

Prof. Dr. Olaf Köller, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN)

Prof. Dr. Petra Stanat, Prof. Dr. Hans Anand Pant, Institut zur Qualitätsentwicklung im Bildungswesen (IQB)

Prof. Dr. Kristina Reiss, School of Education, Technische Universität München

## **Antwort auf den offenen Brief von Prof. Dr. Dr. h.c. Erich Wittmann zur der Lernstandserhebung VerA 3 Mathematik 2010**

In seinem Brief vom 31.05.2010 setzt sich Prof. em. Dr. Dr. h.c. Wittmann mit den Mathematikaufgaben aus dem VERA 3-Durchgang 2010 sehr kritisch, teilweise in polemischer Art und Weise auseinander. Seine Kernaussagen sind:

- Der 2010 eingesetzte Mathematiktest ist zu lang, zu textlastig und zu schwer für Schülerinnen und Schüler der 3. Jahrgangsstufe („Ein einziger Blick ... genügt um zu sehen, dass eine Reihe von Aufgabentexten und die Fülle der Aufgaben überhaupt das Fassungsvermögen von Kindern dieses Alters weit übersteigen.“)
- Es findet eine völlige Überbetonung der Leitidee Daten, Häufigkeit und Wahrscheinlichkeit statt. Breite Teile der Grundschulcurricula und der Bildungsstandards sind in dem Test unterrepräsentiert.
- Primäre Verursacher dieser Mängel sind die Bildungsforschung und Psychometrie.
- Deutschland befindet sich auf einem Weg ähnlich dem in den USA, wo die Testkritik und Evaluierung zu Verschlechterungen im Bildungssystem geführt habe.

Neben diesen Kernaussagen kritisiert Wittmann auch Einzelaufgaben. Bemerkenswert ist allerdings, dass er dabei im Wesentlichen seine Kernaussagen wiederholt (zu schwer, zu viel Daten, Häufigkeit und Wahrscheinlichkeit) und Alternativformulierungen vorschlägt, welche die fachliche Richtigkeit der Testaufgaben **keinesfalls** anzweifeln.

Wir möchten im Folgenden auf die Kernaussagen eingehen und darlegen, dass viele der Kritikpunkte Wittmanns (a) schlichtweg aus Unkenntnis über die Anlage der Lernstandserhebungen in der 3. Jahrgangsstufe resultieren und (b) aus einer subjektiven Theorie über den Mathematikunterricht der 3. Jahrgangsstufe rühren, die durch die empirischen Gegebenheiten in keinsten Weise gestützt wird. Wir beginnen mit einer kurzen Darlegung, welche Funktionen die Lernstandserhebungen in der Grundschule haben und wie Aufgaben für die Tests entwickelt werden.

## **1. Lernstandserhebungen in der Grundschule**

Lernstandserhebungen in Grundschulen (VERA) werden seit vielen Jahren durchgeführt. Sie begannen unter der wissenschaftlichen Ägide der Universität Koblenz-Landau (verantwortliche Leiter: Prof. Dr. Andreas Helmke und Prof. Dr. Ingmar Hosenfeld) und wurden später durch das Institut zur Qualitätsentwicklung im Bildungswesen (IQB: früher verantwortlich: Prof. Dr. Olaf Köller, neuerdings: Prof. Dr. Petra Stanat und Prof. Dr. Hans Anand Pant) fortgesetzt. Mittlerweile werden die Lernstandserhebungen in der Grundschule in allen 16 Ländern in dritten Jahrgangsstufen (VERA 3) durchgeführt. Die 16 Länder haben eine Steuerungs- und Leitungsgruppe eingesetzt, die dafür Sorge tragen, dass die am IQB entwickelten Testhefte für VERA im Einklang mit den Strategien der Länder zur Qualitätssicherung und Qualitätsentwicklung stehen. Mit VERA 3 werden wenigstens zwei Ziele verfolgt:

- 1.) Die Einzelschulen sollen gegen Ende der 3. Jahrgangsstufe Rückmeldungen erhalten, welche Leistungsstände ihre Schülerinnen und Schüler bis dato erreicht haben und ob diese Leistungsstände erwarten lassen, dass die Kinder bis zum Ende der 4. Jahrgangsstufe die in den Bildungsstandards der KMK festgehalten Leistungserwartungen erfüllen werden. Dazu erhalten die Schulen Rückmeldungen über die Leistungen der einzelnen Klassen und Schülerinnen/Schüler. Die erreichten Kompetenzstände werden dabei mit anderen Klassen/Schulen verglichen (teilweise nur mit Schulen, die eine vergleichbare Schülerschaft haben), zusätzlich können die Leistungen der Kinder neuerdings auf dem fünfstufigen Kompetenzmodell des IQB verortet werden. Das Kompetenzstufenmodell lässt eine kriteriale Verortung der Kompetenzstände zu.
- 2.) Die Aufgaben und die Ergebnisse sollen in den Schulen zur Schul- und Unterrichtsentwicklung beitragen, indem Kollegien gemeinsam die Ergebnisse und ihre Ursachen diskutieren und Maßnahmen planen, die der weiteren Förderung der Schülerinnen und Schüler dienen können. Für die weitere Unterrichtsentwicklung werden in VERA 3 didaktische Kommentare zu den Aufgaben entwickelt, die den Lehrkräften zur Verfügung gestellt werden. Darüber hinaus hat sich das in 11 Ländern laufende Projekt SINUS an Grundschulen das Ziel gesetzt, Lehrkräfte explizit im Bereich der Nutzung von Ergebnissen aus Vergleichsarbeiten zu professionalisieren.

## **2. Aufgabenentwicklung in VERA 3**

Bereits die Universität Koblenz-Landau hatte Strukturen etabliert, in denen die Aufgaben für VERA konsequent von Lehrkräften entwickelt wurden. Dazu wurden die Länder gebeten, Lehrkräfte zu nominieren und bereitzustellen, die in der Aufgabenentwicklung äußerst erfahren sind und zugleich die Realität an deutschen Grundschulen aus eigener Praxis sehr gut kennen. In einem weiteren Schritt der Qualitätssicherung wurden in VERA Fachdidaktiker verschiedener Universitäten eingebunden, welche jede Aufgabe hinsichtlich ihrer fachdidaktischen und fachlichen Eignung einschätzten und Überarbeitungsvorschläge machten. Unmit-

telbar vor dieser Erprobung sind darüber hinaus die Fachvertreter aller Länder damit befasst, die Aufgaben zu prüfen und sie dann abzunehmen bzw. Änderungen vorzuschlagen. In einem nächsten Schritt wurden alle Aufgaben in Stichproben hinsichtlich ihrer Eignung für den Einsatz in Grundschulklassen erprobt. Diese Vorgehensweise ist am IQB übernommen worden, neu ist lediglich, dass sich die Aufgaben explizit am durch die KMK verabschiedeten Kompetenzstufenmodell für die Bildungsstandards orientieren. Zusammenfassend gilt für die in VERA 2010 eingesetzten Mathematikaufgaben:

- Alle Aufgaben wurden durch erfahrene und von den Ländern benannte Grundschullehrkräfte entwickelt. Alle Aufgabenentwickler hatten breite Erfahrungen im Unterrichten von Grundschülerinnen und -schülern.
- Alle Aufgaben wurden von ausgewiesenen Professorinnen/Professoren für Mathematikdidaktik begutachtet und (wenn nötig) überarbeitet. Die fachdidaktische Leitung lag bei Frau Prof. Dr. Kristina Reiss (TU München).
- Alle Aufgaben wurden an hinreichend großen Stichproben von Drittklässlern hinsichtlich ihrer Eignung in der 3. Jahrgangsstufe erprobt. Dabei wurde vor allem auch darauf geachtet, dass alle Aufgaben, die in der 3. Jahrgangsstufe von zu wenigen Schülerinnen und Schülern gelöst werden konnten, ausgeschieden wurden. Alle Aufgaben wurden aufgrund ihrer empirisch gewonnenen Schwierigkeiten und einer inhaltlichen Analyse den fünf Stufen des Kompetenzstufenmodells im Fach Mathematik für die Primarstufe zugeordnet.
- Es wurden die Aufgaben so ausgewählt, dass alle Kompetenzstufen durch hinreichend viele Aufgaben repräsentiert wurden.

### **3. Zu Wittmanns Kritikpunkten**

#### **3.1 Der 2010 eingesetzte Mathematiktest ist zu lang, zu textlastig und zu schwer für Schülerinnen und Schüler der 3. Jahrgangsstufe**

Es gehört zu den grundlegenden Erkenntnissen der Bildungsforschung, dass die Leistungen von Grundschülerinnen und Grundschülern derselben Jahrgangsstufe erheblich variieren. Für die Mathematik bedeutet dies beispielsweise, dass man in der 3. Jahrgangsstufe teilweise Schülerinnen und Schüler antrifft, deren Kompetenzen eher den Erwartungen der 1. Jahrgangsstufe entsprechen, aber auch solche Kinder vorfindet, die keine Schwierigkeiten haben, Aufgaben, die typischerweise erst viel später behandelt werden, erfolgreich zu lösen. So gehörte es zu den bemerkenswerten Befunden der TIMS-Studie 2007, dass deutsche Schülerinnen und Schüler der 4. Jahrgangsstufe vergleichsweise gut abschnitten, obwohl der Mathematiktest in TIMSS viele Bruchrechenaufgaben enthielt, die in Deutschland zumeist erst ab der 6. Jahrgangsstufe eingeführt werden..

Die Normierungsbefunde des IQB haben gezeigt, dass die Leistungen der Schülerinnen und Schüler in der 3. Jahrgangsstufe auf der gewählten Skala breit streuen. Diese Skala ist so normiert, dass sie für Grundschülerinnen und -schüler in der 4. Jahrgangsstufe einen Mittelwert von  $M = 500$  und eine Streuung von  $SD = 100$  aufweist. In der 3. Jahrgangsstufe ergab sich folgende Verteilung auf die Kompetenzstufen:

Stufe Ia (unter 310 Punkten): 14%

Stufe Ib (310 – 390 Punkte): 25%

Stufe II (390 – 460 Punkte): 29%

Stufe III (460 – 530 Punkte): 19%

Stufe IV (530 – 600 Punkte): 9%

Stufe V (über 600 Punkte): 3%

Akzeptiert man die große Leistungsheterogenität in der Grundschule, so mutet die Feststellung, Aufgaben seien für die 3. Jahrgangsstufe zu leicht oder zu schwer, anachronistisch an. Solch eine Feststellung geht von der naiven Annahme einer homogenen Schülerschaft aus und unterschätzt das große Potenzial der Grundschulkinder. Gute Tests mit förderdiagnostischem Potenzial in allen Kompetenzgruppen müssen das gesamte Leistungsspektrum einer Jahrgangsstufe abbilden, dazu zählen selbstverständlich auch Aufgaben, die schwer sind und nur von den leistungsstärksten Schülerinnen und Schülern gelöst werden. Auch müssen sehr leichte Aufgaben enthalten sein, um damit Informationen über die Schwächsten zu gewinnen. Genau dies leistet der Test in VERA. Nach der Erprobung der Aufgaben in über 100 3. Klassen wurde eine Testzusammenstellung vorgenommen, bei der alle Kompetenzstufen mit hinreichend vielen Aufgaben (mindestens 6) abgebildet wurden. So konnte gesichert werden, dass diagnostische Informationen für alle Schülergruppen verfügbar waren.

Der große Unterschied zwischen Herrn Kollegen Wittmann und den in VERA 3 involvierten Akteuren liegt ohne Zweifel darin, dass in VERA Aufgaben nicht anhand gefühlter Schwierigkeiten ausgewählt werden, sondern empirisch gewonnene Schwierigkeiten Grundlage der Auswahl sind. Letzteres nimmt expliziten Bezug auf die Realität in deutschen Klassenzimmern und nicht auf naive Theorien über die Leistungsstände von Grundschulkindern.

### **3.2 Völlige Überbetonung der Leitidee Daten, Häufigkeit und Wahrscheinlichkeit, breite Teile der Grundschulcurricula und der Bildungsstandards sind unterrepräsentiert**

Herr Kollege Wittmann erkennt hier völlig richtig, dass im VERA 3-Durchgang im Jahre 2010 ein deutlicher Schwerpunkt auf die inhaltliche Kompetenz Daten, Häufigkeit und Wahrscheinlichkeit gelegt wurde. Dieser Inhaltsbereich stellt ein innovatives Element dar, das lange Zeit in den Lehrplänen der Länder zu wenig verankert war.

Die Konzentration auf einzelne inhaltliche Kompetenzen entspricht der Strategie, die von den Ländern seit Jahren in VERA in der Grundschule verfolgt wird. Es werden Schwer-

punkte auf zwei der fünf in den Bildungsstandards beschriebenen inhaltlichen Kompetenzen gelegt, die dann vertieft getestet werden. Eine gewisse Konzentration ist im Übrigen auch nötig, wenn man reliable und valide Informationen zu Kompetenzständen in Teilbereichen haben möchte. Weiterhin kann damit eine vertiefte Auseinandersetzung mit Leitideen in den Schulen initiiert werden. Im Jahre 2010 wurde als zweiter Schwerpunkt der Bereich Zahlen und Operationen getestet. Nur folgerichtig sind die übrigen Leitideen/inhaltliche Kompetenzen vernachlässigt. Hätte sich Herr Kollege Wittmann auch nur ansatzweise mit der Idee von VERA 3 auseinandergesetzt, so hätte er dies gewusst und keinen Ansatzpunkt für Kritik darin sehen können, dass sich in dem Test nicht die Breite der Curricula und Bildungsstandards wiederfindet.

An dieser Stelle sei weiterhin erwähnt, dass sich die Länder im Durchgang 2010 ganz bewusst für die Leitidee Daten, Häufigkeit und Wahrscheinlichkeit entschieden haben. Damit sollte ein Signal gesetzt werden, diesen Inhaltsbereich in der Grundschule zu stärken. Mit den didaktischen Kommentaren wurden Lehrkräften zugleich Anregungen für die Unterrichtsentwicklung in dieser Leitidee gegeben.

### **3.3 Primäre Verursacher der Mängel in den Tests kommen aus der Bildungsforschung und Psychometrie**

Herr Kollege Wittmann führt seit Jahren einen weitgehend irrationalen Feldzug gegen die Bildungsforschung, indem er argumentiert, Vertreter der Bildungsforschung und Psychometrie würden die Testerstellung mehr oder weniger allein verantworten und mangels Qualifikation dieser Personen seien die Tests schlecht. Auf VERA bezogen argumentiert er: „Bei VerA kann ich mich aber den Eindrucks nicht erwehren, dass die Tests umso schlechter werden, je mehr Einfluss die Bildungsforschung und die Psychometrie auf die Testerstellung gewinnen. Kompetente LehrerInnen und Mathematik-DidaktikerInnen, die das Curriculum, die Kinder und den Unterricht kennen, stellen so miserable Aufgaben nicht. Hier drängt sich schon die Frage auf, wer eigentlich die Qualifikation derjenigen prüft, die an den Agenturen für ‚Qualitätssicherung‘ tätig sind.“

Ob bewusst oder unbewusst will Herr Kollege Wittmann schon seit Jahren nicht zur Kenntnis nehmen, dass alle in den Tests verwendeten Aufgaben aus den Federn von Mathematiklehrkräften, die aufgrund ihrer besonderen Qualifikationen von den Ländern benannt werden, stammen und von qualifizierten Kolleginnen und Kollegen Wittmanns aus der Mathematikdidaktik geprüft und nötigenfalls überarbeitet wurden. Kein Bildungsforscher oder Psychometriker wird sich je anmaßen wollen, Mathematikaufgaben zu schreiben. Interessanterweise haben einige der Verfasser dieses Papiers Herrn Wittmann wiederholt auf diese Fakten hingewiesen. Er kann oder will dies offenbar nicht zur Kenntnis nehmen, stattdessen verunglimpft er ohne jede Grundlage eine Profession, mit der viele seiner Kolleginnen und Kollegen seit Jahren erfolgreich im Bereich der Unterrichtsforschung zusammenarbeiten.

### **3.4 Deutschland befindet sich auf einem Weg ähnlich den USA, wo die Testeritis und Evaluitis zu Verschlechterungen im Bildungssystem geführt habe**

Einmal abgesehen davon, dass es in den USA keine empirische Evidenz dafür gibt, dass die Leistungen amerikanischer Schülerinnen und Schüler infolge der Testprogramme abgenommen hätten, verkennt Wittmann, dass in VERA 3 primär das Ziel der Schul- und Unterrichtsentwicklung verfolgt wird. Es handelt sich bei VERA nicht um *High Stakes Tests* wie in den USA, wo es unbestritten Verwerfungen infolge der vielen Tests gibt. Kritiker von Assessment-Verfahren beziehen sich gern auf die USA und verschweigen gleichzeitig, dass es Nationen gibt, in denen sich über viele Jahre äußerst erfolgreich Systeme der Qualitätssicherung und Entwicklung etabliert haben. Ein Beispiel hierfür sind die Niederlande, in denen sich viele diagnostische Systeme entwickelt haben, die explizit Grundlage von Fördermaßnahmen sind.

### **3.5 Kritik an einzelnen Aufgaben**

Wir möchten im Folgenden nicht Item für Item auf Wittmanns Kritik eingehen, sondern an einigen Stellen exemplarisch antworten:

- Wenn man es auf einen einfachen Nenner bringt, dann kritisiert Herr Kollege Wittmann im Bereich Daten, Häufigkeit und Wahrscheinlichkeit insbesondere die Aufgaben, die sich mit Grundbegriffen der Wahrscheinlichkeit und einfachen Zufallsexperimenten befassen. Man kann durchaus der Meinung sein, dass ein Verständnis in der Grundschule „unnötig früh“ angestrebt wird. Allerdings sind diese Inhalte in den Bildungsstandards verankert und die Anforderungen der Aufgaben haben sich in der Pilotierung als realistisch erwiesen, da sie keineswegs viele Schülerinnen und Schüler überfordern. Eine propädeutische Vorbereitung in der Grundschule scheint also durchaus leistbar zu sein.
- Wir wissen aus zahlreichen empirischen Studien, dass Schülerinnen und Schüler der dritten Jahrgangsstufe in ihren fachlichen Leistungen extrem variieren. Eine Leistungserhebung muss alle Kinder fordern, damit die Grundlagen für eine individuelle Förderung deutlich werden (s.o.). Entsprechend brauchen wir Aufgabenstellungen, bei denen auch schwierigkeitsgenerierende Merkmale variiert werden. Das kann der Kontext sein (z.B. Aufgabe 12), das können auch Nebenbedingungen sein (z.B. Aufgabe 6), das kann aber auch ein schlicht in der Anlage schwieriges Problem sein (z.B. Aufgabe 9).
- Die Kritik zum Inhaltsbereich Zahlen und Operationen betrifft zunächst den Umgang mit Termen. Fraglos gehört dieses Thema in die siebte Jahrgangsstufe, allerdings in einer vertieften und abstrakten Form. Bei den vorliegenden Items aus VERA-3 geht es wesentlich darum, ein einfaches Wissen zum Erstellen von konkreten Termen (oder sollte man in vielen Fällen nicht besser „Berechnungen“ sagen?) explizit zu zeigen. Auch diese Kompetenz findet sich den Bildungsstandards der KMK. So wird hier ge-

fordert, „Beziehungen zwischen der Sache und den einzelnen Lösungsschritten (zu) beschreiben“ (z.B. Aufgabe 6, aber auch die von Herrn Wittmann als unangemessen leicht beurteilte Aufgabe 10), „Grundrechenarten (zu verstehen)“ (z.B. Aufgabe 7) oder „verschiedene Rechenwege (zu) vergleichen und (zu) bewerten (z.B. Aufgabe 14).

- Weitere Kritikpunkte stehen eher isoliert da. So fordert Kollege Wittmann, mehr Einmaleinsaufgaben „abzuprüfen“. Aus dieser Forderung wird ein tiefes Missverständnis deutlich. Es geht nicht darum „abzuprüfen“. Es geht darum, Lehrerinnen und Lehrern anhand der Aufgaben einerseits den Leistungsstand ihrer Klasse vergleichend zu verdeutlichen und ihnen andererseits Ideen für die Weiterentwicklung von Wissen und Kompetenzen der ihnen anvertrauten Schülerinnen und Schüler zu geben. Entsprechend sind wir sicher, dass jede Lehrerin und jeder Lehrer alle anderen Einmaleinsaufgaben mühelos selbst finden und in der Klasse behandeln kann.
- Aufgabe 11 wird als „wenig lebensnah“ kritisiert, „abgesehen von der Schülerzahl, die nicht weit von den heutigen Verhältnissen entfernt ist.“ In der Aufgabenentwicklergruppe wurde lange über dieses Item und seine Aufnahme in den Test diskutiert und das nur aus dem Grund, weil 34 Kinder in einer Grundschulklasse längst nicht mehr der Realität entsprechen. Hier ist Kollege Wittmann ganz offensichtlich nicht mehr auf dem neuesten Stand des Wissens.
- Gänzlich unverständlich ist uns schließlich die harsche Kritik an Aufgabe 15. Es ist nicht nachvollziehbar, warum dieser Aufgabentyp „Gift für die mathematische Bildung“ ist, der von Kollegen Wittmann in seinem Schulbuch benutzte Aufgabentyp Rechendreieck hingegen Gnade findet. Es sei angemerkt, dass sich die Entwicklergruppe explizit um unterschiedliche Aufgabenformate bemüht, um nicht Kinder zu bevorzugen oder zu benachteiligen, die ein bestimmtes Lehrwerk verwenden.

### **3.6 Abschluss**

Die Lernstandserhebungen in der Grundschule und in der Sekundarstufe I basieren auf einer konstruktiven Kooperation zwischen Fachdidaktik und Psychometrie. Damit verfügen die Länder aktuell über das effektivste Mittel zur flächendeckenden Implementation der Bildungsstandards und der mit ihnen verbundenen Kompetenzstufenmodelle. Kritik, wie sie von Herrn Kollegen Wittmann vorgetragen wird, führt auf schulischer wie politischer Ebene zur Verunsicherung. Das Urteil eines Kollegen ist durchaus wünschenswert, doch aus unserer Sicht zählt es dabei zum professionellen Selbstverständnis von Hochschullehrern, Urteile auf der Basis breiter Recherchen und der Auseinandersetzung mit Kolleginnen und Kollegen aus der Wissenschaft zu stützen und nicht auf naive Theorien, die empirisch widerlegt sind. Wir müssen leider feststellen, dass vom Kollegen Wittmann Standpunkte vertreten werden, die an die Zeit vor TIMSS in den 1990er Jahren und früher erinnern und von denen sich breite Teile der Fachdidaktik mit gutem Grund verabschiedet haben.