

Protollnotiz zur Besprechung im Schulentwicklungsinstitut, TU Dortmund,  
am 28.09.2017, 10:00 Uhr bis 12:15 Uhr  
inkl. Nachklärungen per Email am 03.10.2017 und 04.10.2017 sowie weiterer  
Arbeiten, die durch das Gespräch angestoßen wurden

### **Thema: „Zwanzigeins-Hypothese“ und TIMSS 2015**

Teilnehmer: Frau Dr. Heike Wendt (TU Dortmund), Herren Prof. Dr. Lothar Gerritzen,  
PD Dr. Peter Morfeld (Uni Bochum)

1) LG und PM berichten zum Verein „Zwanzigeins e.V.“, der Zwanzigeins-Hypothese, geschichtlichen Aspekten und dass die neue Website des Vereins bald unter <https://zwanzigeins.jetzt> erreichbar sein wird. LG übergibt das Buch: Gerritzen L (2008) Zwanzigeins - für die unverdrehte Zahlensprechweise. Brockmeyer, Bochum. HW übergibt den TIMSS 2015-Berichtsband, die zwei internationalen Bände zu TIMSS 2011 in Papierform und die Doktorarbeit von Frau Hußmann (geb. Walzebug) (2015).

2) Die Gesprächsteilnehmer stimmen überein, dass TIMSS 2015 eine niedrige Leistung der deutschen Schüler in Mathematik ausweist, die allein auf einer schwachen Leistung in Arithmetik beruht (statistisch signifikante Unterschiede in den Vergleichen: Deutschland vs. VGEU, VGOECD).

HW erläutert, dass die Ursache für diesen Befund in den für Deutschland nicht curricular validen Aufgaben im Bereich Arithmetik gesehen wird. Andere Erklärungen werden in TIMSS 2015 nicht angeboten. Die Gesprächsteilnehmer stimmen überein, dass allerdings eine Zusatzauswertung allein mit curricular validen Aufgaben (TCMA: test curriculum matching analysis, vgl. S. 96 des TIMSS 2015-Berichts) zum Ergebnis führte, „dass die Testergebnisse Deutschlands auf der TIMSS-Leistungsskala nicht auf das Testdesign zurückzuführen sind. Modifikationen würden vermutlich allenfalls marginale Unterschiede in der internationalen Rangreihe herbeiführen“ (TIMSS 2015-Bericht, S. 109). HW erläutert, dass es schwierig ist, Aufgaben als „curricular valide“ zu identifizieren. Man hat sich im Wesentlichen an

NRW orientiert, aber die Zuordnung sei letztendlich eine Expertenentscheidung. Zudem seien diese Zusatzauswertungen nicht sehr ins Detail gegangen, so dass z.B. keine Auswertung mit curricular validen Aufgaben zum Teilbereich „Arithmetik“ vorliegt, sondern nur zur Mathematik insgesamt.

PM und LG hatten eine Zusammenfassung erstellt und als mögliche Teilerklärung dieses Defizits die verdrehte Zahlenaussprache im Deutschen angeführt (siehe die Anlagen TIMSS\_2015\_verdrehte Zahlenaussprache\_10Mai2017.pdf, Anhang\_TIMSS\_2015\_verdrehte Zahlenaussprache\_10Mai2017.pdf).

3) Die Gesprächsteilnehmer sind sich grundsätzlich einig, dass TIMSS 2015 nicht zur Prüfung von Einzelfragen geplant wurde und solche Sekundäranalysen daher nur mit Einschränkungen sinnvoll sind. HW berechnet die Schätzungen der Fehleranteile und ihrer Varianzen mit Hilfe des IDB-Analyzer von IEA (Jackknife-Verfahren) für die Gegenüberstellung „Deutschland gegen die 5 asiatischen Top-TIMSS-Staaten“ in Aufgabe M01\_01 („convert words to numerals“). HW stellt die Punktschätzungen und die geschätzten Standardabweichungen für die 6 Staaten zur Verfügung. Es ergeben sich Abweichungen in den Fehlerprozentsätzen zwischen den früheren Rechnungen von PM (z.B. Deutschland: 8,3%) und den jetzigen IDB-Analyzer-Ergebnissen (6,9%). PM sieht zudem ein Problem in der ungewichteten Mittelung zur Schätzung des Fehleranteils für eine Staatengruppe, wie es in TIMSS 2015, Methods and Procedures, Formel (9), S. 4.8 vorgegeben wird. Die Gesprächsteilnehmer vereinbaren, die Rechnungen noch einmal im Detail durchzugehen. PM wird Literatur zur Verfügung stellen, warum der Gesamtschätzer präzisionsgewichtet berechnet werden sollte.

### 3.1 Nachtrag (Email von PM am 03.10.2017):

Im Nachgang zur Besprechung führte PM Berechnungen auf Basis der von HW mit Hilfe des IDB-Analyzers erstellten Punktschätzungen und der geschätzten Standardabweichungen für die 6 Staaten durch und ermittelte den mittleren Fehleranteil der 5 asiatischen Staaten auf zwei Arten:

a) ungewichtet, vgl. Methods and Procedures in TIMSS 2015, Formel (9), S. 4.8

b) präzisionsgewichtet, entsprechend dem Statistik-Lehrbuch von J. Hartung (2009) Statistik, 15. Aufl., Oldenbourg-Verlag, München. Siehe dort den Abschnitt 2.2 auf S. 900 zum Thema „Kombinierte Schätzung des Gesamteffektes“. Nach Hartung 2009 muss präzisionsgewichtet gemittelt werden.

Die Varianz der 5 asiatischen Staaten wurde durch die Summe der Einzelvarianzen, geteilt durch 25 geschätzt (vgl. Methods and Procedures in TIMSS 2015, Formel (9), S. 4.8), die Varianz der Mittelwertdifferenz durch die Summe der Varianzen Deutschland und der 5 asiatischen Staaten. Der Test des Mittelwertunterschieds erfolgt über den Z-Test mit Normalverteilung.

Ergebnis: Der Unterschied in der Fehlerhäufigkeit zwischen Deutschland und den 5 asiatischen Staaten ist eindeutig signifikant. Bei Verwendung des ungewichteten Mittelwertes für die asiatischen Staaten beträgt der zweiseitige P-Wert 1,33 %, bei Verwendung des gewichteten Mittelwertes entsprechend Hartung 2009 liegt der zweiseitige P-Wert bei 0,34 %.

Anmerkung: diese Rechnungen basieren auf den Fehleranteilen F\_IDB (vgl. 3.2).

### 3.2 Nachtrag (Email von PM am 04.10.2017):

In dem Treffen am 28.09. ergaben sich unterschiedliche Fehleranteile für Deutschland. PM hatte 8.3% und HW 6.9% ermittelt. Erklärung: Während in der Rechnung PM alle Fehler mitgezählt wurden, wurde in der Rechnung HW ein Ergebnis nur dann als falsch gezählt, wenn eine falsche Option angekreuzt wurde, d.h., es wurde in der Rechnung HW nicht als ein Fehler gewertet, wenn gar nichts angekreuzt wurde.

Im Folgenden ein Überblick.

Variablen DIFF, OM, NR aus TIMSS 2015 Fourth Grade Almanacs

<https://timssandpirls.bc.edu/timss2015/international-database/>

Aufgabe: M1\_1 (Convert words to numerals)

DIFF = percent correct (Anteil der SchülerInnen, die korrekt angekreuzt haben)

OM = percent ommitted (Anteil der SchülerInnen, die nichts angekreuzt haben)

NR = percent not reached (Anteil der SchülerInnen, die die Aufgabe nicht erreicht haben)

F\_total = 100-Diff (Fehlerprozentsatz: falsch oder nichts angekreuzt)

F\_ohneOM = 100-DIFF-OM (Fehlerprozentsatz: falsch angekreuzt)

F\_IDB = Fehlerprozentsatz laut IDB-Analyzer-Rechnung am 28.09.2017

Country	DIFF	OM	NR	F_total	F_ohneOM	F_IDB
Germany	91.7	1.4	0	8.3	6.9	6.9
Singapore	93.7	0	0	6.3	6.3	6.35
Hong Kong SAR	95.7	0	0	4.3	4.3	4.33
Korea, Rep	96.5	0	0	3.5	3.5	3.5
Japan	97.5	0	0	2.5	2.5	2.45
Taiwan	97.7	.5	0	2.3	1.8	1.8

Da die Spalte DIFF im Almanac nur auf eine Stelle hinter dem Komma angegeben wird, kommt es zu Rundungsabweichungen in der zweiten Stelle hinter dem Komma zwischen F\_ohneOM und F\_IDB.

Da alle SchülerInnen aus allen 6 Staaten die Aufgabe erreichten (durchgängig NR=0), ist es sinnvoll, alle fehlenden Rückmeldungen zur Aufgabe ebenfalls als Fehler zu werten, d.h. es liegt auch dann ein Fehler vor, wenn gar kein Kreuz gesetzt wurde. Die mit dem IDB-Analyzer erstellten Berechnungen der Standardabweichungen sollten für F\_total wiederholt werden, um dann die Auswertungen unter 3.1 mit den korrigierten Fehleranteilsangaben erneut durchzuführen.

4) HW informiert, dass für 2019 eine weiterer TIMSS-Durchgang geplant ist, aber die Projektleitung von der TU Dortmund zu Prof. Schwippert, Universität Hamburg, wechseln wird (siehe <https://www.ew.uni-hamburg.de/ueber-die-fakultaet/personen/schwippert.html>). Prof. Schwippert ist Koautor des Berichts zu

TIMSS 2015. Eine wesentliche Herausforderung für TIMSS 2019 wird der diskutierte Übergang von einer paper/pencil- zu einer elektronischen Aufgabenbearbeitung und -erfassung sein, auf die einige der asiatischen Teilnehmerstaaten drängen.

Inhaltliche Erweiterungen von TIMSS, z.B. zusätzliche Aufgaben zur direkteren Prüfung der Zwanzigeins-Hypothese, scheinen daher nur sehr schwer umsetzbar.

5) HW benennt weitere Datenbasen zum möglichen Testen der Zwanzigeins-Hypothese und in diesem Zusammenhang interessierende Projekte und Gesprächspartner:

5a) In Deutschland wurde als Erweiterung von TIMSS 2015 im 4. Schuljahr ein sog. „Speed-Test“ (Kopfrechenaufgaben) durchgeführt. An diesem Speedtest nahmen alle deutschen Kinder von TIMSS 2015 teil (also nicht nur die durch Rotationen der Aufgaben verringerte Zahl, wie in der Hauptstudie). Zu jedem Kind wurde die zu Hause im Wesentlichen gesprochene Sprache dokumentiert. Die Daten aus diesem Speedtest stehen im Schulentwicklungsinstitut zur Verfügung. Als ein möglicher Vergleich zum Prüfen der Zwanzigeins-Hypothese käme z.B. eine Fehlergegenüberstellung von „türkische Sprache“ vs. „deutsche Sprache“ in Betracht. HW übergibt 4 Blätter mit Aufgaben zu den Grundrechenarten, aus denen geeignete für eine Auswertung ausgewählt werden können. Die Fragestellung müsste präzise erarbeitet werden, bevor Analysen versucht würden. HW wäre bereit an einem solchen Projekt mitzuarbeiten. PM und LG werden die Testaufgaben sichten und eine Rückmeldung geben.

Nachtrag: PM und LG haben die Aufgaben gesichtet. Es liegen danach keine typischen Aufgaben vor, um der Zwanzigeins-Hypothese gezielter nachgehen zu können, z.B. Größenvergleich von zwei Aufgaben mit Zahlendrehern oder wenn die zweite Ziffer der kleineren Zahl größer ist als die Summe der Ziffern der größeren Zahl. Auch ist unklar, welche Hypothese aufgestellt werden kann, da nicht bekannt ist, ob türkische Schüler, die Aufgabenstellungen übersetzen und dann in ihrer Muttersprache rechnen oder ob sie ohne Übersetzung versuchen, die Ausgabe in der deutschen Version mit verdrehter Aussprache zu lösen.

5b) HW berichtet, dass eine umfangreiche Ländervergleichsstudie im 4. Schuljahr vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) an der Humboldt-Universität, Berlin durchgeführt wurde (IQB-Ländervergleichsstudie). Angaben zur Primarstufe Mathematik finden sich bei [https://www.iqb.hu-berlin.de/institut/ab/primar\\_ma](https://www.iqb.hu-berlin.de/institut/ab/primar_ma). Das IQB beinhaltet ein Forschungsdatenzentrum „Das Forschungsdatenzentrum (FDZ) am Institut zur Qualitätsentwicklung im Bildungswesen (IQB) archiviert und dokumentiert die Datensätze aus nationalen und internationalen Schulleistungsstudien (z.B. DESI, IGLU, PISA) und stellt sie für Re- und Sekundäranalysen zur Verfügung. Interessierte Wissenschaftlerinnen und Wissenschaftler können beim FDZ den Datenzugang zu der jeweils gewünschten Studie beantragen“ (siehe: <https://www.iqb.hu-berlin.de/fdz>). HW sagt Unterstützung zu, wenn es darum geht, Daten aus dem IQB-Ländervergleich zur Prüfung der Zwanzigeins-Hypothese heranzuziehen. HW hat Interesse an einer solchen Arbeit mitzuwirken. PM und LG werden versuchen, aus dem Internet Informationen zur IQB-Ländervergleichsstudie zu erhalten, und Rückmeldung geben.

Nachtrag: PM und LG haben die neueste IQB-Auswertung (<https://www.iqb.hu-berlin.de/bt/BT2016/Bericht>) gesichtet. Diese Untersuchung belegt einen Rückgang der Mathematikkompetenz an den deutschen Grundschulen und nennt als Fazit:

„Zentral ist dabei die Frage, wie Schülerinnen und Schüler in den Fächern und Kompetenzbereichen, in denen Optimierungsbedarf besteht, durch eine gezielte Weiterentwicklung des Unterrichts besser gefördert werden können“.

Insofern liegt es nahe, auch der Zwanzigeins-Hypothese gezielt nachzugehen, um unnötige Barrieren im Mathematikunterricht abzubauen.

5c) HW übergibt die Doktorarbeit von Frau Anke Hußmann (geb. Walzebug), Mitarbeiterin am Schulentwicklungsinstitut an der TU Dortmund. In dieser Arbeit aus 2015 werden TIMSS-Daten genutzt, um der Hypothese nachzugehen, inwieweit eine sprachlich bedingte soziale Ungleichheit beim Lösen von mathematischen Testaufgaben vorliegt. PM und LG werden die Arbeit danach durchschauen, ob auch die Zwanzigeins-Hypothese behandelt wird oder sich andere interessante Aspekte

ergeben. HW bietet an zu vermitteln, falls es sinnvoll erscheint, mit Frau Hußmann Kontakt aufzunehmen.

5d) HW erläutert, dass die Bedeutung der Sprache für mathematische Leistungen erst in jüngerer Zeit empirisch aufgegriffen wird, wie in der Arbeit von Frau Hußmann (geb. Walzebug). Als weiteren möglichen Kontakt benennt HW Frau Prof. Dr. Susanne Prediger, Professorin für Grundlagen der Mathematikdidaktik am Institut für Erforschung und Entwicklung des Mathematikunterrichts an der TU Dortmund (<http://www.mathematik.uni-dortmund.de/~prediger/>). Herr Taha Kuzu verfasst derzeit eine Doktorarbeit bei Frau Prediger zum Thema „sprachensitive Mathematik“.

6) LG und PM berichten zu Schulversuchen in der DDR zur Zwanzigeins-Hypothese und zu einer Staatsarbeit an der Universität zu Köln:

Schellenberger M (1953) Zahlwort und Schriftbild der Zahl, VEB  
Bibliographisches Institut, Leipzig

Zehner (1955) Das invertierte Zahlensprechen als pädagogisch-  
psychologisches Problem. Habilitationsschrift, Fakultät für Mathematik und  
Naturwissenschaften der Technischen Hochschule Dresden

Ammareller MK (2006) Die nicht-invertierte Zahlensprechweise im  
arithmetischen Anfangsunterricht am Beispiel der Waldschule Bochum.  
Schriftliche Hausarbeit im Rahmen der Ersten Staatsprüfung für das Lehramt  
Primarstufe, Erziehungswissenschaftliche Fakultät der Universität zu Köln

PM und LG stellen die Daten dieser Arbeiten derzeit zusammen, um eine einheitliche Sekundäranalyse durchzuführen und einen Bericht/Publikation zu verfassen. Sie weisen zudem auf die Veröffentlichung von Prof. Meyerhöfer, Mathematikdidaktik der Universität Paderborn hin:

Meyerhöfer W (2015) Zweizehneins, Zwanzigeins, Einundzwanzig. Skizze  
einer stellenwertlogisch konsistenten Konstruktion der Zahlwörter im  
Deutschen. Pädagogische Korrespondenz 52/15, 21-41

Grundsätzlich besteht Interesse, auch weitere Schulversuche zur Prüfung der Zwanzigeins-Hypothese zu planen. HW verweist in dieser Frage auf Frau Prof. Prediger.

7) PM und LG erklären, dass für gemeinsame Forschungsarbeiten auch an refinanzierte Projekte gedacht werden kann. HW betont, dass die unter 5a und 5b genannten Daten mit öffentlichen Mitteln erstellt wurden und weitere Auswertungen stets sinnvoll erscheinen. PM und LG bedanken sich für die intensive Beratung durch HW und werden nach Durchsicht der genannten Materialien wieder Kontakt aufnehmen, um ein weiteres Treffen zu vereinbaren.