

Hagemeister, Volker

Die TIMSS-Leistungen bleiben zweifelhaft. Eine Antwort auf die Replik von Baumert u.a.

Die Deutsche Schule 92 (2000) 3, S. 322-328



Quellenangabe/ Reference:

Hagemeister, Volker: Die TIMSS-Leistungen bleiben zweifelhaft. Eine Antwort auf die Replik von Baumert u.a. - In: Die Deutsche Schule 92 (2000) 3, S. 322-328 - URN: urn:nbn:de:0111-pedocs-276163 - DOI: 10.25656/01:27616

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-276163>

<https://doi.org/10.25656/01:27616>

Nutzungsbedingungen

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

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

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

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

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

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Volker Hagemeister

Die TIMSS-Leistungen bleiben zweifelhaft

Eine Antwort auf die Replik von Baumert u. a.

In ausführlichen Darlegungen (in den Heften 1/2000 und 2/2000) haben Baumert, Klieme, Lehrke und Savelsbergh sich mit meiner Kritik an etlichen Aufgaben der TIMS-Studien (vgl. Heft 2/99) auseinandergesetzt. Sie glauben, meine Einwände im Wesentlichen und weitgehend zurückweisen zu können. Wenn ich mich im Folgenden noch einmal in dieser Sache zu Wort melde, so geschieht dies nicht aus schierer Rechthaberei, sondern es scheint mir um der Sache willen und angesichts der nachfolgenden Leistungsstudien (PISA) wichtig, dass an diesen Beispielen einige grundlegende Fragen der empirischen Leistungs- und Qualitätsprüfung geklärt und deren Lösungen ggf. verbessert werden.

Bedauerlicherweise arbeiten Baumert u.a. im Wesentlichen mit der Technik, Aussagen widerlegen zu wollen, die ich gar nicht gemacht habe. Weil darin offenbar Missverständnisse zum Ausdruck kommen bzw. in die Welt gesetzt werden könnten, möchte ich auf einige solcher Vorwürfe kurz eingehen.:

Würden bei Kontrolle der Deutscheleistungen die Unterschiede zwischen Hauptschülern und Gymnasiasten verschwinden (vgl. Baumert u.a. 2000, S. 114)?

An keiner Stelle ist bei mir die Meinung nachzulesen, bei TIMSS würden die Unterschiede zwischen Gymnasiasten und Hauptschülern oder Gesamtschülern verschwinden, wenn man die Deutscheleistungen kontrolliert (d.h., nur die Testergebnisse von Schülern vergleicht, die in Deutsch z. B. die gleiche Note haben). Eine solche Behauptung würde ich schon deshalb nicht aufstellen, weil ich ja der Meinung bin, dass ein erheblicher Teil der Science-Items fachliche Fehler enthält oder aus pädagogischer Sicht nicht sinnvoll ist.

Wurden die deutschen Schüler vor allem durch Multiple-Choice-Aufgaben benachteiligt (vgl. Baumert u.a. 2000, S. 108)?

Nirgendwo äußere ich die Vermutung, dass die deutschen Schüler mit den freien Aufgaben bei TIMSS leichter zu Rande gekommen seien als mit den Multiple-Choice-Aufgaben. Diese Vermutung unterstellen mir Baumert und Klieme, weil ich zu einigen Multiple-Choice-Items schildere, welche Probleme meine Interviewpartner aus 8. und 9. Klassen mit diesen MC-Aufgaben hatten. – Ich hatte nur MC-Aufgaben besprochen, weil sich bei den ausgewählten Items die fachlichen und konzeptionellen Fehler besonders gut in kurzer und allgemeinverständlicher Form darstellen ließen. Es kann jedoch keine Rede davon sein, dass ich die freien Aufgaben bei TIMSS für unproblematisch halte. Als Beispiel sei hier das TIMSS-Item K-10 angeführt ((IEA, The Hague, 1994):

„Luft ist farb-, geruch- und geschmacklos. Beschreibe eine Möglichkeit, das Vorhandensein von Luft nachzuweisen!“

Dieses Item kann man aus fachlicher und aus pädagogischer Sicht für gut geeignet halten, um zu prüfen, ob 8-Klässler Einblick in naturwissenschaftliche Arbeitsmethoden gewonnen haben. Bezogen auf die deutschen Schüler muss man jedoch einwenden, dass sie es nicht gewohnt sind, eine solche Frage in 77 Sekunden zu beantworten (bei TIMSS hatte man für 70 Aufgaben 90 Minuten Zeit).

Würde dagegen in einer der bei uns üblichen schriftlichen Lernerfolgskontrollen eine Aufgabe vom Typ K-10 vorkommen, so würde man den Schülern mindestens 5, eher 10 bis 15 Minuten Zeit einräumen, um über eine Antwort nachzudenken und um sie dann auch noch gut formuliert niederzuschreiben. Deshalb wird ein Schüler, der sich bei TIMSS der Aufgabe K-10 in gewohnter Weise zugewendet hatte, in Zeitnot gekommen sein.

Soll man Schüler mit sprachlichen Anforderungen verschonen (vgl. Baumert u.a., S. 115)?

Noch nie habe ich irgendwo gesagt, man solle Schüler im Unterricht mit sprachlichen Anforderungen verschonen. Im Gegenteil, auch in meinem Aufsatz verrete ich die Auffassung, dass im Physikunterricht „gefordert“ werden müsse, dass Schüler „Beobachtungen und Gedanken in eigenen Worten beschreiben“ (S. 174). Trotzdem bin ich der Meinung, dass das Science-Testpaket bei TIMSS zu viel Text enthält. Ein typisches Beispiel ist das TIMSS-Item N-5, wo nach den Ursachen des „Sauren Regens“ gefragt wird:

TIMSS-Aufgabe N-5¹

Eine der Hauptursachen des sauren Regens

- A. ist Säure aus Chemiewerken, die als Abfall in die Flüsse gelangt.
- B. ist Säure aus Chemielabors, die verdunstet.
- C. sind Gase aus der Kohle- und Ölverbrennung, die sich im Wasser der Atmosphäre lösen.
- D. sind Gase aus Klimaanlage und Kühlschränken, die in die Atmosphäre entweichen.

Nach Berliner Chemie-Rahmenplan wird die Entstehung des Sauren Regens in Klasse 9 im Abschnitt „Schwefel- und Stickstoffverbindungen“ angesprochen. Für einen 8-Klässler ist dieses Item eine reine Lese- und Vokabel-Aufgabe. Er müsste sich daran erinnern, dass die Alternative „C“ dem Satz über den „Sauren Regen“, den er im Unterricht irgendwann gehört hat, entspricht, ohne dass er dafür eine Begründung liefern könnte.

So verfehlt die Wort-Akrobatik des Items N-5 für einen Test ist, den Schüler aus 8. Klassen bearbeiten sollen, so unbefriedigend ist ferner an dem Science-Testpaket, dass viele wichtige Gebiete der Physik und Chemie nur mit Textaufgaben abgedeckt werden (zum Thema „Körperfarben“ gibt es nur das Text-

¹ TIMSS, Population 2, Science Content, Grouping I, © IEA, The Hague, 1994

Item R-2, zur Schallausbreitung nur die „Weltraumaufgaben“ L-7 oder zum Thema „Ausdehnung bei Erwärmung lediglich die Text-Aufgabe I-16 usw.).

Es besteht gar kein Dissens darüber, dass bei uns im Physik- und Chemie-Unterricht die „Verschriftlichung komplexer Gedankengänge“ vernachlässigt wird. Deshalb bin ich ja der Meinung, dass die textlastigen Testpakete der TIMS-Studie für den Einsatz in Deutschland schlecht geeignet waren, weil sie dem bei uns üblichen Unterricht nicht entsprochen haben. – Aber auch, wenn bei uns im Physikunterricht „Scientific Literacy“ nicht mehr vernachlässigt würde und wenn über Bedeutung und Grenzen naturwissenschaftlichen Arbeitens gemeinsam reflektiert und diskutiert würde, so dürfte ein Physik-Test trotzdem nicht nur Textaufgaben enthalten. Denn „Forschungsergebnisse deuten darauf hin, dass die Papier-und-Bleistift-Methode“, die in Tests aus Kostengründen überall angewendet wird, mit der „direkten Beobachtung“ von Schülern beim Experimentieren höchstens 10% an gemeinsamer Varianz erfasst (vgl. Shavelson/Ruiz-Rrimo 1999).

Nun könnte man ja der Meinung sein, dass ein Test gerade dann viele Textaufgaben enthalten sollte, wenn die Arbeit mit Texten im Unterricht vernachlässigt wird. Die Idee mag sehr verlockend erscheinen, lang ersehnten Schulreformen mit Hilfe überregionaler Tests zum Durchbruch zu verhelfen. Man muss dann jedoch offen sagen, dass nun Schulreform von ganz weit oben betrieben wird durch demokratisch kaum kontrollierbare Experten-Gruppen, die die Tests entwickeln. Wie schwierig es ist, die Ziele und Ergebnisse dieser Experten-Arbeit öffentlich zur Diskussion zu stellen, zeigt das Beispiel TIMSS: Zunächst – vor dem Testeinsatz – werden die Aufgaben geheim gehalten und wer die Aufgaben nach dem Testeinsatz kritisiert, der wird zum natürlichen Feind der Testinstitute. Deshalb haben Baumert und Klieme sich über viele Seiten hinweg so große Mühe gegeben, mich als inkompetent abzuqualifizieren.

Lehrt Alltagserfahrung denn nicht, dass sich Schall durch Eisen oder durch einen straff gespannten Bindfaden ausbreiten kann?

Zur Weltraumaufgabe L-7 führen Baumert und Klieme (DDS 2000, Heft 2) etliche Belege dafür an, dass Schüler sehr wohl auf der Basis von Alltagserfahrungen zu der Erkenntnis gelangen könnten, dass sich Schall durch Materie ausbreiten kann (etwa in einer Eisenbahnschiene). – Dies ist aber von mir an keiner Stelle bezweifelt worden. Ich hatte vielmehr die Meinung vertreten, dass es keine Situationen im Alltag gibt, in denen ein 14-Jähriger die Erfahrung machen könnte, dass sich Schall im luftleeren Raum nicht ausbreiten kann. Licht beispielsweise kann sich ja auch sowohl in Materie wie auch im Weltall ausbreiten. Warum sollte dies beim Schall nicht ebenfalls möglich sein? Außerdem machen Schüler heutzutage keineswegs nur Alltagserfahrungen, die zeigen, dass Materie Schallausbreitung begünstigt. Verwirrende Auswirkungen auf die Entwicklung eines tragfähigen Konzeptes zur Schallausbreitung hat hier die im Zeitalter leistungsstarker HiFi-Anlagen wichtige „Schalldämmung“, die ja durch dicke Material-Schichten bewirkt wird. Im Alltag von 14-Jährigen wird also die Schallausbreitung (ähnlich wie die Lichtausbreitung) durch Materie mal behindert, mal befördert.

In Heft 2/99 hatte ich angesprochen, dass das *Experiment mit der Klingel, die unter einer Glasglocke liegt*, sehr hilfreich ist, wenn man zu der Erkenntnis

hinführen will, dass Schallausbreitung an Materie gebunden ist. Um diese Position abzuwerten, zitieren Baumert und Klieme aus „einem kürzlich veröffentlichten Survey über amerikanische Physikbücher“. In der Zeitschrift *The Physics Teacher* wird ausgeführt, dass es falsch sei zu sagen, der Schall würde in dem luftverdünnten Raum, der unter der Glasglocke hergestellt wird, nicht weitergeleitet. „However, it isn't a matter of sound not travelling in a low-pressure region.“ (*The Physics Teacher*, 1999, S. 299)

Was Baumert und Klieme im Eifer des Gefechts übersehen haben, ist, dass die Position, die sie aus „*The Physics Teacher*“ zitieren, sehr gut zu meiner Kritik an dem TIMSS-Weltraum-Item L-7 passt: Dass nämlich Schüler mit diesem Item dann Schwierigkeiten haben könnten, wenn im Unterricht besprochen wurde, dass „*Vakuum*“ eine Fiktion ist. Sie würden dann möglicherweise zu der (zutreffenden) Einsicht gelangen, dass Antwort D „es gibt keine Luft im Weltraum in der sich der Schall fortbewegen kann“ bei der Raumschiffaufgabe nicht korrekt formuliert ist. Entsprechend würde ich z. B. auch bei dem Experiment mit Klingel und Glasglocke nie davon sprechen, dass unter der Glasglocke „*Vakuum*“ herrsche, wo sich Schall nicht mehr ausbreiten könne. Denn auch nach langem Pumpen wird nur ein luftverdünnter Raum geschaffen. Irgendwann nimmt der Luftdruck nicht weiter ab. Wie tief der Luftdruck unter der Glasglocke sinkt, hängt vom Pumpentyp ab (z. B. vom Zustand der Dichtungen).

Bei diesem Experiment kann man, sofern es real durchgeführt und nicht nur als Gedankenexperiment besprochen wird, sehr schön zeigen, *dass man sich vor voreiligen Schlussfolgerungen hüten muss*: Wenn die Klingel nach einiger Zeit des Abpumpens im Klassenraum nicht mehr zu hören ist, dann sagen Schüler z. B.: „Jetzt ist keine Luft mehr da. Deshalb hört man nichts mehr.“ – Lässt man nun die Schüler unmittelbar an der Glasglocke lauschen, dann hören sie in der Regel noch ein leises Klingeln. Dies führt dann z. B. zu der Hypothese, dass auch in stark verdünnter Luft noch Schallausbreitung stattfindet (wie in „*The Physics Teacher*“ 1999 auf S. 299 dargestellt).

Das Experiment mit Glasglocke und Klingel ist nicht deshalb schlecht, weil es Menschen gibt, die es falsch interpretieren. Es wäre traurig um den Physikunterricht bestellt, wenn man alle Experimente wegließe, die von irgend jemandem falsch durchgeführt oder ausgewertet werden könnten. Vermutlich dürfte man dann gar nicht mehr Experimentieren.

Welche Funktion hat das Zitat aus „*The Physics Teacher*“ bei Baumert und Klieme? – Damit wird mein Eintreten für das Klingel-Experiment *pauschal abqualifiziert*, so als ob dieses Experiment immer nur falsch interpretiert werden könnte.

Darf man das englische „simple“ mit simpel übersetzen?

Baumert u. a. schreiben dazu: „Hagemeister verwechselt aufgrund eines Übersetzungsfehlers Testdimensionen und Aufgabenschwierigkeit... Wer simple information in Abgrenzung zu complex information mit „simpler Information“ übersetzt ... zeigt, dass er die Grundstruktur des gesamten Klassifikationssystems der TIMSS-Items nicht verstanden hat... Einzelwissen (simple information) und Zusammenhangwissen (complex information) haben theoretisch nichts mit dem Schwierigkeitsgrad eines Items zu tun...“ (Baumert u. a. 2000, S. 104).

In meinem „Duden Oxford Englisch. Großwörterbuch“ (1990) steht bei dem englischen Wort „simple“: einfach, schlicht, kinderleicht, simpel. „Simple“ mit simpel zu übersetzen, wie ich es getan habe, entspricht also durchaus dem Üblichen.

Aber hatten Baumert und Lehmann, als sie 1997 die „Deskriptiven Befunde“ zu TIMSS veröffentlichten, das „Klassifikationssystem der TIMSS-Items“ womöglich auch noch „nicht verstanden“? Denn im Jahre 1997 wird bei Baumert/Lehmann (1997, S. 49) „Understanding simple Information“ noch mit „Verstehen einfacher Informationen“ übersetzt. Erst drei Jahre später ersetzen Baumert und Klieme (2000, S. 105) in derselben Tabelle „Verstehen einfacher Information“ durch den Begriff „Einzelwissen“.

Diese Veränderung in der Beschriftung der Tabelle ist aber sicherlich nicht nur vollzogen worden, um über mich sagen zu können, ich hätte „aufgrund eines Übersetzungsfehlers Testdimension und Aufgabenschwierigkeit“ verwechselt. Ich vermute, dass die Testdimension „Einzelwissen“ vor allem deshalb nachträglich eingeführt wurde, weil die Auswertung der Testergebnisse gezeigt hat, dass *etliche Items, die vorab von den Testkonstruktoren unter die Kategorie „Understanding simple Information“ eingeordnet worden waren, von nur wenigen Schülern gelöst worden sind, formal also einen hohen Schwierigkeitsindex haben.* So hat man nun „Einzelwissen“ über die erste Zahlenkolonne der Tabelle „Naturwissenschaftliche Testaufgaben nach Sachgebiet ...“ geschrieben, damit hier Aufgaben mit hohem Schwierigkeitsindex nicht als störend auffallen.

Dass die amerikanischen Testkonstruktoren allerdings bei „Understanding Simple Information“ ursprünglich nicht etwa an anspruchsvolles „Einzelwissen“, sondern eher an das gedacht haben, was wir „Faktenwissen“ nennen würden, zeigen die Items Q-16 und O-12:

TIMSS-Aufgabe Q-16²

Wie lange braucht das Licht vom nächstgelegenen Stern (außer unserer Sonne), um zur Erde zu gelangen?

- | | |
|--------------------------|----------------------|
| A. Weniger als 1 Sekunde | C. Ungefähr 1 Monat |
| B. Ungefähr 1 Stunde | D. Ungefähr 4 Jahre. |

Auch dieses Item haben die Testautoren der TIMS-Studie unter „Understanding Simple Information“ eingeordnet. In der Tat geht es hier um simples Faktenwissen. Ein Schüler, der nicht auswendig weiß, dass der nächste Fixstern etwa 4 Lichtjahre von der Erde entfernt ist, kann keinen Algorithmus aktivieren, um die richtige Zahl zu berechnen, denn dazu müßte er unter anderem wissen, wie groß die Entfernung zum nächsten Fixstern in Kilometern ist, was nicht nur 8-Klässlern unbekannt ist.

Da vielen Schülern das Vokabelwissen „4 Jahre“ für Aufgabe Q-16 nicht präsent war, haben nur relativ wenige bei diesem Item „D“ angekreuzt.

2 TIMSS, Population 2 Science Content Grouping B, © IEA, The Hague, 1994

Um das Abfragen einer Vokabel geht es auch bei der *TIMSS-Aufgabe O-12*:³

Luft besteht aus mehreren Gasen. Welches Gas ist am meisten enthalten?

- | | |
|---------------|-----------------|
| A. Stickstoff | C. Kohlendioxid |
| B. Sauerstoff | D. Wasserstoff |

Auch bei diesem Item wird Faktenwissen abgefragt. Man muss wissen, dass die richtige Vokabel „Stickstoff“ heißt.

Die vorne wiedergegebene *Aufgabe N-5*, wo nach den Ursachen des „Saurer Regens“ gefragt wird, gehört nach Meinung der Testautoren ebenfalls in die Kategorie „Understanding Simple Information“. – Übersetzt man nun, wie jetzt von Baumert und Klieme gefordert, „Simple Information“ mit „Einzelwissen“, dann wird man damit dem Item N-5 keineswegs besser gerecht, denn das Thema „Saurer Regen“ lässt sich nicht zu „Einzelwissen“ linearisieren. Welches „Einzelwissen“ sollte ein Schüler denn aktivieren, um bei den verschiedenen Alternativen des Items N-5 begründen zu können, was falsch und was richtig ist?

Richtig gelöst wurden von den weltweit getesteten 8-Klässlern das *Fixstern-Item Q-16* und das *Luft-Item O-12* aus der Kategorie „Understanding Simple Information“ nur von ca. 25%, wohingegen z. B. das *Stromkreis-Item M-12* aus der Kategorie „Theorizing, Analysing and Solving Problems“ von ca. 50% und das *Trägheits-Item O-13* aus der Kategorie „Understanding Complex Information“ sogar von ca. 56% „richtig“ gelöst wurden⁴.

Wenn nun für das Luft-Item O-12 oder für die Fixstern-Aufgabe Q-16 relativ hohe „Schwierigkeits-Indizes“ bei der Datenanalyse errechnet worden sind, so folgt daraus nicht, dass für diese Items intellektuell anspruchsvolles „Einzelwissen“ benötigt wurde, sondern dass das Faktenwissen, das für die Items O-12 und Q-16 präsent sein musste, *in vielen Ländern im Unterricht nicht angesprochen* worden ist.

Mit dem jetzt präsentierten „Klassifikationssystem der TIMSS-Items“ (Baumert u.a. 2000), wird das Dilemma nicht aus der Welt geschafft, dass Aufgaben der Kategorie „Understanding simple Information“, wie die Items O-12 und Q-16, weltweit nur von relativ wenigen 8-Klässlern richtig gelöst wurden. Die Items O-12 und Q-16 sind nicht intellektuell besonders anspruchsvoll, sondern es handelt sich um Items von *weltweit schlechter Curriculums-Validität*.

Resümee

Die Texte von Baumert und Klieme zeigen, dass ein wissenschaftlicher Diskurs über die Brauchbarkeit eines Tests kaum gelingt, wenn durch den Test Staaten, Schultypen, Lehrer oder Schüler in Rangreihen gebracht werden. Dann darf (außer einigen „Experten“) niemand die Items vor ihrem Einsatz kennen und nach ihrem Einsatz können die Testaufgaben nicht mehr zur Disposition gestellt werden, denn sonst wären ja Entwicklung und Einsatz der Tests umsonst gewesen. Wer die Tests der TIMS-Studie nachträglich kritisiert, muss mit

3 TIMSS, Population 2 Science Content Grouping A, © IEA, The Hague, 1994
4 <http://www.csteep.bc.edu/TIMSS1/Items.htm>.

Vehemenz in die Schranken verwiesen werden, so dass sich niemand mehr traut, Kritik an TIMSS vorzutragen.

Ich möchte hier noch einmal dafür plädieren, *Testentwicklung nur in enger Kooperation mit den betroffenen Lehrern zu betreiben*. Darin stimme ich den Forderungen zu, die Schlömerkemper (2000, S. 8) in seinem Plädoyer für eine bessere „Kommunikationskultur“ erhebt. Tests, die sorgfältig und in kleinen Schritten unter Mitwirkung der betroffenen Lehrerschaft entwickelt worden sind, können für Diagnosezwecke von großem Wert sein. Andererseits ist es falsch anzunehmen, dass durch den überregionalen Einsatz standardisierter Tests Schüler-Zeugnisse gerechter werden, denn die unterschiedlichen Leistungen der Lehrer und die ungleichen sozialen Rahmenbedingungen gehen in nicht abtrennbarer Weise in alle Testergebnisse ein (vgl. Hagemeister 2000).

Literatur

Baumert, Jürgen, Rainer Lehmann u.a. 1997: TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske+Budrich

Baumert, Jürgen; Eckhard Klieme, Manfred Lehrke und Elwin Savelsbergh 2000: Konzeption und Aussagekraft der TIMSS-Leistungstests. Zur Diskussion um TIMSS-Aufgaben aus der Mittelstufenphysik. In: Die Deutsche Schule. 92, 2000, 1, (Abschnitt 1.1 „Testkonzeption“)

Hagemeister, Volker 1999: Was wurde bei TIMSS erhoben? In: Die Deutsche Schule. 91, 1999, 2, S. 160-177

Hagemeister, Volker 2000: Irrwege und Wege zur Testkultur. In: Die Deutsche Schule. 92, 2000, 1, S. 87-101

Schlömerkemper, Jörg 2000: Konsens und Beteiligung. Ein Plädoyer für mehr Demokratie in der Bildungspolitik. In: Die Deutsche Schule. 92, 2000, 1, S. 6-9

Shavelson, Richard J.; Maria A. Ruiz-Rrimo 1999: Leistungsbewertung im naturwissenschaftlichen Unterricht. In: Unterrichtswissenschaft. 27, 1999, 2, S. 102-127

Volker Hagemeister, geb. 1940, Dr., Studium der theoretischen Physik; Mathematik- und Physiklehrer; seit 1977 Wiss. Dir. am Pädagogischen Zentrum bzw. Berliner Institut für Lehrerfort- und -weiterbildung und Schulentwicklung;
Anschrift: Alte Jakobstr. 12, 10969 Berlin;
E-mail: verwaltung@bil.verwalt-berlin.de