

Positionspapier

des Netzwerks Empiriegestützte Schulentwicklung (EMSE)

zu:

Zentrale standardisierte Lernstandserhebungen

(5. EMSE-Tagung, Berlin 08.12.2006)

Die Begriffswahl für standardorientierte Überprüfungsarbeiten – „Lernstandserhebung“, „Kompetenztest“, „Orientierungsarbeit“, „Diagnosearbeit“, „Jahrgangsstufentest“ oder „Vergleichsarbeit“ – ist über die Bundesländer hinweg zurzeit nicht einheitlich festgelegt. Einzelne Bundesländer haben hier eine unterschiedlich lange Tradition mit teilweise unterschiedlicher Gewichtung der Funktionen und Verfahrensweisen.

Geplant ist, ab dem Schuljahr 2007/2008 *Vergleichsarbeiten* erstmals bundesweit in der Grundschule (Ende Klassenstufe 3 Deutsch und Mathematik) und ab dem Schuljahr 2008/2009 *Lernstandserhebungen* in der Sekundarstufe I (Mitte / Ende Klassenstufe 8; Deutsch, Englisch / Französisch und Mathematik) bundesweit durchzuführen. Daneben werden auch weiterhin einzelne Länder für sich oder im Verbund mit anderen (z.B. Lernstand 6, Mathematiktest Klasse 8) in unterschiedlichen Jahrgangsstufen Lernstandstests durchführen.

Die Diskussion um die Implementation der bundesweiten Maßnahmen und die praktischen Erfahrungen aus bundeslandspezifischen Projekten haben dazu geführt, sich im Kontext der EMSE grundsätzlich über Funktionen und Verfahrensweisen zu verständigen.

1 Funktionen von Lernstandserhebungen

Die im bundesweiten Netzwerk EMSE (**E**mpiriegestützte **S**chulentwicklung) versammelten Mitarbeiterinnen und Mitarbeiter aus der Bildungsplanung und Schulentwicklung unterstützen nachdrücklich die mit der so genannten „empirischen Wende“ verbundene daten- und wirkungsorientierte Ausrichtung von Qualitätsentwicklung und Qualitätssicherung im Bildungswesen. Dabei werden Messen, Prüfen und Sichern einerseits sowie Entwickeln, Fördern und Unterstützen andererseits als zwei Seiten ein und derselben Medaille betrachtet; beide Seiten sind aufeinander zu beziehen und bilden einen unauflöslichen Zusammenhang.

Zentrale standardisierte Lernstandserhebungen / Vergleichsarbeiten haben folgende Funktionen:

- *auf der Ebene der Schulen:*
 - zusätzliche Informationen über erreichte Lernstände von Schülerinnen und Schülern und die Wirksamkeit der schulischen Arbeit
 - stärkere kriteriale Orientierung der Lehrkräfte und Schulen an den Standards (Lehrplänen und Richtlinien) bzw. an den durch die Lernerfolgsüberprüfung indirekt definierten Standards (z.B. Kompetenzniveaus)
 - Orientierung für Lehrkräfte und Schulen, wo ihre Schülerinnen und Schüler im Hinblick auf definierte Anforderungen und im Vergleich zu anderen stehen (Verortung und Rechenschaftslegung)
 - Grundlage für weitergehende Lernbedarfsdiagnosen
 - Rückmeldungen und Dialoge über Lernen und Lernentwicklungen mit Eltern und Schülerinnen und Schülern („Feedback-Kultur“)
 - Weiterentwicklung des Unterrichts (Reflexion von Zielen, Inhalten und Methoden auf der Grundlage von empirisch ermittelten Kompetenzniveaus und Vergleichswerten; ggf. Revision schuleigener curricularer Schwerpunktsetzungen; Revitalisierung der Fachgruppen-

und Fachkonferenzarbeit; Anregungen durch Testaufgaben für die Erarbeitung von Lernaufgaben; Entwicklung schuleigener Prüfungs- und Klassenarbeiten)

- *auf der Systemebene:*
 - Informationen über Leistungen des Bildungssystems (u. a. hinsichtlich bestimmter Lernbereiche, Schulformen, Schulstufen sowie spezifischer Schülergruppen)
 - Gewinnung von Planungs- und Steuerwissen (insbesondere Folgerungen für die Lehrplan-/Standardentwicklung und Schulsystementwicklung)
 - Identifizierung von Unterstützungs- und Entwicklungsbedarf (Ausbildung, Fortbildung, regionalisierter Support)

2 Grundsätzliche Überlegungen zur Vorgehensweise

- a) Zentrale Lernstandserhebungen/Vergleichsarbeiten überprüfen Lernergebnisse im Hinblick auf Standards, wie sie in den *länderübergreifenden Bildungsstandards* bzw. den daran gekoppelten verbindlichen *Anforderungen/Kerncurricula* der Länder formuliert sind. Sie zielen darauf, nach vorgegebenen Aufgaben und Beurteilungsmaßstäben Aussagen über die zu fest definierten Zeitpunkten von Schülerinnen und Schülern erreichten Lernergebnisse und damit über erreichte Kompetenzniveaus machen zu können.
- b) Zentrale Lernstandserhebungen/Vergleichsarbeiten beziehen sich auf umfassendere Lern- bzw. Kompetenzentwicklungen (bisher erreichte Fertigkeiten, Fähigkeiten, Kenntnisse) und nicht wie Klassenarbeiten auf das unmittelbar vorangegangene Unterrichtsgeschehen. Es sind damit kompetenzorientierte Tests, die im Vorfeld im Hinblick auf ihre Skalierbarkeit hinreichend erprobt sind (Präpilotierung, systematische Pilotierung).
- c) Der Schwerpunkt des Vergleichs zielt auf die *kriterielle Orientierung* (Wie verhält sich die erbrachte Leistung bezogen auf das, was Schülerinnen und Schüler am Ende der Klassenstufe ... gelernt haben sollten? Welches Kompetenzniveau wird in welchen fachlichen Teilbereichen erreicht?).
- d) Durch die zentrale Auswertung einer repräsentativ angelegten Stichprobe oder der Vollerhebung wird der Bezug zu einer schulübergreifenden Norm herstellbar (Benchmarking; Kontextuierung von Schulen und Zuordnung zu Gruppen mit ähnlichen Bedingungen und Ausgangslagen; Vergleichsmöglichkeit: Welche Teilleistungen / Ergebnisse haben Klassen / Kurse in vergleichbaren Schulen / im Landesschnitt / in anderen Schulformen ... ?).
- e) Die Leistungsergebnisse von Schülerinnen und Schülern sind in hohem Maße von Kontextbedingungen abhängig. Diese sind bei Vergleichen zu berücksichtigen, will man zu aussagefähigen Ergebnissen gelangen („faire Vergleiche“). Solange entsprechende elaborierte Verfahren (etwa Indizes auf Schülerindividualebene) noch nicht vorliegen, ist hilfsweise auf Schülerangaben oder auf Einschätzungen der Schulen zurückzugreifen. Im Grundschulbereich könnten - solange Schuleinzugsbereiche klar definiert sind - Daten des jeweiligen Einzugsbereichs Berücksichtigung finden.

Perspektivisch müssen Verfahren zur Berücksichtigung der Kontextbedingungen weiterentwickelt werden (Sozialindizes für Schulen), damit Schulen ihre Vergleichswerte angemessen einschätzen können (value-added-Perspektive).

- f) Als kompetenzorientierte Tests, die nach den Grundsätzen der klassischen und probabilistischen Testtheorie angelegt und ausgewertet werden, geben zentrale Lernstandserhebungen/Vergleichsarbeiten – unter Berücksichtigung quantifizierbarer Messfehler – hinreichend zuverlässige Informationen zum erreichten Lernstand in Teilbereichen der Fächer auf der Ebene von Klassen/Kursen, auf der Ebene von Jahrgangsstufen und auf der Ebene von Schulformen/Bildungsgängen.

- g) Lernstandserhebungen/Vergleichsarbeiten sind kein originäres Instrument der Individualdiagnostik. Auf Individualebene sind die skalierten Ergebnisse deutlich messfehlerbehaftet, wobei die Messfehler in den Randbereichen des Leistungskontinuums besonders zum Tragen kommen. Die einzelschülerbezogenen Ergebnisse von Lernstandsbestimmungen sind somit mit Unsicherheit behaftete Schätzungen, die einen Aussagewert auf Individualebene erst im Kontext weiterer diagnostischer Verfahren bekommen.
Ergebnisse einzelner Schülerinnen und Schüler können für die Berichtsfunktion von Zeugnissen genutzt werden, wenn (a) hinreichende unterrichtliche Lerngelegenheiten für die Bewältigung der Testanforderungen und (b) diagnostisch begründbare Verknüpfungen zu den unterrichtlichen Leistungsfeststellungen gegeben sind.
- h) Bei den Lernstandserhebungen/Vergleichsarbeiten werden alle Testaufgaben zentral entwickelt. Die Auswertung und Beurteilung liegt – unterstützt durch standardisierte Manuale (Kriterien, schwierigkeitsbestimmende Parameter, typische Fehler ...) – bei den Schulen.
- i) Bei der Testentwicklung und Testkommentierung müssen fachdidaktische Perspektiven berücksichtigt und fachdidaktische Expertise einbezogen sein.
Im Interesse einer standardorientierten Unterrichtsentwicklung ist zusätzlich die Veröffentlichung der Testaufgaben nach der Durchführung der Lernstandserhebungen / Vergleichsarbeiten mit fachdidaktischen Hinweisen zur Weiterarbeit sinnvoll und notwendig.
- j) Für eine sinnvolle unterrichtsbezogene Verwertung der Leistungsergebnisse sind zentrale standardisierte Lernstandserhebungen durch alltagspraktische, curriculumbezogene „diagnostische Materialien“¹ zu ergänzen.
- k) Das Instrument Lernstandserhebungen/Vergleichsarbeiten ist primär adressiert an eigenverantwortliche bzw. selbstständige Schulen und an die Professionalität dort unterrichtender Lehrkräfte. Insofern Schulen bzw. Lehrpersonen bei der Auswertung der Lernstandserhebungen und bei der Umsetzung von Konsequenzen fachkundigen Rat und begleitende Unterstützung benötigen, bedarf es eines darauf ausgerichteten Unterstützungssystems.

3 Lernstandserhebungen im Kontext der Diskussion um die Veröffentlichung von Leistungsdaten

Die Einführung von zentralen standardisierten Lernstandserhebungen ist ein Ausdruck der „empirischen Wende“ in der Bildungspolitik und Bildungsplanung. Diese Wende stößt in der Schulpraxis auf Vorbehalte. Ihr ist durch ein gleichermaßen aufklärerisches wie behutsames Vorgehen zu begegnen. Dabei dürfen berechnete Befürchtungen eines Daten-„shaming“ einerseits und Fragen nach den curricular-didaktischen Verknüpfungen andererseits nicht ignoriert werden.

Schulleistungen sind zu einem erheblichen Teil auf kognitive, soziale, kulturelle und ökonomische Voraussetzungen von Schülerinnen und Schülern bzw. ihrer Elternhäuser zurückzuführen. Dementsprechend ist die Variation schulischer Testleistungsergebnisse zu einem erheblichen Anteil auf außerschulische Faktoren zurückzuführen.² Insofern sind globale Vergleiche zwischen Schulen wenig aussagekräftig. Dies gilt umso mehr für „Rankingverfahren“, die bislang die schulischen Kontextbedingungen und individuellen Lernausgangslagen unberücksichtigt lassen.

Rankings sind wenig ergiebig oder gar kontraproduktiv, weil sie die Lehr- und Lernbedingungen außer Acht lassen und damit die pädagogischen Anstrengungen von Kollegien bzw. Lehrpersonen nicht hinreichend beachten. Schulische Rankings sind ohne Berücksichtigung der Kontextvariablen zu einem erheblichen Anteil Ausdruck der Schülerzusammensetzung einer Schule; nichtadjustierte Rankings ge-

¹ Vgl. z. B. die in verschiedenen SINUS-Ländervorhaben eingesetzten curriculumbezogenen diagnostischen Materialien für das Fach Mathematik aus Schweden und Australien / Neuseeland (vgl. z.B.: <http://www.learn-line.nrw.de/angebote/sinus/projekt5/index.html>).

² Jaap Scheerens & Roel Bosker (1997, 2005): The foundations of educational effectiveness. Oxford u. a.: Pergamon.

ben deshalb keine Auskunft über Schulqualität. International ist man daher dazu übergegangen, zum einen Kontextvariablen in das Ranking einzubeziehen und zum anderen die Kompetenzzuwächse einer Schule als Qualitätsindikator zu verwenden.

Eine entscheidende Rolle für die Akzeptanz und für den Umgang mit den Daten in der Schulpraxis spielt das Ausmaß ihrer Veröffentlichung. Wegen der noch bestehenden Vorbehalte sollten Ergebnisse aus Lernstandserhebungen auf Schul-, Klassen- und Individualebene zunächst nur für den dienstinternen Gebrauch und pädagogischen Bedarf verwendet werden. In der Perspektive einer eigenverantwortlichen Schule sind jedoch die pädagogischen Leistungen einer Schule öffentlich zu verantworten („Freiheit und Verantwortung“).

In einem internationalen OECD-Symposium kamen bereits 1995 die dort versammelten Experten verschiedener Länder zu folgender Bewertung von standardisierten Leistungsmessungen:³ Nur dann, wenn Leistungstests bestimmte Bedingungen erfüllen, können sie eine pädagogisch sinnvolle Ergänzung zu den bereits bestehenden Qualitätssicherungsmaßnahmen darstellen und einen produktiven Beitrag zur Qualitätssicherung und Qualitätsentwicklung von Unterricht leisten: wenn sich diese Tests nicht nur auf Kenntnisse, sondern auch auf fachspezifische Fähigkeiten und Fertigkeiten beziehen, wenn sie ferner nicht für ungerechtfertigte Disziplinierungsmaßnahmen missbraucht werden, wenn die summativen Verfahren mit formativen Verfahren kompatibel sind, wenn diese Tests mit einer diagnostischen Funktion für die einzelne Lehrperson eng verkoppelt werden, wenn die Tests von der Lehrerschaft als hilfreich erlebt und akzeptiert werden und wenn es Interpretationshilfen, substantielle Beratungs- sowie Fortbildungsangebote gibt.

Die Vorlage zu diesem Positionspapier wurde von Rainer Peek, Ulrich Steffens und Olaf Köller erarbeitet, auf der 5. EMSE-Tagung in Berlin am 8. Dezember 2006 diskutiert und mit kleineren Änderungen (insbesondere auf Vorschlag von Karl-Heinz Arnold, Christoph Burkard, Peter Dobbstein und Peter May) verabschiedet.

³ Paul Black & J. Myron Atkin (Ed.) (1996): Changing the Subject. Innovations in Science, Mathematics and Technology Education. London & New York: Routledge, Chapter 4.
Vgl. dazu aktuell: OECD (2005): Informal meeting of Ministers of Education (28-29 April 2005). School accountability and transparency of information. Summary record. St. Gallen, Switzerland.