

Ländergemeinsame
Vergleichsarbeiten



Ländergemeinsame Vergleichsarbeiten in Deutschland (VERA-3 und VERA-8)

Kritikanalyse



Stand: 06. April 2009

StD Martin Sachse, Bayerisches Staatsministerium für Unterricht und Kultus,
Tel.: (089) 2186-2789, Fax: (089) 2186-2839 Mail: martin.sachse@stmuk.bayern.de

Inhalt

<u>Vorbemerkung</u>	3
<u>Ländergemeinsame Vergleichsarbeiten: Kritikanalyse</u>	6
1 Warum werden überhaupt Tests durchgeführt?	6
2 Warum werden ländergemeinsame Tests durchgeführt, die dann nicht ländervergleichend ausgewertet werden?	6
3 Werden nicht zu viele Tests, Prüfungen etc. an Schulen durchgeführt? („Testeritis“)	7
4 Wie können Vergleichsarbeiten in Jahrgangsstufen geschrieben werden, für die keine KMK-Bildungsstandards vorliegen?	8
5 Dient VERA dazu, die Arbeit von Lehrkräften zu beurteilen?	9
6 Die Teilnahme an den Vergleichsarbeiten (Vorbereitung, Durchführung und Nachbereitung [Korrektur und Dateneingabe]) stellt eine zusätzliche, unverhältnismäßig große Belastung der Lehrkräfte dar.	9
7 Lehrpläne und Lehrplaninhalte werden bei VERA nicht berücksichtigt.	11
8 Der Umgang mit den Ergebnissen und ihr Adressatenkreis ist unterschiedlich.	11
9 Die Tatsache, dass in manchen Ländern keine Benotung vorgenommen wird, erschwert in diesen Ländern eine verantwortungsbewusste Teilnahme.	13
10 Die Testhefte sind zu umfangreich.	13
11 Die Durchführung der Vergleichsarbeiten „blockiert“ schulorganisatorisch ganze Wochen.	14
12 Scheinbar unklare oder als ungewöhnlich empfundene Formulierungen verunsichern die Lehrkräfte bei der Korrektur (z. B. Nicht-Berücksichtigung der Rechtschreibung bei Grammatikaufgaben).	15
13 Die Korrekturanweisungen sind zu restriktiv.	16
14 Die Vergleichsarbeiten werden von Psychometrikern erstellt, die von den tatsächlichen Erfordernissen an den Schulen nichts wissen.	16
15 Es werden nur Kompetenzen und Problemlösungen geprüft, keine Inhalte oder Rechenfertigkeiten.	17
16 Die Terminierung der Vergleichsarbeiten ist a) zu früh oder b) zu spät.	17
17 Die Eingabe der Daten (Schülerergebnisse) ist nur online möglich.	18
18 Viele Kinder, insbesondere solche mit Migrationshintergrund oder aus bildungsfernen Schichten, verstehen die Aufgabenstellungen nicht.	19
19 Bei den Vergleichsarbeiten verfälschen manche Lehrkräfte die Testergebnisse, indem sie die Tests vorbesprechen oder während der Prüfung Hilfestellung geben.	19
20 Die zentrale, bundesweite Durchführung von Vergleichsarbeiten ist nicht geeignet, individuellen Denk- und Lernprozessen gerecht zu werden.	20
21 Die Tests sind auf ein mittleres Niveau hin ausgerichtet, leistungsschwächere Schülerinnen und Schüler müssen also notwendigerweise schlecht abschneiden.	21

Vorbemerkung

In den zurückliegenden Jahren haben sich alle Länder in Deutschland intensiv mit den Ergebnissen der internationalen Schulleistungsstudien IGLU und PISA auseinandergesetzt. Dabei hat sich gezeigt, dass erfolgreiche Staaten die Ziele, die ihre Schulen erreichen sollen, verbindlich und klar festlegen, meist in Form von nationalen Bildungsstandards für Unterrichtsfächer. Das Erreichen dieser Zielvorgaben wird in diesen Staaten in bestimmten Abständen überprüft. Dies geschieht z. B. durch Vergleichsarbeiten bzw. zentrale Lernstandserhebungen.

Ziel solcher Lernstandserhebungen ist es,

- allgemeine Informationen über das Kompetenzniveau im jeweiligen Fach in einer bestimmten Jahrgangsstufe zu gewinnen, also festzustellen, was die Schülerinnen und Schüler zu diesem Zeitpunkt können, und
- den Schulen die Möglichkeit zu geben, den Leistungsstand ihrer Schülerinnen und Schüler an einem objektiven, übergreifenden Maßstab zu verorten und somit zur Sicherung und Verbesserung der Qualität des Unterrichts beizutragen.

Die Kultusminister der Länder haben deshalb den Beschluss gefasst, auch in Deutschland für ausgewählte Fächer an Schnittstellen des Bildungswesens nationale Bildungsstandards entwickeln zu lassen und sie verpflichtend anzuwenden.

Am 2. Juni 2006 wurde dieser Beschluss ergänzt durch die „Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring“, die folgende vier zentrale Verfahren und Instrumente in einen Gesamtzusammenhang einordnet:

- internationale Schulleistungsuntersuchungen (z. B. TIMSS, PISA, PIRLS [= IGLU in Deutschland]);
- zentrale Überprüfungen des Erreichens der Bildungsstandards in einem Ländervergleich (Nachfolgevorhaben von PISA-E);
- Vergleichsarbeiten in Anbindung oder Ankoppelung an die Bildungsstandards zur landesweiten oder länderübergreifenden Überprüfung der Leistungsfähigkeit aller Schulen (VERA) und
- gemeinsame Bildungsberichterstattung von Bund und Ländern.

Als Konsequenz daraus sollen nunmehr in regelmäßigen Abständen ländergemeinsame Vergleichsarbeiten durchgeführt werden. Dabei mag der Name vielleicht irreführend sein, denn das erklärte Ziel dieser Arbeiten ist es gerade nicht, die Länderergebnisse miteinander zu vergleichen oder gegeneinander aufzurechnen, sondern im Sinne einer klar definierten Qualitätsentwicklung Schüler-, Klassen- und Schulleistungen innerhalb eines Landes zu vergleichen, um den Unterricht zu verbessern. Dabei geht es darum, eine objektive Rückmeldung darüber zu erhalten, auf welchem Leistungsstand die Schülerinnen und Schüler sind, welchen Nachhol- oder Förderbedarf es bei einzelnen gibt und wo der Unterricht unter Umständen verändert werden muss.

In der Praxis hat sich herausgestellt, dass aufgrund länderspezifischer Bildungstraditionen, wegen aktueller gesetzlicher bzw. schulrechtlicher Vorgaben oder aus anderen Gründen teilweise deutliche Unterschiede im Hinblick auf Vorbereitung, Durchführung und Weiterarbeit mit den Ergebnissen existieren. So werden die ermittelten Schülerleistungen in einigen Ländern beispielsweise benotet oder werden bei Übertrittsentscheidungen mit herangezogen; in anderen Ländern werden die Ergebnisse klassen- bzw. schulintern als Diagnoseinstrument und damit als Grundlage für eine individuelle Förderung genutzt, ohne dass dies unmittelbare schulrechtliche Konsequenzen (beispielsweise Zeugnisrelevanz) hätte.

Diese VERgleichsArbeiten, kurz VERA genannt, finden in den Jahrgangsstufen 3 (VERA-3: bundesweit), 6 (VERA-6: HH, HE, MV, SN, SH, TH) und 8 (VERA-8: bundesweit außer BW, SL ab 2009/10) statt, da die Länder der Überzeugung sind, dass mit den sog. Vergleichsarbeiten in der Bildungspolitik konsequent der Weg der Qualitätssicherung fortgesetzt werden kann, der mit der Teilnahme an den internationalen Vergleichsstudien eingeschlagen wurde.

Je nach Zeitpunkt, Umfang oder Art der erstmaligen Durchführung dieser Vergleichsarbeiten in einem Land wurden an die jeweils Verantwortlichen Kritikpunkte herangetragen, die auf mehreren Ebenen vorgebracht wurden, u. a.:

- persönliche Kontakte mit betroffenen Schülerinnen und Schülern, Lehrerinnen und Lehrern, Eltern;

- schriftliche Stellungnahmen an Kultusministerien, Landesinstitute oder Qualitätsagenturen;
- Berichterstattung in öffentlichen Medien, z. B. Verbandszeitschriften¹ und
- Landtagseingaben.

Häufig zeigt sich bei diesen Reaktionen, dass die Vorwürfe auf Unkenntnis der Sachlage oder unzureichenden Informationen basieren. Im Folgenden werden deshalb die hauptsächlichen Kritikpunkte aufgegriffen und in ihrer Stichhaltigkeit beurteilt. Aufgrund der längeren und bundesweiten Durchführung konzentrieren sich die Vorwürfe insbesondere auf die Vergleichsarbeiten in Jahrgangsstufe 3.

Die Argumente der Kritiker, so scheint es, fokussieren häufig Extrempositionen und sind mithin kaum geeignet, zu einer konstruktiven Auseinandersetzung beizutragen. Annemarie von der Groebens Vorwurf (siehe Anm. 1), der Staat misstraue seinen Pädagogen generell und spreche den Schulen das Vermögen ab, guten Unterricht und gute Pädagogik umsetzen zu können, vermag als Beispiel zu dienen für eine klare Fehleinschätzung der Intentionen von VERA – und sowohl gesamtdeutscher als auch länderspezifischer Bildungspolitik.

Zu hoffen ist deshalb auf eine sachliche Auseinandersetzung mit den Intentionen, der Organisation und den Formalia der Vergleichsarbeiten, wozu die folgenden Ausführungen einen Beitrag leisten wollen.

Es sei an dieser Stelle nochmals darauf hingewiesen, dass bei bestimmten Aspekten länderspezifischen Regelungen bzw. Abweichungen Rechnung getragen werden muss (Beispiel: Benotung der Vergleichsarbeiten nur in einigen Ländern).

¹ Nachfolgend eine kleine Auswahl der aktuellen Berichterstattung über VERA: Bartnitzky, Horst: VERA Deutsch 2008: Wie Kindern durch Vergleichsarbeiten Unrecht geschieht. In: GS aktuell 103, September 2008, S. 4-8; Gudjons, Herbert: Bessere Schulqualität durch zentrale Lernstandserhebungen: Pro. In: Pädagogik 7-8/08, S. 82; Hecker, Ulrich: VERA 2008 – Realität in Deutschland: „... ein immer hektischeres Teaching to the Test“. In: GS aktuell 103, September 2008, S. 3; Hecker, Ulrich: Zusammenstellung von Reaktionen auf eine Anfrage des Grundschulverbandes: VERA 08 – Ansichten und Einsichten. In: GS aktuell 103, September 2008, S. 15-17; Schulte, Thomas: Idiotische Benutzerführung. Leserbrief. In: Bayerische Schule 7/8 2008, S. 29; Vieth-Entus, Susanne: Politik will Mogeleyen von Lehrern bei Vergleichstests unterbinden. In: Der Tagesspiegel vom 28.10.2008; von der Groeben, Annemarie: Bessere Schulqualität durch zentrale Lernstandserhebungen: Contra. In: Pädagogik 7-8/08, S. 83; Weinrich, Klaus: VERA 2008 Mathematik: Viel Aufwand – wenig Nutzen. In: GS aktuell 103, September 2008, S. 9-14.

Ländergemeinsame Vergleichsarbeiten:

Kritikanalyse

1 *Warum werden überhaupt Tests durchgeführt?*

Die (KMK-)Gesamtstrategie bzw. die Plöner Beschlüsse vom 02.06.2008 legen die verbindliche Teilnahme der Länder an Vergleichsarbeiten fest. Eine Nicht-Beteiligung würde bedeuten, sich aktuellen internationalen positiven Erfahrungen² und Empfehlungen im Bereich der empirischen Bildungsforschung zu verschließen. Zudem belegen TIMSS und andere Schulleistungsstudien, dass die sog. neue Aufgabekultur wie auch weitere Verbesserungen des Bildungssystems vor allem über Tests (bzw. darin enthaltene Items) eingeführt wurden und werden. Schließlich bedarf eine verstärkte Outcome-Orientierung, bedingt auch durch die Einführung von Bildungsstandards, einer effektiven Überprüfung der Ergebnisse.

2 *Warum werden ländergemeinsame Tests durchgeführt, die dann nicht ländervergleichend ausgewertet werden?*

Der Begriff „Vergleichsarbeiten“ ist in der Tat irreführend, denn das erklärte Ziel von VERA ist gerade nicht ein Vergleich der Länder untereinander und kann aufgrund der spezifischen Anlage von VERA auch nicht wissenschaftlich bzw. psychometrisch fundiert über die Ländergrenzen hinweg durchgeführt werden.

Der Vorteil einer ländergemeinsamen Erstellung und zeitgleichen Durchführung liegt jedoch in der Kostenersparnis, indem die Testhefte, Rückmeldeformate und methodisch-didaktischen Materialien für (fast) alle Länder von nur einer Institution

² Staaten, die bei internationalen Large-Scale-Assessments gut abschneiden, verfügen zumeist über ein Schulsystem, das eine Balance gefunden hat zwischen einerseits größtmöglicher Freiheit und Selbständigkeit für die einzelne Schule sowie andererseits regelmäßiger Rechenschaftsablegung über Bildungs- und Erziehungsarbeit anhand von vorgegebenen Standards – in Deutschland beispielsweise durch Vergleichsarbeiten.

(VERA-3: Universität Koblenz-Landau; VERA-8: Institut zur Qualitätsentwicklung im Bildungswesen [IQB], Berlin; Zusammenführung beim IQB seit 2009) erarbeitet werden. Bereits die Unterschiede bei der Verpflichtung zur Teilnahme (Anzahl der Fächer, Niveau der ausgewählten Testhefte, teilnehmende Schularten) und der Umfang von schulrechtlichen Konsequenzen (Benotung, Relevanz für Schullaufbahneempfehlung) belegen jedoch länderspezifische Eigenarten, zumal einige Länder auch eigene Rückmeldeformate entwickeln.

Damit zählt VERA auch nicht zur Gattung der Ländervergleiche, sondern dient (vor allem vor dem Hintergrund weiter führender Materialien) der Diagnose und Förderung bzw. der Unterrichtsentwicklung. Ein nationaler Referenzrahmen, der mit den bundesweit verbindlichen KMK-Bildungsstandards grundgelegt wurde, steht gleichzeitig als Referenzgröße für den eigenen Unterricht zur Verfügung und kann von der Bildungsadministration zur Überprüfung genutzt werden, ob und inwieweit die KMK-Bildungsstandards an den Schulen umgesetzt bzw. erreicht werden.

3 *Werden nicht zu viele Tests, Prüfungen etc. an Schulen durchgeführt? („Testeritis“)*

Tatsächlich ist in den letzten Jahren an den Schulen in Deutschland eine Häufung von Prüfungen, Tests und Leistungsnachweisen zu beobachten. Diese Entwicklung geht jedoch einher mit Maßnahmen, die das Qualitätsniveau von Schule stärken und deshalb eine klare Kontrolle der Ergebnisse erfordern (vgl. 1). Einige Länder reagieren auf die Zunahme ländergemeinsamer bzw. zentraler Tests durch eine Flexibilisierung der Durchführung weiterer Prüfungen: Bislang verpflichtende andere Tests werden zunehmend fakultativ vorgeschlagen, können ganz entfallen oder durch die Teilnahme an Vergleichsarbeiten kompensiert werden. Die Durchführung von VERA hat in vielen Ländern zur Folge, dass bisher verbindliche andere Prüfungen (z. B. Jahrgangsstufentests) nun auf freiwilliger Basis durchgeführt oder durch VERA ersetzt werden können.

Zudem sind Leistungsmessungen in ihrer Intention und Wirkung zu differenzieren:

- internationale Vergleichsstudien (z. B. PISA);

- nationale Vergleichsstudien zur zentralen Überprüfung der KMK-Bildungsstandards („Ländervergleiche“, bislang PISA-E);
- Verbesserung der Unterrichtsentwicklung (durch neue Aufgabenkultur; VERA);
- Diagnose und Förderung (durch zusätzliche didaktische Materialien; VERA);
- Lehrplanüberprüfung (Klassenarbeiten, Schulaufgaben, Stegreifaufgaben, Jahrgangsstufentests etc.)

Die ländergemeinsamen Vergleichsarbeiten (VERA) liefern somit Ergebnisse, die in dieser Form durch keine anderen Prüfungen erbracht werden. Gerade auch die Möglichkeit, einen Vergleich verschiedener Regionen und Stadtteile (ggf. in Kontrastierung zu vergleichbaren Schulen [„fairer Vergleich“]) und einen chronologischen Längsschnitt durchzuführen, wenn VERA regelmäßig wiederholt wird, gewährleistet den Nutzen der Durchführung.

Zudem haben Schülerinnen und Schüler ein Recht darauf, ihre Lernanstrengungen und -bemühungen gewürdigt zu sehen. Eine entsprechende Überprüfung kann verhindern, dass Kenntnisse und Kompetenzen, die nicht gefordert und überprüft werden, unnütz erscheinen.

Eine allgemein verbindliche Aussage, wie viele Tests sinnvoll sind, ist nicht zu treffen, da dies von diversen Faktoren abhängig ist (Umfang der Prüfungen, teilnehmende Jahrgangsstufe etc). Pauschalierungen sind insbesondere in diesem Zusammenhang zu vermeiden.

4 *Wie können Vergleichsarbeiten in Jahrgangsstufen geschrieben werden, für die keine KMK-Bildungsstandards vorliegen?*

Die [KMK-Bildungsstandards](#) liegen für die Jahrgangsstufen 4, 9 und 10 vor. Dennoch ist ihre Überprüfung auch in den Jahrgangsstufen 3, 6 und 8 möglich, denn

- die Items und Lösungserwartungen der Vergleichsarbeiten wurden in Bezug auf die vorliegenden Bildungsstandards für die jeweiligen Jahrgangsstufen adaptiert (Prämisse: Wenn die Schülerin bzw. der Schüler in Jahrgangsstufe 10 eine

bestimmte Kompetenz(-stufe) erreichen soll, dann muss er in Jahrgangsstufe 8 die abgeprüfte, reduzierte Kompetenz(-stufe) erreicht haben.);

- es handelt sich um eine sog. intermediäre Erhebung, die untersucht, wie weit man auf dem Weg hin zur Erreichung der Bildungsstandards ist.

Damit ergeben sich für die Verantwortlichen Möglichkeiten zur rechtzeitigen Nachsteuerung, um ggf. das Erreichen der Bildungsstandards an Gelenkstellen des Bildungswesens noch positiv beeinflussen bzw. sicher stellen zu können.

5 Dient VERA dazu, die Arbeit von Lehrkräften zu beurteilen?

Die Teilnahme an Vergleichsarbeiten und ihre spezifische Anlage, die sich weniger an sozialen Bezugsnormen als an objektiven psychometrischen Erkenntnissen ausrichtet, ist nicht als Kritik am bisherigen Beurteilungs- bzw. Bewertungssystem der Lehrkräfte zu verstehen. Es handelt sich dabei um eine Ergänzung bisheriger Praxis; die Kompetenz der Lehrkräfte, verantwortungsvoll guten Unterricht zu planen, durchzuführen und zu evaluieren, wird dadurch nicht in Frage gestellt, sondern findet eine sinnvolle Erweiterung. Insofern versteht sich VERA nicht als Korrektiv, sondern explizit als Hilfestellung für die Arbeit der Lehrkräfte, d. h. es geht nicht um eine Beurteilung der Lehrkräfte.

6 Die Teilnahme an den Vergleichsarbeiten (Vorbereitung, Durchführung und Nachbereitung [Korrektur und Dateneingabe]) stellt eine zusätzliche, unverhältnismäßig große Belastung der Lehrkräfte dar.

Es ist unbestreitbar, dass die Teilnahme an Vergleichsarbeiten auf den ersten Blick einen zusätzlichen Arbeitsaufwand für Lehrkräfte bedeutet. Neben dem erforderlichen Abwägen zwischen Nutzen und Aufwand sind jedoch auch weitere Aspekte zu berücksichtigen:

- Die Teilnahme an VERA bedingt in einigen Ländern den Wegfall anderer, bislang verpflichtender Prüfungen. Manchmal ergänzen sich die Vergleichsarbeiten in Kombination mit weiteren (zentralen, schulinternen oder klassenspezifischen) Tests zu einer validen Beurteilungsmöglichkeit.

- Bei sinnvoller Verortung in den Unterrichtsverlauf und entsprechend überlegter Nachbereitung der Ergebnisse lassen sich die Vergleichsarbeiten organisch im Unterricht verankern. Sie bieten damit die Möglichkeit, über den Termin der unmittelbaren Durchführung hinaus Impulse für die Unterrichtsentwicklung zu setzen. Insbesondere die Verfügbarkeit methodisch-didaktischer Materialien mit ihrem umfangreichen Beratungs- und weiterführenden Aufgabenfundus ist geeignet, die Lehrkräfte zu entlasten. VERA gibt damit konkrete Hilfestellung für den eigenen Unterricht.
- Diagnose und (individuelle) Förderung gehören zu den Kernaufgaben der Lehrtätigkeit. Gerade durch ihre wissenschaftlich abgesicherte Anlage vermögen die Vergleichsarbeiten hier Unterstützung zu bieten, die eine je eigene, lediglich auf individueller Prognose abgesicherte Einschätzung nur eingeschränkt leisten könnte.
- Auch die Korrektur von Prüfungsaufgaben bzw. Schülerleistungen gehört zu den Aufgabenbereichen, die dem Lehrberuf inhärent und teilweise in Ländergesetzen oder -verordnungen festgeschrieben sind.
- Die Beschäftigung mit den Ergebnissen und der gemeinsame Austausch darüber in den Kollegien, beispielsweise im Rahmen von Fachkonferenzen oder Jahrgangsstufenbesprechungen, wird überwiegend nur anfangs als zusätzliche Belastung empfunden. Durch konstruktiven Austausch von Informationen und Materialien ist dieser Aspekt geeignet, mittelfristig Synergieeffekte herbeizuführen und damit langfristig zu einer Entlastung beizutragen.

Nicht vergessen werden darf, dass die Durchführung der Vergleichsarbeiten (insbesondere in ihrer ländergemeinsamen Anlage) noch relativ neu ist, so dass sich immer wieder Optimierungsmöglichkeiten aufzeigen. In diesem Zusammenhang ist auch der von vielen Ländern inzwischen angebotene zentrale Druck und Versand zu sehen, der den tatsächlich hohen Arbeitsaufwand an Schulen durch das Kopieren der kompletten Testhefte spürbar reduziert.

Bedenkt man den Nutzen der Vergleichsarbeiten, so erscheint ein objektiv ermittelter Korrekturaufwand von durchschnittlich vier Stunden für eine gesamte Klasse als durchaus vertretbar und angemessen. Gerade die Anlage als Test erleichtert eine zügige Korrektur im Vergleich zu herkömmlichen Proben an Grundschulen oder

Schulaufgaben an weiterführenden Schulen; vergleichbare Zeiten sind an den weiterführenden Schulen festzustellen.

7 Lehrpläne und Lehrplaninhalte werden bei VERA nicht berücksichtigt.

Im Fokus der Vergleichsarbeiten steht die (Über-)Prüfung von Kompetenzen, die die Schülerinnen und Schüler zu einem bestimmten Zeitpunkt erreicht haben. Aufgrund ihrer ländergemeinsamen Anlage können sich die Aufgaben nicht explizit auf konkrete Lehrpläne oder Lehrplaninhalte eines bestimmten Landes beziehen, sondern sind so ausgerichtet, dass sie Aspekte thematisieren, die durch ihre Festlegung in den KMK-Bildungsstandards länderübergreifend als bedeutsam erkannt sind und damit eben auch Lehrplan relevant sind.

Die Vergleichsarbeiten zeigen den individuellen Entwicklungsstand und das Entwicklungspotenzial auf, da beispielsweise auch Ergebnisse zu Themen- bzw. Inhaltsbereichen vorliegen, die im Unterricht noch nicht behandelt wurden. Damit ist sichergestellt, dass auch gute Schülerinnen bzw. Schüler erkannt und entsprechend ihrem individuellen Leistungsstand gefördert werden können.

Hervorzuheben ist jedoch, dass ohne entsprechend fundierte Kenntnisse bzw. ohne vorhandenes Wissen keine Kompetenzen geprüft werden können: Insofern ist der Ansicht entgegenzutreten, es ginge lediglich um Kompetenzen, nicht (mehr) um Inhalte: Nur auf einer soliden inhaltlichen Grundlage ist es möglich, kompetenzorientiert zu arbeiten.

8 Der Umgang mit den Ergebnissen und ihr Adressatenkreis ist unterschiedlich.

Bedingt durch länderspezifische Regelungen sind hier teilweise große Unterschiede möglich. Zum einen bedingt bereits die Anlage der Rückmeldeformate, welche Informationen in welchem Umfang (zurück-)gemeldet werden. Hier bieten sich mehrere Möglichkeiten an, deren Einbezug oder Ausschluss häufig länder-

spezifischen Gesetzen oder Traditionen geschuldet sind; es ist daher nicht möglich und auch nicht intendiert, generell die eine oder andere Alternative als beste oder einzig sinnvolle darzustellen.

Die Rückmeldeformate können beispielsweise folgende Informationen beinhalten:

- Darstellung der Leistungen der einzelnen Schülerin bzw. des einzelnen Schülers auf einem überindividuellen Kompetenzstufenmodell;
- Wiedergabe der Lösungshäufigkeiten eines Items innerhalb einer Klasse;
- Vergleich der Klassenleistung bei einer Domäne mit der Parallelklasse und/oder
- Abbildung der Schulleistung vor dem Hintergrund des Landesergebnisses.

Insbesondere die übersichtliche Veranschaulichung in Graphiken ist geeignet, Interpretationen und Schlussfolgerungen Ziel gerichtet und rasch vorzunehmen.

Daneben ist zu unterscheiden zwischen den verschiedenen Adressaten(-gruppen) der Rückmeldungen:

- teilnehmende Schülerin bzw. teilnehmender Schüler;
- betroffene Eltern, ggf. Elternvertretung;
- beteiligte Lehrkraft;
- Schulleitung und bzw. oder
- Bildungsadministration.

Je nach ausgewählter Zielgruppe kann der Informationsumfang differenziert werden; so können beispielsweise Lehrkräfte, Schulleitung und Bildungsadministration alle Daten erhalten, die Schülerinnen und Schüler bzw. ihre Eltern nur die jeweiligen Individualergebnisse und die Elternvertretung nur das Schulergebnis. Damit wird zum einen den Belangen des Datenschutzes Rechnung getragen, zum anderen aber den einzelnen Gruppen die Möglichkeit gegeben, ihren Aufgaben bzw. Interessen gerecht Schlussfolgerungen zu ziehen und geeignete Schritte einzuleiten.

9 Die Tatsache, dass in manchen Ländern keine Benotung vorgenommen wird, erschwert in diesen Ländern eine verantwortungsbewusste Teilnahme.

Tatsächlich findet in einigen Ländern keine Benotung der Ergebnisse statt, unmittelbare schulrechtliche Konsequenzen (z. B. Laufbahneempfehlungen bei VERA-3) sind dort nicht mit den Vergleichsarbeiten verbunden. Dies ergibt sich beispielsweise aus der Tatsache, dass in diesen Ländern Prüfungsinhalte, die sich nicht explizit auf den in der teilnehmenden Schulart zugelassenen Lehrplan beziehen, nicht bewertet werden dürfen. Zudem raten die beteiligten Wissenschaftler und Psychometriker davon ab, die Vergleichsarbeiten zum Zweck der Benotung heranzuziehen.

Sinnvoll erscheint es auch, die Schülerinnen und Schüler zunehmend dahingehend zu motivieren, Leistung auch ohne Notendruck zu erbringen. Dieses Ziel geht einher mit den aktuellen Erkenntnissen der psychologischen und pädagogischen Wissenschaft, dass gute Leistungen vor allem in einer benotungs- und damit stressfreien Atmosphäre zu erbringen sind: So sollen Arbeitsphasen und Prüfungsphasen getrennt voneinander abwechseln.

10 Die Testhefte sind zu umfangreich.

Um jede Leitidee umfassend und Erkenntnis führend überprüfen zu können, bedarf es immer einer nicht zu knappen Anzahl an Items. Dies wiederum führt zu einer bestimmten Testlänge, die allerdings in Vorstudien (Pilotierung, Normierung) erhoben und deshalb als durchaus lösbar ermittelt wird.

Im Hinblick auf die Anzahl der nötigen Kopien bzw. die mit Druck und Kopieren verbundenen Kosten dringen die Länder bei den verantwortlichen Institutionen dennoch nachdrücklich auf eine Reduzierung der Seitenzahl. Leider konnte dieses Anliegen im aktuellen Schuljahr nur bedingt berücksichtigt werden, da sachliche Gründe die momentane Gestaltung der Testhefte erforderten:

- Jede Aufgabe beansprucht aufgrund ihres Schwierigkeitsgrades eine feste Position innerhalb des Testheftes. Aufgaben können demzufolge nicht beliebig getauscht werden, um Platz zu sparen.
- Die Kombination der Lese- und Grammatikaufgaben im Fach Deutsch erfordert darüber hinaus eine besondere Positionierung der einzelnen Aufgaben.
- Bestimmte Fächer – Deutsch, Englisch, Französisch – bedingen allein aufgrund ihrer Inhalte (Texte) einen bestimmten Umfang, während in den Testheften Mathematik enthaltene Graphiken deutlich lesbar und damit in entsprechender Größe abgebildet werden müssen.

Da VERA ein ländergemeinsames Projekt ist, ist es den Ländern auch nur eingeschränkt möglich, die Testhefte individuell zu verändern, zumal dies den Korrekturanweisungen nicht entsprechen und Testergebnisse verfälschen würde.

Es ist jedoch davon auszugehen, dass bei längerfristiger Durchführung und zunehmender Erfahrung eine Reduzierung der Seitenzahl erfolgen kann.

11 Die Durchführung der Vergleichsarbeiten „blockiert“ schulorganisatorisch ganze Wochen.

Die Vergleichsarbeiten an der Grundschule (VERA-3) werden in zwei Fächern durchgeführt (Deutsch und Mathematik). Tatsächlich sind also zwei Tage (ab 2010: drei Tage, da die Domäne „Lesen“ aufgrund ihres Umfangs ausgegliedert wird) von der Durchführung dieser 60minütigen Tests betroffen.

In den weiterführenden Schulen können bis zu drei Tage von der Durchführung der Vergleichsarbeiten betroffen sein (Deutsch, erste Fremdsprache, Mathematik), was häufig eine Verteilung auf zwei Wochen bedeutet.

Da diese Tests jedoch nicht als Additum zu verstehen sind, sondern sich organisch in die Unterrichtsverpflichtungen der Lehrerinnen und Lehrer integrieren lassen, ist keineswegs von einer „Blockade“ z. B. durch unterrichtsfremde Aktivitäten

auszugehen, sondern bei bewusster Nutzung der Ergebnisse im Gegenteil von einer Bereicherung.

An manchen Schulen wird inzwischen auch die Möglichkeit ergriffen, die Vergleichsarbeiten im Skilager, im Schullandheim oder an vergleichbaren Orten durchzuführen, so dass keine in diesem Zeitraum traditionell geplanten Aktivitäten ausfallen müssen.

12 *Scheinbar unklare oder als ungewöhnlich empfundene Formulierungen verunsichern die Lehrkräfte bei der Korrektur (z. B. Nicht-Berücksichtigung der Rechtschreibung bei Grammatikaufgaben).*

Von Lehrkräften vielleicht als ungewöhnlich empfundene Korrekturanweisungen (z. B. Nicht-Berücksichtigung der Rechtschreibung bei der Aufgabenlösung im Bereich „Sprache und Sprachgebrauch“) ergeben sich aus testtheoretischen bzw. testpraktischen Erfordernissen, die aber durchaus durchdacht und gewollt sind.

So ist es einsichtig, dass Items, die darauf ausgerichtet sind, Kompetenzen im Bereich der Grammatik abzuprüfen, nicht zugleich Fertigkeiten aus der Domäne „Schreiben“ testen sollen und können. Die Anweisungen zur Korrektur, die eine rasche und exakte Eingabe sowie Auswertungsmöglichkeit vorsehen, sind deshalb so ausgelegt, dass keine Vermischung bzw. Kombination der Domänen möglich ist. Andernfalls könnte keine domänenspezifisch genaue Diagnose und damit auch keine Ziel gerichtete individuelle Förderung erfolgen.

Nur so lässt sich über Klassen und Schulen hinweg eine einheitliche Beurteilung der Ergebnisse gewährleisten und nur so haben die daraus abgeleiteten und zur Verfügung gestellten didaktischen Materialien Sinn. Im Übrigen entsprechen die Tests in ihrer gesamten Konzeption damit auch den groß angelegten internationalen Vergleichsstudien wie PISA und PIRLS/IGLU³.

³ Internationale Grundschule-Leseuntersuchung.

13 Die Korrekturanweisungen sind zu restriktiv.

Manchmal wird kritisch kommentiert, dass beispielsweise keine (Teil-)Punkte vergeben werden, wenn Teilergebnisse richtig sind oder Teilaufgaben korrekt gelöst wurden (Beispiel 1: zwei Reimpaare gefragt, ein Reimpaar richtig; Beispiel 2: sechs Verben gefragt, fünf Verben richtig; in beiden Fällen: Lösung komplett falsch).

Die Bewertung von Tests, die den Anspruch erheben, valide (= Messung genau dessen, was überprüft werden soll) und reliabel (= formale Genauigkeit) zu sein, bedingt eine eindeutige Anleitung der Lehrkräfte bei der Korrektur, d. h. es wird ausschließlich „richtig/falsch“ bewertet, Zwischenstufen gibt es nicht. Nur so lässt sich über die Klassen und Schulen hinweg eine einheitliche Beurteilung der Ergebnisse gewährleisten – und nur so haben die daraus abgeleiteten und zur Verfügung gestellten didaktischen Materialien Sinn, indem sie eine präzise Diagnose und eine darauf abgestimmte individuelle Förderung ermöglichen.

Im Übrigen entsprechen die Tests in ihrer gesamten Konzeption auch den groß angelegten internationalen Vergleichsstudien wie PISA und PIRLS/IGLU.

14 Die Vergleichsarbeiten werden von Psychometrikern erstellt, die von den tatsächlichen Erfordernissen an den Schulen nichts wissen.

Selbstverständlich sind in die jeweiligen Gruppen, die die Aufgaben bzw. Testhefte entwickeln, neben den Wissenschaftlern auch „Praktiker“ integriert. In beiden Fällen, also sowohl bei VERA-3 als auch bei VERA-8, sind Lehrkräfte auch aus Schulen bzw. aus den Landesinstituten maßgeblich beteiligt. An das IQB wurden Lehrkräfte auch längerfristig abgeordnet, um sicherzustellen, dass Erfahrungen aus der Unterrichtspraxis und dem alltäglichen Lehrberuf ausreichend gewürdigt werden.

Da darüber hinaus auch Vertreter der wissenschaftlichen Fachdidaktiken an der Aufgabenentwicklung beteiligt sind, ist ein deutlicher Praxisbezug gewährleistet, der es den Lehrkräften an den Schulen ermöglicht, die Ergebnisse der Vergleichsarbeiten unmittelbar für ihren Unterricht fruchtbar zu machen.

15 Es werden nur Kompetenzen und Problemlösungen geprüft, keine Inhalte oder Rechenfertigkeiten.

Im Unterschied zu traditionellen Leistungsnachweisen, die sich auf das Abprüfen von Lehrplaninhalten konzentrieren, fokussieren Vergleichsarbeiten ergebnisorientiert auf die Überprüfung von vorhandenen Kompetenzen. Sie orientieren sich deshalb an Kompetenzmodellen und sind so konzipiert, dass sie unterschiedliche Kompetenzen in den betroffenen Fächern systematisch überprüfen.

Nicht korrekt ist jedoch, dass Inhalte bzw. Rechenfertigkeiten keine Rolle spielen: Die Vergleichsarbeiten sind so angelegt, dass auch basale Rechenfähigkeiten oder Sprachfertigkeiten beurteilt werden oder dass Kenntnisse abgeprüft werden, die ein altersgemäßes bzw. schulartspezifisches Wissen voraussetzen.

16 Die Terminierung der Vergleichsarbeiten ist a) zu früh oder b) zu spät.

In der Einschätzung des Termins für die Vergleichsarbeiten gehen die Meinungen auseinander: Während die einen Kritiker den Termin der Durchführung – unabhängig von VERA-3 oder VERA-8 – als zu früh empfinden, plädieren andere für eine spätere Testung.

- Als zu früh wird der Testzeitpunkt häufig dann eingeschätzt, wenn man davon ausgeht, dass bestimmte Lehrplaninhalte zu diesem Zeitpunkt noch nicht vermittelt wurden. Diese Annahme lässt aber das Faktum außer acht, dass sich die Vergleichsarbeiten ohnehin nicht an den länderspezifischen Lehrplänen orientieren. Naturgemäß müssten dann auch die Länder, deren Schuljahr insgesamt später begonnen hat (also vor allem Baden-Württemberg und Bayern), immer signifikant schlechter abschneiden – und dies bei allen länderübergreifenden Tests, was definitiv nicht der Fall ist. Insofern kann sowohl bundes- wie auch bayernweit durchaus von einer Gleichheit der Bedingungen und Anforderungen ausgegangen werden.
- Als zu spät wird der Testzeitpunkt häufig dann eingeschätzt, wenn befürchtet wird, bis zum Ende des Schuljahres stünde nicht mehr genügend Zeit zur Verfügung, um mit den Ergebnissen adäquat zu arbeiten. Diese Meinung

vernachlässigt jedoch zum einen die zeitnahe Ergebnismeldung – und damit Interventionsmöglichkeit –, zum anderen die (erwünschte) Kooperation der Lehrkräfte auch über die Schuljahresgrenzen hinaus.

Generell erscheint es prinzipiell sinnvoller, den Testzeitpunkt möglichst früh im Schuljahr zu terminieren, um wirklich ausreichend Zeit für Diagnose und individuelle Förderung zur Verfügung zu haben und um ggf. vorhandene Schwächen beseitigen und existierende Stärken weiter ausbauen zu können. Derzeit finden entsprechende Planungen mit dieser Intention statt, doch bedingt eine frühere Testung auch frühere Normierungs- und Pilotierungsphasen, so dass eine Umstellung nur allmählich vollzogen werden kann.

17 Die Eingabe der Daten (Schülerergebnisse) ist nur online möglich.

Dies ist korrekt und erleichtert die zeitnahe Auswertung in vielfacher Hinsicht. Jede andere Form ist als Rückschritt zu betrachten, will man eine zeitnahe Rückmeldung erreichen, die – siehe oben – auch eine ausreichende anschließende Zeit der Förderung ermöglicht. Zudem erlaubt diese Datenverarbeitung eine Rückmeldung in verschiedenen Variationen, die auf länderspezifische Anforderungen einzugehen vermag.

Möglich erscheint vielmehr eine Ausdehnung der digitalen Bearbeitung bereits auf den Bereich der Prüfung, d. h. die Schülerinnen und Schüler bearbeiten die Aufgaben am Computer. Sowohl aufgrund nicht ausreichend vorhandenen technischen Equipments (z. B. Anzahl der Computer-Terminals) an den Schulen als auch aufgrund fehlender Qualifikationen der Schülerinnen und Schüler wird davon derzeit aber noch abgesehen.

18 *Viele Kinder, insbesondere solche mit Migrationshintergrund oder aus bildungsfernen Schichten, verstehen die Aufgabenstellungen nicht.*

Diese Problematik bezieht sich nicht auf VERA (allein), sondern ist eine generelle Herausforderung für das deutsche Bildungssystem. Die Aufgabenstellungen bzw. die Items – gemeint ist hier in erster Linie VERA-3 – weichen nicht von vergleichbaren und üblichen Fragen und Aufgaben in Lernmitteln ab (Schulbücher, Arbeitshefte etc.). Insofern scheitern Kinder, die die Testhefte der Vergleichsarbeiten nicht lesen bzw. nicht verstehen können, auch an anderen Texten.

Zudem werden die Aufgabenersteller durch Pilotierung und Normierung bereits im Vorfeld auf eklatante Schwierigkeiten beim Leseverständnis aufmerksam; entsprechende Items können dann – im Unterschied zu Lernmitteln – noch vor dem Testzeitpunkt modifiziert werden.

19 *Bei den Vergleichsarbeiten verfälschen manche Lehrkräfte die Testergebnisse, indem sie Tests vorbereiten oder während der Prüfung Hilfestellungen geben.*

Auch dies ist kein Phänomen, das speziell den Vergleichsarbeiten anzulasten ist, geschweige denn flächendeckend auftritt.

Gegebenenfalls ist hier verstärkt Aufklärungsarbeit insofern zu leisten, als sich Zweck und Nutzen von VERA nur dann erschließen, wenn eine für Prüfungen adäquate Durchführung gewährleistet ist. Andernfalls handelt es sich tatsächlich um eine sinnlose Mehrarbeit, deren Stellenwert gering einzuschätzen ist.

20 Die zentrale, bundesweite Durchführung von Vergleichsarbeiten ist nicht geeignet, individuellen Denk- und Lernprozessen gerecht zu werden.

Im Unterschied zu prüfungsfreien Zeiten und Phasen, in denen individuelles Lernen im Vordergrund steht, muss es Möglichkeiten geben, den persönlichen Lern- und Leistungsstand an objektiven Maßstäben zu messen. Für die Schülerin bzw. den Schüler stellt die Verortung seiner individuellen Leistungsfähigkeit auf einer überindividuellen Skala („Kompetenzstufenmodell“) immer auch eine Möglichkeit dar, sich eigener Stärken und Schwächen bewusst zu werden.

Dabei ist es einsichtig, dass nicht zugleich individuelles Verstehen gefördert und kollektives Wissen abgeprüft werden können (zu unterscheiden ist in diesem Zusammenhang beispielsweise auch eindeutig zwischen Übungs- und Testaufgaben). Verständlich ist aber, dass die Vergleichsarbeiten nicht diesen beiden Zielen dienen: Die Phase individueller Förderung muss den Vergleichsarbeiten ebenso vorausgehen wie sie – nach der individuellen Diagnose durch VERA – wieder fortgesetzt wird.

Zudem empfinden Kinder und Jugendliche „Konkurrenz“ in erster Linie als anregend. Sportliche Aktivitäten und Leistungssteigerungen sind anders gar nicht denkbar. Das Messen an objektiven, anerkannten Normen ist eine Gelegenheit, sich seiner selbst und seiner Möglichkeiten bewusst zu werden. Freilich ist es dafür erforderlich, den Nutzen von Leistung und Leistungsbereitschaft wieder stärker auch im Unterricht zu verankern: So lange Leistungsstarke den Eindruck haben, in erster Linie würden Leistungsschwächere gefördert und Differenzierung würde vor allem bedeuten, Schwächen auszugleichen, so lange besteht die Gefahr, eigene Erfolge als selbstverständlich abzutun oder gar als peinlich zu empfinden.

21 Die Tests sind auf ein mittleres Niveau hin ausgerichtet, leistungsschwächere Schülerinnen und Schüler müssen also notwendigerweise schlecht abschneiden.

Die Ausrichtung der Vergleichsarbeiten an Kompetenzstufenmodellen bedingt, dass eine Verortung der Leistungen von Schülerinnen und Schülern an (international gebräuchlichen) Skalen stattfindet. Dabei beinhalten die Testhefte Items auf verschiedenen Kompetenzniveaustufen, sind also geeignet, verschiedenen Lerntypen und Leistungsvermögen gerecht zu werden. Die Prüfung verschiedener Domänen erlaubt zudem, unterschiedliche Stärken in den Vordergrund zu stellen bzw. sich vorhandener Schwächen bewusst zu werden. Es wäre unverantwortlich, die gesellschaftlichen Ansprüche an Bildung für Schülerinnen und Schüler an den Schwächsten zu messen, ohne diesen die Möglichkeit und Anreize zu geben, sich zu verbessern.

Will man erreichen, dass die Schülerinnen und Schüler nach Abschluss ihrer schulischen Laufbahn auch auf nationaler oder gar internationaler Ebene reüssieren, so kann man sich dem Anliegen nicht verschließen, sie auf ein Arbeitsleben und eine Karriere im Vergleich zu anderen – auch Besseren – vorzubereiten.

Anmerkung:

Eine ausführlichere Version dieser Kritikanalyse, die auf Einzelvorwürfe bzw. teilweise nur auf die spezifische Situation in Bayern eingeht (Pressemitteilungen, Zeitschriftenartikel), kann angefordert werden unter martin.sachse@stmuk.bayern.de.