

Brügelmann, Hans

"Erster, Zweiter, Dritter – Letzter"

Grundschule aktuell : Zeitschrift des Grundschulverbandes (2017) 137, S. 36-39



Quellenangabe/ Citation:

Brügelmann, Hans: "Erster, Zweiter, Dritter – Letzter" - In: Grundschule aktuell : Zeitschrift des Grundschulverbandes (2017) 137, S. 36-39 - URN: urn:nbn:de:0111-pedocs-234745 - DOI: 10.25656/01:23474

<http://nbn-resolving.org/urn:nbn:de:0111-pedocs-234745>

<http://dx.doi.org/10.25656/01:23474>

in Kooperation mit / in cooperation with:



www.grundschulverband.de

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.

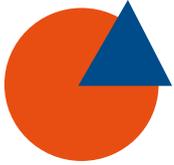
By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

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

Mitglied der





Grundschule **aktuell**

Zeitschrift des Grundschulverbandes · Heft 137



Sprache **Bildung** Sprachbildung



Tagebuch

- S. 2 Kennen Grundschul Kinder ihre Rechte? (U. Carle)

Thema: Sprache • Bildung • Sprachbildung

- S. 3 »Sprachen lernt man, indem man sie spricht«
(R. Tracy/H. Brügelmann)
- S. 6 Bildungssprache – auch im Fachunterricht
(M. Gutzmann)
- S. 9 Schreiben mit der Laut-Tabelle – auch für mehrsprachige Kinder? (E. Brinkmann)
- S. 14 Mehrsprachiges Lernen
(R. Benati/M. J. Sánchez Oroquieta)
- S. 17 Kinder testen oder Kinder stärken? (H. Bartnitzky)

Praxis: Sprachbildung – Bildungssprache

- S. 21 Eine Sache zur Sprache bringen (M. Gutzmann)
- S. 25 Fachbezogene Sprachförderung im Mathematikunterricht (L. Verboom)
- S. 29 Sprachliches Lernen im Kunstunterricht (R. Pols)
- S. 33 Eine fremde Sprache kennen lernen (R. Urbanek)

Aus der Forschung

- S. 36 IQB, TIMSS, PISA u. a. (H. Brügelmann)

Rundschau

- S. 40 Vergleichsarbeiten 2017 (H. Brügelmann)
- S. 42 Comenius-Netzwerk Projekt »TdiverS«
(K. Merz-Atalik/K. Weber)

Landesgruppen aktuell – u. a.:

- S. 44 Baden-Württemberg: »Endlich wieder überall Fehler anstreichen«
- S. 46 Brandenburg: Ein inklusives Theaterprojekt
- S. 48 Nordrhein-Westfalen: Grundschultag
- S. 49 Thüringen: Qualitätskriterium Zusammenarbeit

www. ► grundschule-aktuell.info

Hier finden Sie Informationen zu »Grundschule aktuell« sowie Zusatzmaterialien zu den Beiträgen in der Print-Ausgabe der Zeitschrift des Grundschulverbandes.

► Herausgeber und Redaktion respektieren die Vielfalt geschlechtlicher Identitäten. Manche Autorinnen und Autoren bringen dieses Anliegen durch besondere schriftsprachliche Zeichen zum Ausdruck. Eine allgemein anerkannte Lösung für das Problem »gendersensibler« (Schrift-)Sprache gibt es zurzeit nicht. Daher gilt für diese Zeitschrift: Jede Autorin/jeder Autor verwendet in ihrem/seinem Text ihre oder seine bevorzugte Form.

»Sprachliche Kompetenzen fördern« war das Thema des mit Begeisterung aufgenommenen Vortrags von Prof. Rosemarie Tracy bei der Herbsttagung des Grundschulverbandes im November in Kassel. Über Mehrsprachigkeit und ihr Konzept »alltagsintegrierter Sprachförderung« sprach Hans Brügelmann mit Rosemarie Tracy. ► S. 3

Bildungssprache – auch im Fachunterricht: Sprachbildung ist eine Querschnittsaufgabe aller Lernbereiche. Wie bildungssprachliche Handlungskompetenz entstehen kann, beschreibt Marion Gutzmann. ► S. 6



Schreiben mit der Anlaut-Tabelle ist – ausgehend von Baden-Württemberg – wieder in die öffentliche Debatte gekommen. Über Sinn und Möglichkeiten besonders auch für mehrsprachige Kinder schreibt Erika Brinkmann. ► S. 9

Zwei Beiträge runden die Rubrik »Thema« ab: Rosella Benati und María José Sánchez Oroquieta berichten über Kölner Erfahrungen mit »Mehrsprachigem Lernen« ► S. 14 und Horst Bartnitzky fragt, ob die überprüfbaren Kompetenzen am Ende das Maß für Schulqualität sind: »Kinder testen oder Kinder stärken?« ► S. 17

Impressum

GRUNDSCHULE AKTUELL, die Zeitschrift des Grundschulverbandes, erscheint vierteljährlich und wird allen Mitgliedern zugestellt.

Der Bezugspreis ist im Mitgliedsbeitrag enthalten. Das einzelne Heft kostet 9,00 € (inkl. Versand innerhalb Deutschlands); für Mitglieder und ab 10 Exemplaren 5,00 €.

Verlag: Grundschulverband e.V., Niddastraße 52, 60329 Frankfurt/Main, Tel. 0 69/77 60 06, Fax: 0 69/7 07 47 80, www.grundschulverband.de, info@grundschulverband.de

Herausgeber: Der Vorstand des Grundschulverbandes

Redaktion: Ulrich Hecker, Hülsdonker Str. 64, 47441 Moers, Tel. 0 28 41 / 2 17 14, ulrich.hecker@gmail.com

Fotos: Allegro-Grundschule Berlin (Titel), Laborschule Bielefeld (Titel), B. Butzke (S. 17), Luisa Greco (S. 17, 19); Autorinnen und Autoren, soweit nicht anders vermerkt

Herstellung: novuprint, Tel. 0511 / 9 61 69-11, info@novuprint.de

Anzeigen: Grundschulverband, Tel. 0 69/77 60 06, info@grundschulverband.de

Druck: Beltz Bad Langensalza, 99974 Bad Langensalza

ISSN 1860-8604 / Bestellnummer: 6077

Beilage: Grundschule aktuell SPEZIAL »Alle Kinder mitnehmen« zur Verleihung des Erwin-Schwartz-Grundschulpreises 2016 an Annemarie von der Groeben

Hans Brügelmann

»Erster, Zweiter, Dritter – Letzter«

Ende 2016 haben drei Schulleistungsstudien wieder mal für Aufregung in den Medien gesorgt. Was sagen uns IQB-Bildungstrend, TIMSS, PISA und andere Ländervergleiche?

Als erstes kam Baden-Württemberg unter die Räder. Im »IQB-Bildungstrend 2015« (Stanat u. a. 2016) wurden vor allem die Deutsch-Leistungen von 9.-Klässler*innen in den verschiedenen Bundesländern verglichen. Zu ihrer Erhebung wurden Tests in den Bereichen »Lesen«, »Zuhören« und »Orthographie« eingesetzt. Beim Vergleich der Ergebnisse wurde Baden-Württemberg unter den 16 Bundesländern auf den Rängen 10, 14 und 12 eingestuft. Nach Plätzen im oberen Bereich bei früheren Erhebungen war die Aufregung groß. Wie der Bildungsforscher Ulrich Trautwein im SPIEGEL (Nr. 44/2016), so malte auch Heike Schmoll in der FAZ v. 2.11.2016 die Schulentwicklung in Baden-Württemberg aufgrund der Daten des Ländervergleichs in düsteren Farben. Zu Recht – oder mediale Brandstiftung?

Erklärungen für den angeblichen »Leistungsabfall« hatten die Kritiker*innen auch gleich zur Hand: Abschaffung der verbindlichen Schulempfehlung nach Klasse 4, Einführung der Gemeinschaftsschulen, Schreiben mit der Anlauttabelle im Anfangsunterricht. Und das, obwohl die Autor*innen (a. a. O., S. 185) selbst ausdrücklich anmerken, »... dass die Gemeinschaftsschule als neue Schulart ... noch nicht berücksichtigt werden« konnte. Und zur Grundschule sind die 9.-Klässler*innen noch zu Zeiten der CDU/FDP-Koalition gegangen. Insofern läuft der Versuch, aus dem IQB-Trend 2015 ein Grün-Rot-Bashing zu konstruieren, leer.

Im Übrigen lohnt es, sich die Ergebnisse genauer anzuschauen. Denn was bedeuten die unterschiedlichen Rangplätze konkret?

Im Lesen erreicht Baden-Württemberg den zehnten Platz mit 496 Punkten, Schleswig-Holstein auf dem zweiten Platz erreicht 514 (Stanat u. a. 2016, S. 337). Sind diese 18 Punkte (oder rund

3 Prozent) Unterschied eine inhaltlich bedeutsame Differenz? Wir wissen es nicht, denn die Forscher*innen teilen nicht mit, welchen Realunterschieden (gelesene Textmenge pro Minute, Anteil der falsch gelösten Aufgaben) diese Werte entsprechen. Das gilt auch für die Differenz zu den baden-württembergischen Ergebnissen von 2009, als das Land 521 Punkte erreichte.

Kasten 1: Medaillenverteilung nach hundertstel Sekunden

Bei den alpinen Ski-Weltmeisterschaften 2015 gewann der Schweizer Patrick Küng die Abfahrt. Er brauchte für die Strecke in Beaver Creek 1 Minute, 43 Sekunden und 18 Hundertstel. Elfter wurde sein Landsmann Didier Défago. Dank der eingesetzten Präzisionsgeräte konnte die elektronische Zeitmessung einen Rückstand von 71 Hundertstel feststellen. Nicht mal eine Sekunde langsamer – bezogen auf eine Gesamtzeit von über 100 Sekunden. Der Dritte brauchte sogar nur sieben Hundertstel mehr als der Zweite. Das sind gerade mal 0,06 Prozent mehr – für die »Ski-Kompetenz« im Alltag belanglos. Bedeutsam ist der Unterschied nur für Wettbewerbe unter einzelnen Spitzenfahrern. Merke: Mit Präzisionsverfahren messbare Unterschiede sagen noch nichts über ihre alltagspraktische Bedeutung aus.

bergischen Ergebnissen von 2009, als das Land 521 Punkte erreichte.

Zwar übersetzen die Autor*innen die Punktdifferenzen in zeitliche Unterschiede und sagen zum Beispiel (a. a. O., S. 536), in dieser Untersuchung entsprechen 20 Punkte ungefähr dem Lernfortschritt von einem Schuljahr. Aber ist der in diesem Alter wirklich bedeutsamer, als die in Kasten 1 berichteten Unterschiede in der Ski-Abfahrt? Lesen durchschnittliche 9.-Klässler*innen in derselben Zeit 2 Prozent oder 20 Prozent mehr Text als 8.-Klässler*innen? Und lösen sie mit 15 Jahren nur 4 Pro-

zent der Aufgaben falsch, mit 14 aber noch 10 Prozent oder auch nur 5 Prozent? Die Bedeutung der Punktunterschiede für Leistungen im Alltag bleibt also offen.

Und noch eine zweite Einschränkung: Die Ergebnisse stammen aus Stichproben, mithilfe derer die Verhältnisse in der Grundgesamtheit lediglich geschätzt werden. Eine solche Schätzung geht immer mit einem gewissen Schätzfehler einher, der es streng genommen nur zulässt, einen Wertebereich anzugeben, das sogenannte Vertrauens- (»Konfidenz-«) Intervall, in dem sich der tatsächliche Wert in der Grundgesamtheit mit einer bestimmten Wahrscheinlichkeit befindet. In der Rechtschreibung beispielsweise liegt der tatsächliche Wert für die Grundgesamtheit aller baden-württembergischen 9.-Klässlerinnen auf Platz 12 mit 95%iger Wahrscheinlichkeit irgendwo innerhalb des Vertrauensintervalls von 491 bis 506, für die Zweitplatzierten aus Sachsen irgendwo zwischen 500 und 514 (a. a. O., S. 340). Die Überschneidungen der beiden Vertrauensintervalle zeigen, wie unsicher die berichteten Differenzen und damit die zugewiesenen Plätze 2 und 12 sind, denn es könnte auch sein, dass der tatsächliche Wert in der sächsischen Grundgesamtheit 501 beträgt und in der baden-württembergischen 505.

Veränderungen der Rangplätze von 2009 nach 2015 sind ähnlich wenig verlässlich zu interpretieren. Zwar erreichte die baden-württembergische Stichprobe in der Orthographie 2009 noch 516 Punkte (a. a. O., S. 351). Aber die Vertrauensintervalle (mit immer noch 5 Prozent Fehlerwahrscheinlichkeit) für die tatsächlichen Werte (in der Grundgesamtheit) liegen für 2015 bei 499 bis 513 und für 2009 bei 509 bis 523. Also überschneiden sich auch hier die Vertrauensintervalle. Zudem: Rangplätze sind relative Bewertungen. So sinken sie auch ohne Verschlechterung der Leistungen, wenn andere Länder gewonnen haben.

Umso verwunderlicher das Mediengetöse. Die Autor*innen stellen dagegen in der durchschnittlichen Rechtschreibkompetenz sachlich »... keine signifikanten Unterschiede zwischen den Jahren 2009 und 2015« fest (a. a. O., S. 351). Warum aber soll dann im Anfangsunterricht z.B. das Schreiben mit Anlauttabellen verboten werden?

Ganz zu schweigen davon, dass niemand weiß, welchen Anteil das lautorientierte Schreiben in baden-württembergischen Grundschulen tatsächlich hat (bundesweit lag z.B. Reichens Marktanteil nie über 1 Prozent). Und noch weniger ist bekannt, in welchen methodischen Kombinationen (etwa mit Grundwortschatzarbeit) es in der Regel auftritt und wie lange es durchschnittlich andauert (vgl. die Varianten in den Beiträgen zu Brinkmann 2015). Positive Effekte auf die spätere Rechtschreibung sind für das synthetisch-analytische Konstruieren von Wörtern in der Anfangsphase im deutschen wie im angelsächsischen Raum nachgewiesen (s. u. a. Richter 1992, S. 150ff.; National Early Literacy Panel 2008).

Wie oberflächlich empirische Befunde in vielen Medien dargestellt und kommentiert werden, konnte man auch nach der Veröffentlichung des internationalen Grundschulvergleichs in Mathematik und den Naturwissenschaften (TIMSS) beobachten. So war auf FAZ-online am 29.11.2016 in einer dpa-Meldung zu lesen: »Deutschlands Grundschüler haben anscheinend große Probleme mit Mathematik. In diesem wichtigen Unterrichtsfach sind sie laut der Bildungsstudie TIMSS im internationalen Vergleich mit 522 Punkten tief ins Mittelfeld gerutscht und liegen nun unterhalb des EU-Durchschnitts von 527 Punkten« (Hervorhebung brü).

Schaut man sich die Daten genauer an, stellt man fest, dass die deutschen Viertklässler*innen 2007 bei 525 Punkten lagen, auf diesen Basiswert bezogen also bis 2011 drei Punkte gewonnen und bis 2015 drei Punkte verloren haben. Demnach kein erkennbarer Trend, sondern Schwankungen – statistisch durchaus im Zufallsbereich für einen Trend. Zudem beziffern die Autor*innen des TIMSS-Berichts den Lernfortschritt für ein Schuljahr auf rund 30 Testpunkte; selbst sechs Punkte Differenz (2015 vs.

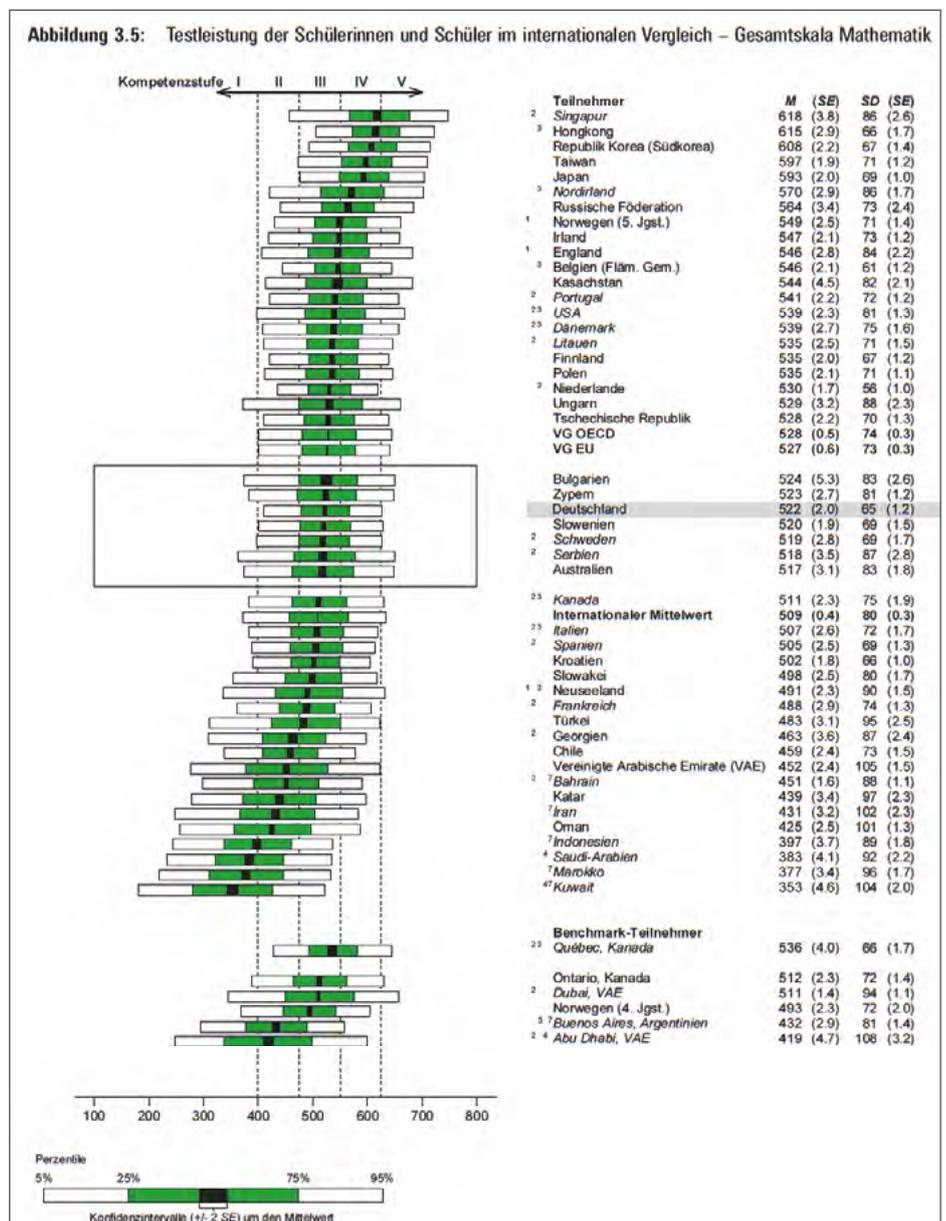
2011) entsprechen also gerade mal zwei Monaten. »Große Probleme«?

Entsprechend zu relativieren ist die Differenz von fünf Punkten zum europäischen Durchschnitt, der bei 527 Punkten liegt. Denn die Verteilungen der Leistungen in den einzelnen Ländern überlappen sich erheblich. Dass Deutschland nicht mehr über dem EU-Durchschnitt liegt, kann zudem nicht einfach als Folge schlechterer Leistungen in unserem Land interpretiert werden (»tief ... gerutscht«), sondern ist zumindest auch einer Verbesserung der Testergebnisse in anderen Ländern geschuldet: Der europäische Durchschnitt lag 2011 bei 519 Punkten.

Vor allem aber wurde in den Medien kaum zur Kenntnis genommen, dass

sich die deutsche Schülerpopulation von 2011 bis 2015 bedeutsam verändert hat, so dass Veränderungen nicht ohne Weiteres auf den Unterricht zurückgeführt werden können. In den Worten der Autor*innen des TIMSS-Berichts: »In Mathematik sind unter Berücksichtigung von Veränderungen in der Schülerschaft die durchschnittlichen Leistungen von dem Jahr 2007 zu dem Jahr 2011 statistisch signifikant um 11 Leistungspunkte gesunken und von TIMSS 2011 zu 2015 statistisch signifikant um 8 Punkte gestiegen. Damit wurde in TIMSS 2015 wieder das Leistungsniveau von TIMSS 2007 erreicht« (S. 375, Hervorhebung brü).

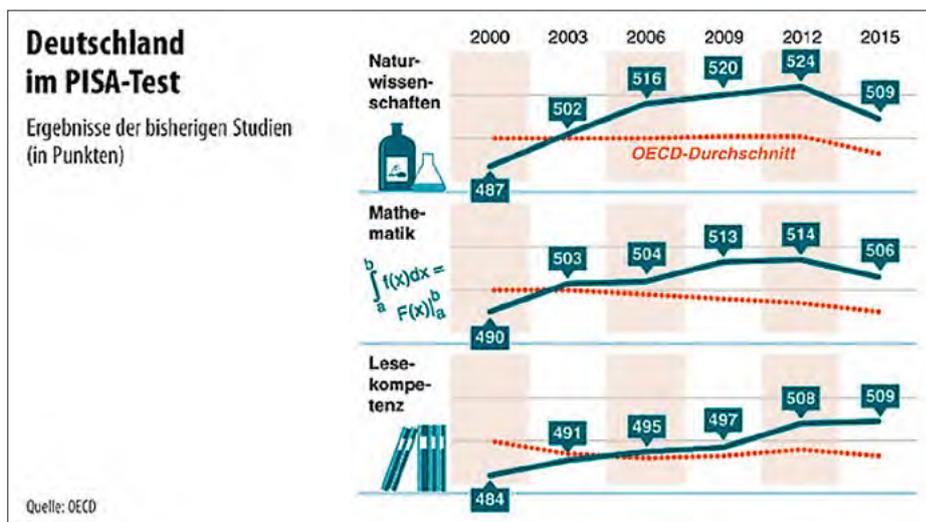
Das Bild ist also viel komplizierter, als in der öffentlichen Diskussion verhan-



Kasten 2: Mittelwertsunterschiede und Vertrauensintervalle: Überlappende Leistungsverteilungen bei TIMSS 2015 (aus Abb. 3.5 in Wendt u. a. 2016, S. 107)

delt. Und: Vorschnelle Ursachen- und Schulzuweisungen können zu fatalen Fehlschlüssen bei der Entscheidung für die erforderlichen Maßnahmen führen – wie sich schon oben bei den Fehlinterpretationen der IQB-Trendstudie gezeigt hat.

Und dann kam PISA – zum sechsten Mal seit 2000. Nach den Erfolgsmeldungen der letzten Runden, in denen die deutschen 15-Jährigen jeweils Punkte gewonnen und hier und da Plätze gut gemacht hatten, dieses Mal »Stabilisierung auf hohem Niveau« (so KMK-Präsidentin Claudia Bogedan): Im Lesen + 1 Punkt, in Mathematik – 8 und in den Naturwissenschaften – 15 Punkte. Die rücksichtsvolle Bewertung erstaunt, wurden doch in den Vorjahren schon Zuwächse von 4 oder 9 Punkten zu »Fortschritten« hochgejubelt und als »Beweis« für eine erfolgreiche Bildungspolitik verkauft: die Einführung von Bildungsstandards und regelmäßigen Tests habe die »Qualität« der Schulen gesteigert. Nun heißt es, solche Differenzen seien »statistisch nicht signifikant«.



Kasten 3: Unterschiede ohne Basisbezug (dpa/ FAZ-online v. 6.12.2016)

Das stimmt, aber dann muss man sich das Ganze doch etwas grundsätzlicher anschauen. Bekanntlich werden auch kleine Unterschiede »statistisch signifikant«, wenn die Stichproben nur groß genug sind. Bei PISA umfassen sie 5.000 bis 10.000 Schüler*innen pro Land. Wenn dann Unterschiede von 5 bis 10 Punkten (je nach Messfehler) oft »nicht

signifikant« werden, können sie auch inhaltlich nicht besonders bedeutsam sein. Bei PISA 2000 konnte schon die Lösung einer einzigen Aufgabe zusätzlich zu 20 Punkten Zugewinn führen. Die Umrechnung der Punkte auf eine 500er-Skala lässt selbst kleine Unterschiede groß erscheinen – visuell zusätzlich verzerrt, wenn in den Kurven die unteren

Kasten 4: Wann sind Unterschiede in Testergebnissen »bedeutsam« für Bildungspolitik und für den pädagogischen Alltag?

Rangplätze in einer Stichprobe sagen nicht viel über reale Abstände aus. Darum muss man die Unterschiede zwischen den Punktwerten in mehrfacher Hinsicht prüfen.

1. Ist angegeben, ob Punktdifferenzen zwischen Gruppen »statistisch signifikant« sind? Nur dann sind die zwischen den Stichproben gefundenen Unterschiede mit großer Wahrscheinlichkeit auch zwischen den entsprechenden Grundgesamtheiten (bspw. alle 15-jährigen Schüler*innen Deutschlands und Finnlands) zu finden. [Technische Anm.: Die realen Punktwerte liegen mit (immer noch 5%igem Fehlerrisiko) im Bereich von rund +/- zwei Standardfehlern um die Werte in den Stichproben (in dem »Vertrauensintervall«).] Die deutschen Schüler*innen könnten demnach bei einer Wiederholung von PISA statt auf Platz 10 (bezogen auf die OECD-Länder) wegen der Messfehler auch auf Platz 6 oder Platz 13 landen, in 5 Prozent der Fälle sogar außerhalb dieses Bereichs (und das auch, falls sich die Werte anderer Länder ebenfalls verändern). Schon die Plätze auf den Ranglisten sind also sehr unsicher ...

2. Sind Unterschiede statistisch signifikant, sind sie damit aber noch nicht inhaltlich bedeutsam. Denn bei gleichen Durchschnittswerten spielt es eine Rolle, wie breit die Einzelwerte innerhalb einer Gruppe um den Mittelwert streuen. Scharen sich die Werte in beiden Gruppen eng um den Durchschnitt, ist der Unterschied zwischen den Mittelwerten von zwei Gruppen trennschärfer, als wenn sich beide Verteilungen breit überlappen. [Technische. Anm.: Ein Maß für diese Trennschärfe ist die *Effektstärke* – meist berechnet als Differenz der beiden Mit-

telwerte geteilt durch die gemittelte Streuung in den beiden Gruppen (zur einfachen Berechnung aus den üblichen Tabellenwerten s. das Formular unter www.soerenwallrodt.de). Dieser Wert sollte mindestens 0.2 – 0.3 (Schwelle für einen sogar nur »schwachen Effekt«), besser sogar mehr als 0.5 (»mittlerer Effekt«) erreichen. Differenzen selbst von 10 Plätzen in den hier berichteten Studien erreichen meist nur Werte von 0.1 bis 0.2.]

3. Letztlich kommt es aber auf das *fachlich begründete Urteil* an, welche Bedeutung man Punktdifferenzen in persönlicher Verantwortung zuerkennt. Werden Originaldaten angegeben, kann man die Bedeutung von Unterschieden für den Alltag selbst abschätzen: Macht z.B. die eine Gruppe auf 1.000 Wörter 81 Rechtschreibfehler und die andere 79, spielt die Differenz von 2 Fehlern für den Alltag sicher keine große Rolle. Oft werden aber – wie bei PISA oder TIMSS – statt der Originaldaten nur umgerechnete Werte berichtet. Manchmal veranschaulichen die Autor*innen die Bedeutung von Unterschieden durch den Vergleich mit Lernfortschritten über die Zeit hinweg. Aber auch dann muss man aufpassen: Im Lesen verbessert sich die Leistung über ein Schuljahr hinweg auf Klassenstufe 1/2 erheblich, auf Klassenstufe 9/10 dagegen nur noch wenig. Geht es andererseits um neue Inhalte (z.B. in Mathematik), kann auch auf Klassenstufe 9/10 ein Schuljahr einen bedeutsamen Unterschied darstellen.

Mehr zu den technischen Details bei Lind (2016): www.uni-konstanz.de/ag-moral/pdf/Lind-2016_Effektstaerke-Vortrag.pdf

480 Punkte »abgeschnitten« werden (s. Kasten 3).

Und in der Tat erreichen selbst 10–20 Punkte auf der PISA-Skala in der Regel nicht die Schwelle auch nur »kleiner Effektstärken«, die für Mittelwertunterschiede bei $d = 0.2$ bis 0.3 angesetzt werden (s. Kasten 4).

Nachdenklich stimmt auch, dass Korea von 2012 bis 2015 in den Naturwissenschaften 30 Punkte verloren hat, die Schüler*innen also binnen drei (!) Jahren um ein ganzes Schuljahr schwächer geworden sein sollen. Schweden andererseits hat von 2003 bis 2012 insgesamt 31 Punkte verloren, dann aber bis 2015 schon wieder 16 Punkte aufgeholt, was angeblich einem halben Schuljahr entspricht. Werden in solchen Schwankungen tatsächlich reale Veränderungen abgebildet?

Es gibt auch noch andere Erklärungen. So mussten die Aufgaben bis 2012 mit Bleistift auf Papier bearbeitet werden – 2015 aber am Computer. In einem Vergleich beider Formen stellten Robitzsch u. a. (2016) fest, dass deutsche Schüler*innen bei denselben Aufgaben am PC deutlich (um 10 und mehr Punkte) schlechtere Ergebnisse erzielten als auf dem Papier. Dafür, dass dieser Medienwechsel einen gewichtigen Einfluss auf die Leistungsvergleiche hat, spricht auch ein Ergebnis aus Österreich: Dort haben sich im Lesen die Mädchen verschlechtert, die Jungen etwas verbessert (s. Nimmervoll 2016).

Literatur

Brinkmann, E. (Hrsg.) (2015): Rechtschreiben in der Diskussion – Schriftspracherwerb und Rechtschreibunterricht. Beiträge zur Reform der Grundschule, Bd. 140. Grundschulverband: Frankfurt.

Brügelmann, H. (2015): Vermessene Schulen – standardisierte Schüler. Zu Risiken und Nebenwirkungen von PISA, Hattie, VerA & Co. Beltz: Weinheim/ Basel.

Dammer, K.-H. (2015): Vermessene Bildungsforschung. Wissenschaftsgeschichtliche Hintergründe zu einem neoliberalen Herrschaftsinstrument. Schneider Hohengrehen: Baltmannsweiler.

Jahnke, T./Meyerhöfer, W. (Hrsg.) (2006): Pisa & Co. Kritik eines Programms. Hildesheim: Franzbecker

National Early Literacy Panel (Hrsg.) (2008): Developing early literacy: A scientific synthesis of early literacy development and implications for intervention. National Institute for Literacy & The Partnership for Reading. Jessup, Maryland.

Wenn aber schon der Wechsel der Aufgabenform solche Unterschiede ausmachen kann – was bedeutet das erst für die Aussagekraft der künstlichen Testsituation für Leistungen unter Alltagsbedingungen? Und welche Bedeutung hat der Zeitdruck für den Abruf vorhandener Kompetenzen, welche Rolle spielt dabei z. B. die unterschiedliche Textlänge der Aufgabe in verschiedenen Sprachen, wie stark überlagern zudem die Leseanforderungen der textgebundenen Aufgabenstellung die mathematischen und naturwissenschaftlichen Leistungen? Viele Fragezeichen.

So ist auch denkbar, dass die Leistungszuwächse vorher, also nach 2000, auf eine wachsende Vertrautheit der Schüler*innen mit den Aufgabenformaten und deren stärkere Nutzung im Unterricht zurückzuführen sind. Außerdem verändern sich die Maßstäbe für die Bewertung erreichter Punkte, nämlich die Mittelwerte von EU, OECD und internationaler Stichprobe, von Termin zu Termin, weil nicht immer dieselben Länder teilnehmen, so dass sich Rangplätze selbst bei gleich bleibenden Ergebnissen verändern können.

Das PISA-Ranking und das mit ihm verbundene Verständnis von Forschung ist demnach hoch problematisch (ausführlicher: Jahnke/Meyerhöfer 2006; Brügelmann 2015; Dammer 2015). Interessanter als die Ergebnisse solcher Studien sind Fragen, die sie aufwerfen: So erreichen bei PISA 2015 die deut-

schen Schüler*innen in den Naturwissenschaften – bezogen auf Unterrichts- und Hausaufgabenzeit – mehr Punkte als alle anderen Länder außer Finnland. Ein Indiz für besonders effektives Lernen, für besonders erfolgreiche Lehrer*innen? Und anders als erwartet sind sie im Fachwissen besonders gut – im methodischen Denken eher schwach. Letzteres beherrschen angeblich die asiatischen Schüler*innen besser, deren Schulen immer wieder reiner Drill vorgeworfen wird. An der Untersuchung solcher Überraschungen sollte Bildungsforschung ansetzen. Und vor allem sollte sie untersuchen, wie es manchen Schulen gelingt, selbst unter schwierigen Bedingungen erfolgreich zu arbeiten. Aber das erfordert einen anderen Stil von Forschung: Lernbiographien von Schüler*innen, Beobachtungen der Interaktionen im Unterricht aus verschiedenen Perspektiven, Dokumentation und Analyse von Umsetzungsvarianten desselben Programms/derselben Methode, Fallstudien der Entwicklung von Schulen und Lehrer*innen. Die nächste PISA-Olympiade kann dafür noch zehn Jahre warten. □

Prof. Dr. Hans Brügelmann

Prof. i. R., Grundschulpädagoge und
Schriftsprachdidaktiker, Fachreferent
im Grundschulverband;
hans.bruegelmann@
grundschulverband.de

Nimmervoll, L. (2016): Statistiker hinterfragt Österreichs PISA-Verschlechterung.

In: Der Standard v. 8.12.2016. Download: <http://derstandard.at/2000048953546/Statistiker-hinterfragt-Oesterreichs-Pisa-Verschlechterung>

Reiss, K. u. a. (Hrsg.) (2016): PISA 2015. Eine Studie zwischen Kontinuität und Innovation. Waxmann: Münster/ New York.

Richter, S. (1992): Die Rechtschreibentwicklung im Anfangsunterricht und Möglichkeiten der Vorhersage ihrer Störungen. Phil. Diss. FB 12 der Universität Bremen. Verlag Dr. Kovac: Hamburg.

Robitzsch, A. u. a. (2016): Herausforderungen bei der Schätzung von Trends in Schulleistungsstudien. Eine Skalierung der deutschen PISA-Daten. Online am 6.12.2016: <http://econtent.hogrefe.com/doi/full/10.1026/0012-1924/a000177>

Stanat, P. u. a. (Hrsg.) (2016): IQB-Bildungstrend 2015. Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten

Ländervergleich. Waxmann: Münster / New York.

Wendt, H. u. a. (Hrsg.) (2016): Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Waxmann: Münster u. a.