bestimmter räumlicher Gebilde und Konstellationen geschult, die Tests bedienen sich aber völlig anderer Gebilde und Kon-

stellationen. Der mit diesem Vorgehen verbundene Vergleich dieser beiden Wege, sich dem Problem des räumlichen Vorstellungsvermögens zu widmen, entwickelt gleichzeitig unser Bewußtsein für die Begrenztheit unseres Tuns - sowohl beim Unterrichten als auch beim Testen.
b) Das Problem der Meßunschärfe entsteht bei TIMSS und PISA durch die fehlende Operationalisierung von Fähigkeiten, die man messen möchte. Größere Meßschärfe ließe sich erreichen, wenn man zunächst deutlich macht, was man eigentlich messen möchte. Höheren und komplexen Fähigkeiten würde man sich dann durch Itemkombinationen annähern. Der wichtigste Gedanke scheint hierbei das *Bewußtsein für die Grenzen des Konstrukts* zu sein. Dieses Bewußtsein steht allerdings im Widerspruch zu einer Kultur des Testens, welche den Vergleich nach Rangfolgen in den Vordergrund stellt. Wenn man vorrangig Populationen in Rangfolgen bringen möchte, wird es zweitrangig, nach welchen Kriterien man es tut.

c) Man könnte Meßunschärfe auch dadurch vermindern, dass man Lösungswege ausschließt, indem man sie verhindert. Hierzu sind verschiedenste Konstruktionen denkbar (z.B. Auswahl des Zahlenmaterials; expliziter Ausschluß von Lösungswegen, wenn der Lösungsweg Bestandteil des zu Testenden ist). Das Problem hierbei ist, dass unser Wissen über das kognitive Geschehen beim Lösen einer konkreten Aufgabe noch sehr begrenzt ist. Wir wissen nicht, auf welche Weise Individuen die Probleme lösen können, die wir ihnen stellen. Wir können lediglich die sachlogisch im Problem steckenden Lösungsoptionen benennen, so wie es jeweils am Anfang der ausführlichen Interpretationen geschehen ist. Dabei deuten sich Zusammenhänge und Übergänge zwischen diesen Optionen an. Wie sie sich konkret kognitiv gestalten, wissen wir noch kaum.

d) Die PISA-2003-Gruppe hat im Vortest die Schüler nach ihren Lösungswegen gefragt. Dabei erfaßt man allerdings nicht das Lösungsverhalten der Schüler. Man erfaßt nicht einmal, was die Schüler über ihr Lösungsverhalten glauben. Man erfaßt lediglich, wie die Schüler sich in die Lösungswegkategorien der Tester einordnen. Wir erfahren weder, ob die Schüler die Kategorien sinnvoll bzw. im Sinne der Tester erfassen, noch erfahren wir etwas über Mischwege. Nähere Kenntnis der Leistung des Schülers erlangt man über explizites Abfragen des Lösungsweges. Hier verändert man natürlich den Meßgegenstand. Es verbleibt selbst bei solchem Abfragen das Problem der Auswertung: Kodierer können hier zwar eine äußerlich gleichwertige Bewertung vornehmen, sind aber im Erfassen des latenten Sinns des Schülertexts weniger aussagekräftig als der Lehrer mit seinen Alltagsdeutungen der Texte seiner Schüler. Hier läge also eine schlechtere - aber äußerlich einheitlichere - Variante einer zentralen Prüfung vor.

e) In den Interpretationen zeigten sich vielfältige Probleme des Auseinanderlaufens und Gegeneinanderlaufens von latenten und manifesten Textelementen, welche den Meßprozeß stören. Diese Störungen liegen vor, obwohl sowohl bei TIMSS als auch bei PISA ein aufwändiges Verfahren der Beurteilung der Aufgaben durch eine Vielzahl von Experten stattgefunden hat. Das Expertenverfahren hat also die in dieser Arbeit freigelegten Probleme nicht ausreichend bearbeitet. Der Gedanke liegt nahe, objektiv-hermeneutische Aufgabeninterpretationen in den Prozeß der Testentwicklung einzubeziehen. Allerdings wird sich erst zu zeigen haben, ob die hier aufeinanderprallenden Wissenschaftskulturen produktiv miteinander verzahnt werden können.

Literatur

Adorno, Theodor W. (1955/1973): Zum Verhältnis von Soziologie und Psychologie. In: Aufsätze zur Gesellschaftstheorie und Methodologie. Frankfurt/Main.

Adorno, Theodor W. (1981): Einleitung zu: Adorno, Theodor W. (u.a.): Der Positivismusstreit in der deutschen Soziologie. 9.Aufl. Darmstadt/Neuwied, S.7-79

Aebli, Hans (1980): Denken: Das Ordnen des Tuns. Band 1: Kognitive Aspekte der Handlungstheorie. Stuttgart: Klett

Aebli, Hans (1994): Über das Verstehen von Witzen – eine kognitionspsychologische Analyse und einige pädagogische Schlussfolgerungen. in: Kurt Reusser, Marianne Reusser-Weyenet (Hrg.): Verstehen. Psychologischer Prozeß und didaktische Aufgabe. Verlag Hans Huber, Bern 1994, S. 89-112

Artelt, Cordula; Stanat, Petra; Schneider, Wolfgang; Schiefele, Ulrich: Lesekompetenz: Testkonzeption und Ergebnisse. In: Deutsches PISA-Konsortium (Hrsg.) 2001, S.69-140

Baumert, Jürgen; Lehmann, Rainer u.a. (1997): TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Leske + Budrich, Opladen

Baumert, Jürgen, Eckhard Klieme, Manfred Lehre und Elwin Savelsbergh (2000): Konzeption und Aussagekraft der TIMSS-Leistungstests. Zur Diskussion um TIMSS-Aufgaben aus der Mittelstufenphysik. in: Die Deutsche Schule, 92.Jahrgg. 2000, Heft 1 (S.102-115), Heft 2 (S.196-217)

Baumert, Jürgen; Stanat, Petra; Demmrich, Anke: PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In: Deutsches PISA-Konsortium (Hrsg.) 2001, S.15-68

Blum, Werner (1985): Anwendungsorientierter Mathematikunterricht in der didaktischen Diskussion. in: Mathematische Semesterberichte, Jg.32(1985), S.195-232

Blum, Werner (1996): Anwendungsbezüge im Mathematikunterricht – Trends und Perspektiven – In: G.Kadunz, H.Kautschtisch, G.Ossimitz, E.Schneider (Hrsg.): Trends und Perspektiven - Beiträge zum 7.Internationalen Symposium zur Didaktik der Mathematik in Klagenfurt Sept. 1994 (S.15-38). Wien: Hölder-Pichler-Tempsky

Cohors-Fresenborg, Elmar (1996): Mathematik als Werkzeug zur Wissensrepräsentation. In: G.Kadunz, H.Kautschtisch, G.Ossimitz, E.Schneider (Hrsg.): Trends und Perspektiven - Beiträge zum 7.Internationalen Symposium zur Didaktik der Mathematik in Klagenfurt Sept. 1994 (S.85-90). Wien: Hölder-Pichler-Tempsky

Deutsches PISA-Konsortium (Hrsg.) (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Leske + Budrich, Opladen

Deutsches PISA-Konsortium (Hrsg.) (2002): PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. Leske + Budrich, Opladen 2002

Führer, Lutz (1998): Mathematikunterricht nach dem 7.Schuljahr - warum eigentlich für alle? in: Neue Sammlung 38/1998 S.489 -

Freud, Sigmund (1916/17): Vorlesungen zur Einführung in die Psychoanalyse. In: Freud - Studienausgabe. Band I. Frankfurt/Main 1982, S.37-445

Hagemeister, Volker (1999): Was wurde bei TIMSS erhoben? Eine Analyse der empirischen Basis von TIMSS. in: Die Deutsche Schule, 91.Jahrgg. 1999, Heft 2, S.160-177

Harada, Nobuyuki (1998): Open Lessons of the Subject „Life Environment Studies“ in Japanese Primary Schools. in: Publications of the Japanese-German Center Berlin, Volume 15: Symposium Mathematics and Elementary Science Education, 3.-5.12.1997

Jahnke, Thomas; Herget, Wilfried; Kroll, Wolfgang (2000): Produktive Aufgaben für den Mathematikunterricht der Sekundarstufe I. Cornelsen Verlag, Berlin.

Johnson-Laird, P.N. (1983): Mental models. Cambridge, MA: Harvard University Press

Kaiser, Gabriele (1995): Realitätsbezüge im Mathematikunterricht – Ein Überblick über die aktuelle und historische Diskussion. In: Graumann, Günther u.a. (Hrsg.): Materialien für einen realitätsbezogenen Mathematikunterricht, Band 2. Hildesheim: Franzbecker. S.66-84

Kaiser, Gabriele (1998): TIMSS - woher und wohin? In: mathematik lehren. Heft 90 Oktober 1998

Kaiser, Gabriele (2000): Internationale Vergleichsuntersuchungen im Mathematikunterricht – Eine Auseinandersetzung mit ihren Möglichkeiten und Grenzen. in: Journal für Mathematik-Didaktik 21(2000) Heft 3/4, S.171-192

Kaiser, Gabriele; Knoche, Norbert; Lind, Detlef; Zillmer, Wolfgang (Hrg.) (2001): Leistungsvergleiche im Mathematikunterricht, ein Überblick über aktuelle nationale Studien. Franzbecker. Hildesheim, Berlin

Keitel, Chistine; Kilpatrick, Jeremy (1998): Mathematikunterricht zwischen Wissenschaft und Politik. Rationalität und Irrationalität internationaler vergleichender Studien. in: [Journal] Neue Sammlung. Vierteljahres-Zeitschrift für Erziehung und Gesellschaft. (1998) v. 38(4) S. 513-532.

Kintsch, Walter (1974): The representation of meaning in memory. Hillsdale, NJ: Lawrence Erlbaum

Kintsch, Walter (1994): Kognitionspsychologische Modelle des Textverstehens: Literarische Texte. in: Kurt Reusser, Marianne Reusser-Weyenet (Hrg.): Verstehen. Psychologischer Prozeß und didaktische Aufgabe. Verlag Hans Huber, Bern, S.39 – 54

Kintsch, W.; van Dijk, T.A. (1978): Toward a model of text comprehension and production. Psychological Review, 85, 363-394

Klieme, Eckhard; Bos, Wilfried (2000): Mathematikleistung und mathematischer Unterricht in Deutschland und Japan. In: Zeitschrift für Erziehungswissenschaft 3/2000, S.359-379

Klieme, Eckhard; Maichle, Ulla (1989): Zum Training von Techniken des Textverstehens und des Problemlösens in Naturwissenschaften und Medizin. In: Günter Trost (Hrg.): Test für medizinischen Studiengänge (TMS): Studien zur Evaluation (13.Arbeitsbericht), Bonn: Institut für Test- und Begabungsforschung. S.188-247

Klieme, Eckhard; Maichle, Ulla (1990): Ergebnisse eines Trainings zum Textverstehen und zum Problemlösen in Naturwissenschaften und Medizin. In: Günter Trost (Hrg.): Test für medizinischen Studiengänge (TMS). 14.Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung, S.258-307

Klieme, Eckhard; Neubrand, Michael; Lüdtke, Oliver: Mathematische Grundbildung: Testkonzeption und Ergebnisse. In: Deutsches PISA-Konsortium (Hrsg.) 2001, S.141-191

Knoche, Norbert; Lind, Detlef (2000): Eine Analyse der Aussagen und Interpretationen von TIMSS unter Betonung methodologischer Aspekte. in: Journal für Mathematik-Didaktik 21(2000), Heft 1, S.3-27

Köller, Olaf; Trautwein, Ulrich (2001): Evaluation mit TIMSS-Instrumenten. Untersuchungen in der 8.Jahrgangsstufe an fünf Gesamtschulen. in: Die Deutsche Schule

Levi-Strauss, Claude (1993): Der Blick aus der Ferne. Frankfurt/M.

Lewis, Catherine C. (1998): Improving Japanese Science Education. How „Research Lessons" Build Teachers, Schools, and a National Curriculum. in: Publications of the Japanese-German Center Berlin, Volume 15: Symposium Mathematics and Elementary Science Education, 3.-5.12.1997

Meyerhöfer, Wolfram (2001): Was misst TIMSS? Einige Überlegungen zum Problem der Interpretierbarkeit der erhobenen Daten. in: http://pub.ub.uni-potsdam.de/2001meta/0012/door.htm

Meyerhöfer, Wolfram (2002): Jeder Arbeiter verlegt gleich viel. In: mathematik lehren 114, Oktober 2002

Meyerhöfer, Wolfram (2004): Das Kompetenzstufenmodell von PISA – Eine empirische Dekonstruktion. erscheint 2004

National Council of Teachers of Mathematics (NCTM) (1989): Curriculum and evaluation standards for school mathematics. Reston, VA: NCTM

National Council of Teachers of Mathematics (NCTM) (2000): Principles and standards for school mathematics. Reston, VA: NCTM

NISA (1998): National Institute on Student Achievement, Curriculum, and Assessment: The Educational System in Japan: Case Study Findings. in: http://www.ed.gov/pubs/JapanCaseStudy/title.html (June 1998)

NISA (1999a): National Institute on Student Achievement, Curriculum, and Assessment: The Educational System in the United States: Case Study Findings. in: http://www.ed.gov/pubs/USCaseStudy/index.html (March 1999)

NISA (1999b): National Institute on Student Achievement, Curriculum, and Assessment: The Educational System in Germany: Case Study Findings. in: http://www.ed.gov/pubs/GermanCaseStudy/ (June 1999)

Nôda, Nobuhiko (1998): Teaching and Learning with the „Open-Approach-Method" in Classroom Mathematic Activities. in: Publications of the Japanese-German Center Berlin, Volume 15: Symposium Mathematics and Elementary Science Education, 3.-5.12.1997

OECD (2001): PISA 2000. Beispielaufgaben aus dem Mathematiktest. unter: http://www.mpib-berlin.mpg.de/pisa/index-pisa_sg3.htm

OECD (2003): PISA 2000 Deutschland. Beispielaufgaben Mathematik aus der nationalen Ergänzung zu PISA in Deutschland. Veröffentlichungsstatus unklar.

Oevermann, Ulrich (1983): Zur Sache. Die Bedeutung von Adornos methodologischem Selbstverständnis für die Begründung einer materialen soziologischen Strukturanalyse. in: L. v. Friedeburg und J.Habermas (Hrsg.): Adorno-Konferenz. Suhrkamp Frankfurt, S.234-289

Oevermann, Ulrich (1991): Genetischer Strukturalismus und das sozialwissenschaftliche Problem der Erklärung der Entstehung des Neuen. In: Stefan Müller-Doohm (Hrsg.): Jenseits der Utopie: Theoriekritik der Gegenwart. Frankfurt/M., S.267-336

Oevermann, Ulrich (1993): Die objektive Hermeneutik als unverzichtbare methodologische Grundlage für die Analyse von Subjektivität. Zugleich eine Kritik der Tiefenhermeneutik. in: T.Jung und S.Müller-Doohm (Hrsg.): Wirklichkeit im Deutungsprozeß. Verstehen und Methoden in den Kultur- und Sozialwissenschaften, Frankfurt/M. 1993, S.106-189

Oevermann, Ulrich (1996): Konzeptualisierung von Anwendungsmöglichkeiten und praktischen Arbeitsfeldern der objektiven Hermeneutik. (Manifest der objektiv hermeneutischen Sozialforschung) unveröffentlichtes Manuskript, März 1996

Öhlschläger, Günther (1974): Einige Unterschiede zwischen Naturgesetzen und sozialen Regeln. in: Hans Jürgen Heringer (Hrg.): Seminar: Der Regelbegriff in der praktischen Semantik. Suhrkamp. Frankfurt/Main

PISA 2000: Deutsches PISA-Konsortium (Hrsg.) (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Leske + Budrich, Opladen

Reusser, Kurt (1989): Vom Text zur Situation zur Gleichung. Kognitive Simulation von Sprachverständnis und Mathematisierung beim Lösen von Textaufgaben. Habilitationsschrift, Universität Bern. (Neudruck Zürich 1995)

Reusser, Kurt (1989b): From text to situation to equation: cognitive simulation of understanding and solving mathematical word problems. in: H.Mandl, N.Bennett, E. de Corte & H.R.Friedrich (Eds.): Learning and Instruction in an international context (Vol. II and III). Oxford: Pergamon

Reusser, Kurt (1996): From Cognitive Modelling to the Design of Pedagogical Tools. in: S.Vosuiadou, E. de Corte, R.Glaser, H.Mandl (Eds.): International Perspectives on the Design of Technology-Supported Learning Environments. New Jersey: Lawrence Erlbaum Associates. S.81-103

Rohlen, Thomas P. (1980): The JUKU Phenomenon. in: Journal of Japanese Studies 6 (1980), S. 204 - 242

Schriewer, Jürgen (1999): Vergleich und Erklärung zwischen Kausalität und Komplexität. in: Hartmut Kaelble, Jürgen Schriewer: Diskurse und Entwicklungspfade. Der Gesellschaftsvergleich in den Geschichts- und Sozialwissenschaften. Campus Verlag, Frankfurt/Main; New York

Schümer, Gundel (1997): in: Juergen van Buer, Rainer H. Lehmann u.a. (Hrg.): Erweiterte Autonomie für Schule - Qualität von Schule und Unterricht. Berlin.

Schümer, Gundel (1998): Mathematikunterricht in Japan - Ein Überblick über den Unterricht in öffentlichen Grund- und Mittelschulen und privaten Ergänzungsschulen. in: Unterrichtswissenschaft. 26/1998, Heft 3, S.195-228

Schümer, Gundel (1999): Lehrer in Japan. Arbeitsbedingungen und Arbeitsethos. In: Pädagogik 6/1999 S.30-35

Schupp, Hans (1988): Anwendungsorientierter Mathematikunterricht in der Sekundarstufe I zwischen Tradition und neuen Impulsen. in: Der Mathematikunterricht 34(1988) Heft 6, S.5-16

Searle, John R.(1983, Original 1969): Sprechakte. Ein sprachphilosophischer Essay. Suhrkamp. Frankfurt/Main

Staub, F.C.; Reusser, Kurt (1995): The role of presentational structures in understanding and solving mathematical word problems. In: C.A.Weaver; S.Mannes; C.R.Fletcher (Eds.): Discourse comprehension. Essays in honor of Walter Kintsch (pp. 285 – 305). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stern ; Elsbeth (1998): Die Entwicklung des mathematischen Verständnisses im Kindesalter. Pabst Science Publishers, Lengerich, Berlin, Düsseldorf u.a.

Stevenson und Nerison-Low (1997): To sum it up: Case studies of Education in Germany, Japan and the United States. Michigan: University of Michigan

Stevenson, H.W.; Stigler, J.W. (1992): The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education. New York: Simon and Schuster

Stigler, James W.; Gonzales, Patrick; Kawanaka, Takako; Knoll, Steffen; Serrano, Ana (1999): The TIMSS Videotape Classroom Study. Methods and Findings from an Exploratory Research Project on Eigth-Grade Mathematics Instruction in Germany, Japan, and the United States. U.S.Department of Education. National Center for Education Statistics. Download: http://nces.ed.gov/timss

Stigler, James W.; Hiebert, James (1999): The Teaching Gap. Best Ideas from the World`s Teachers for Improving Education in the Classroom. New York

Van Dijk,T.A.; Kintsch,W. (1983): Strategies of Discourse Comprehension. N.Y.: Academic Press

Wagner, Hans-Josef (2001): Objektive Hermeneutik und Bildung des Subjekts. Velbrück

Weber Max: (1907/1982): R.Stammlers „Überwindung“ der materialistischen Geschichtsauffassung. in: Gesammelte Aufsätze zu Wissenschaftslehre. J.C.B.Mohr, Tübingen 1982

Weigand, Hans-Georg (1997): Überlegungen zur TIMS-Studie. Mathematik in der Schule 35 (1997) 10 S.512-526

Wernet, Andreas (2000): Einführung in die Interpretationstechnik der Objektiven Hermeneutik. Leske + Budrich, Opladen

Wittgenstein, Ludwig (1953/2001): Philosophische Untersuchungen. Dieses Werk wurde 1945/46 fertig gestellt und 1953 erstmals veröffentlicht. Ich zitiere aus der Kritisch-genetischen Edition (Hrg.: Joachim Schulte) die Spätfassung. Suhrkamp, Frankfurt/Main 2001

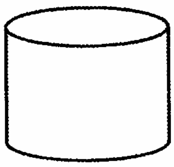
Zais, Thomas; Grund, Karl-Heinz (1991): Grundpositionen zum anwendungsorientierten Mathematikunterricht bei besonderer Berücksichtigung des Modellierungsprozesses. In: Der Mathematikunterricht. Jg.37, Heft 5/1991, S.4-17

Anhang

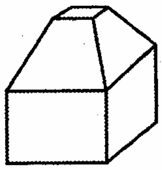
PISA-Aufgabe

BEHÄLTER

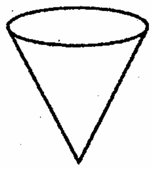
In der Abbildung siehst du drei Arten von Behältern. Diese werden bei gleichmäßigem Zufluss mit Wasser gefüllt.



ZYLINDER



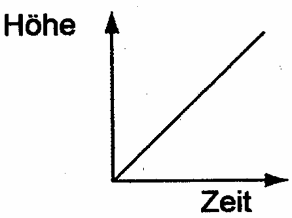
UNREGELMÄSSIGER BEHÄLTER



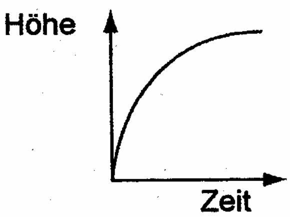
KEGEL

Frage: BEHÄLTER

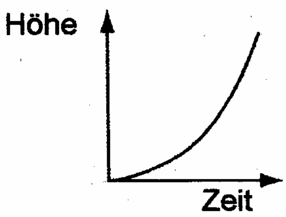
Hier sind mehrere Graphen, die zeigen, wie der Wasserspiegel (Höhe des Wassers) in sechs verschiedenen Behältern beim Einfüllen steigt.



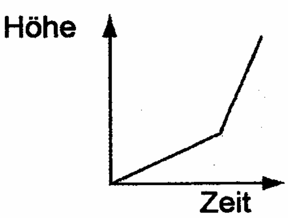
A



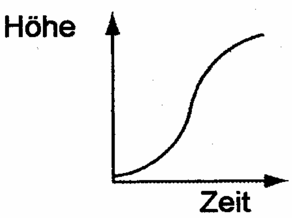
B



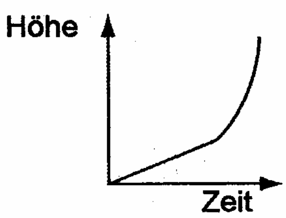
C



D



E



F

Welcher Graph passt jeweils zu den Behältern?

Zylinder: Graph________

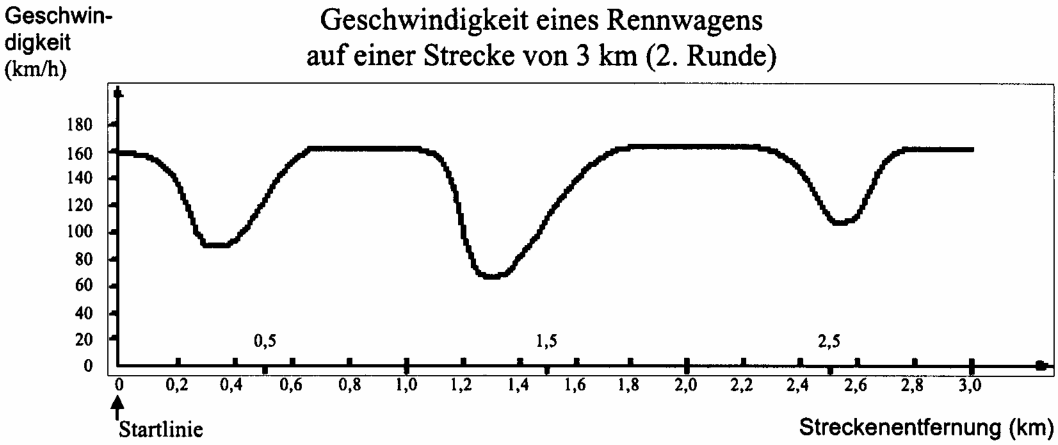
Unregelmäßiger Behälter: Graph________

Kegel: Graph________

PISA-Aufgabe:

GESCHWINDIGKEIT EINES RENNWAGENS

Dieser Graph zeigt, wie die Geschwindigkeit eines Rennwagens während seiner zweiten Runde auf einer drei Kilometer langen ebenen Rennstrecke variiert.



Frage 5: GESCHWINDIGKEIT EINES RENNWAGENS

Wie groß ist die ungefähre Entfernung von der Startlinie bis zum Beginn des längsten geradlinigen Abschnitts der Rennstrecke?

A 0,5 km
B 1,5 km
C 2,3 km
D 2,6 km

Frage 6: GESCHWINDIGKEIT EINES RENNWAGENS

Wo wurde während der zweiten Runde die geringste Geschwindigkeit gemessen?

A an der Startlinie
B bei etwa 0,8 km
C bei etwa 1,3 km
D nach der halben Runde