

Hans Brügelmann

Tests statt Noten?

Warum PISA, VERA & Co kein Modell für die Leistungsbeurteilung von SchülerInnen sein können¹

Zu Beginn ein kurzer Überblick über meine Argumentation im folgenden Beitrag:

- Die Not mit den Noten: kurze Wiederholung einer lange bekannten Kritik
- Tests statt Noten? Überschätzung der Leistungsfähigkeit eines neuer „Allzweck-Instruments“
- „PISA, bleib bei deinem Leisten!“: Was beim System Monitoring hilfreich ist, muss deshalb nicht schon für die Schulevaluation oder gar für die Beurteilung individueller Leistungen geeignet sein
- Stärkung der Evaluationskompetenz vor Ort: Die Prozess-Standards der „Blick-über-den-Zaun“-Schulen und das Konzept „Pädagogische Leistungskultur“ des Grundschulverbands
- Verbalgutachten - eine Alternative zur Ziffernnoten? Eine Änderung der Form bei gleicher Funktion reicht nicht

Fazit: Notwendig ist ein System- und Mentalitätswechsel - von der Auslese zur Förderung!

*

* *

¹ Vortrag auf dem Symposium der Software AG -Stiftung „Aus PISA gelernt?“ am 24.11.2006 in Berlin; vgl. vertiefend die Kap. 46-48, 56-62 in: Brügelmann, H. (2005): Schule verstehen und gestalten - Perspektiven der Forschung auf Probleme von Erziehung und Unterricht. Libelle: CH-Lengwil (fortlaufend aktualisiert unter: www.agprim.uni-siegen.de/schuleverstehen).

Vor einigen Monaten haben wir für den Grundschulverband ein Gutachten erarbeitet zu der Frage: „Sind Noten nötig - und nützlich?“². Es gab zwar schon in den 1980er und 1990er Jahren einige vorsichtige Schritte in Richtung entwicklungsorientierter Verbalgutachten. Aber parallel - und zum Teil unter Berufung auf - PISA erleben wir zurzeit einen *roll-back*. Und das, obwohl alle Argumente gegen Noten - sowohl gegen ihre Funktion als auch gegen ihre Form - seit langem bekannt sind. Sie sind jetzt im Gutachten noch einmal durchgehend bestätigt wurden.

Die Not mit den Noten

Ich will die wichtigsten Ergebnisse unserer Analysen deshalb - sozusagen zur Erinnerung - nur in ein paar Stichworten zusammenfassen: Ziffernnoten suggerieren eine technische Präzision der Leistungsbeurteilung, sie sind aber....

- nicht valide, d. h. nicht vorhersagekräftig bezogen auf zukünftigen Schul-, Ausbildungs- und Berufserfolg;
- nicht verlässlich, d. h. nicht stabil, sondern stark beeinflusst von äußeren Umständen;
- nicht objektiv, d. h. nicht unabhängig von den beurteilenden Personen;
- nicht vergleichbar, da sich LehrerInnen auf sehr unterschiedliche Maßstäbe und Schwellenwerte beziehen;
- nicht fair wegen der meist als Norm unterstellten Normalverteilung für jede Klasse, obwohl die Leistungsniveaus breit streuen.

Noten erfüllen also schon die eigenen Versprechen nicht - von externer Kritik, auf die ich am Ende noch zu sprechen komme, ganz zu schweigen.

² Arbeitsgruppe Primarstufe (2006): Sind Noten nützlich und nötig? Zifferzensuren und ihre Alternativen im empirischen Vergleich. Eine wissenschaftliche Expertise der Grundschulverbandes, erstellt von der Arbeitsgruppe Primarstufe an der Universität Siegen (Hans Brügelmann mit Axel Backhaus u. a.). Grundschulverband e.V.: Frankfurt. Weitere Informationen → <http://www.agprim.uni-siegen.de/notengutachten.htm>

Seit vielen Jahren sind diese Befunde bekannt, aber gegen eine Abschaffung von Noten gab es heftigen Widerstand³ wegen

- der Ökonomie ihrer Erstellung für LehrerInnen
- ihrer Akzeptanz bei SchülerInnen und Eltern
- ihrer Symbolkraft für „Leistung“ in der Öffentlichkeit.

Der bildungspolitische Kontext hat sich allerdings verändert durch „PISA & Co“ - d. h. durch die Verbreitung (inter)nationaler Testprogramme.

Tests statt Noten?

Wenn wir wissen wollen, ob PISA & Co. Modell für eine Alternative zu Ziffernnoten sein können, müssen wir bei der Argumentation zwei Ebenen und Schritte unterscheiden:

- Wo liegen die Stärken und Schwächen der PISA-Studie in ihrer Funktion als System-Evaluation? (nur kurz)
- Lässt sich das Instrumentarium einer System-Evaluation ohne Weiteres für die Individualdiagnose nutzen?

Als positive Erträge und damit als Stärken von PISA verbuche ich

- Einführung eines System-Monitoring
- Orientierung auf Ergebnisse von Unterricht (aber als „output“?)
- öffentliche Aufmerksamkeit für Schule
- Konfrontation mit verdrängten Tatbeständen:
 - mit einem Drittel Sitzenbleiben usw. hohe Selektivität
 - extreme Schwierigkeiten von MigrantInnen
 - starke Schichtabhängigkeit des Schulerfolgs
 - fachbezogene Geschlechterdifferenzen

³ Das gilt allerdings nicht für den Grundschulbereich und vor allem nicht für Schulen, in denen eine entwicklungsorientierte Verbalbeurteilung schon etabliert und in ein entsprechendes Konzept von Unterricht eingebettet ist.

Diese bildungspolitische Wirksamkeit von PISA ist nicht gering zu schätzen. Das Irritierende ist aber, dass die wesentlichen Befunde aus dieser Studie inhaltlich vorher schon bekannt waren und dass man viel Differenzierteres durch eine (auch preiswertere) Befragung von LehrerInnen, Schulaufsichtsbeamten usw. hätte erfahren können. Aber erst durch die öffentlichkeitswirksame Darstellung der Scheinpräzision von Testergebnissen sind diese Botschaften wahrgenommen, sind die Probleme in einem kommunikationstheoretischen Sinn „real“ geworden.

Dabei ist PISA selbst nicht unproblematisch⁴. Ich nenne nur drei Punkte, die verdeutlichen, dass auch Statistiken in hohem Maße interpretationsbedürftig sind:

- die internationale Vergleichbarkeit der Stichproben durch unterschiedliche Teilnahmequoten (vgl. etwa Niederlande, UK, USA und das viel diskutierte Südtirol) und von Untergruppen (z. B. „MigrantInnen“ in Canada vs. Deutschland , in Bayern vs. Nordrhein-Westfalen)
- die Repräsentativität im unteren Leistungsbereich, insbesondere durch unterschiedliche Drop-out-Quoten in den verschiedenen Schulsystemen
- die Auswahl statistischer Verfahren, in die jeweils eine Reihe von Annahmen eingehen, die durchaus kontrovers diskutiert werden.

Noch wichtiger für unser Thema:

- Die Beschränktheit der (wegen der Standardisierung möglichen) Aufgabentypen - und die dennoch unvermeidliche Mehrdeutigkeit ihrer Lösungen⁵.

Nun kann man sagen: Ist doch gut, wenn PISA dazu beiträgt, dass die richtigen Botschaften auch verbreitet werden. Andererseits gibt es aber auch bedenkliche Nebenwirkungen: Indem die Alltagserfahrung die Plausibilität vieler PISA-Befunde bestätigt, stärkt sie auch die Autorität der Instrumente, mit denen diese gewonnen

⁴ Vgl. die Kritik u. a. von: Wuttke, J. (2006): Fehler, Verzerrungen, Unsicherheiten in der PISA-Auswertung. In: Jahnke, T./ Meyerhöfer, W. (Hrsg.) (2006): PISA & Co - Kritik eines Programms. Franzbecker: Hildesheim (S. 101-154):

⁵ Vgl. zur inhaltlichen Kritik der Aufgabentypen in standardisierten Tests u. a. Barnitzky, H. (2004): VERA Deutsch 2004: Ungeeignet und bildungsfern. In: Grundschulverband aktuell, H. 89, 10-16. Rumpf, H. (2006): Was greifen und messen sie wirklich? Zwei PISA-Aufgaben auf dem Prüfstand. Mit einem Postscriptum von Peter Buck. In: Erziehungskunst, 70. Jg., Sonderheft - Oktober 2006, 14-22.

wurden. Die Tests und statistischen Auswertungen scheinen die eigentlichen Quellen der Erkenntnis zu sein, während es doch die Intelligenz und das profunde Wissen der Forscher ist, das aus den Testwerten und ihrer Verrechnung in Korrelationen usw. inhaltlich überzeugende Botschaften macht. In der öffentlichen Wahrnehmung kommt an: Tests sind klüger als Erfahrung - und sei sie noch so professionell.

Damit keine Missverständnisse entstehen: Ich gehöre zu denen, die LehrerInnen ausdrücklich nahe legen, normierte Tests in ihr Repertoire der Leistungserfassung einzubeziehen. Deshalb zunächst der Hinweis auf vier Vorteile, die ich als besonderes Potenzial von Tests sehe:

- Fokussierung der Datenerhebung
- Transparenz der Anforderungen
- Kalibrierung der Maßstäbe (Bezug auf Normstichproben)
- Aufdecken blinder Flecken in Unterricht und Beurteilung.

Aber Tests können das Lehrerurteil nicht ersetzen.

- sie müssen sich auf bestimmte Aufgabentypen und Auswertungsformen beschränken („Standardisierung“);
- sie konzentrieren sich inhaltlich auf wenige Ausschnitte des Leistungsprofils (Messbarkeit; Ökonomie);
- ihre Ergebnisse sind durch bei punktueller Erhebung ebenfalls nicht verlässlich;
- ihre prognostische Validität ist ähnlich unsicher wie die der Noten;
- auf Individualebene ist der Messfehler, der sich in großen Stichproben ausgleicht, zu groß, um biografisch nachhaltig wirksame Entscheidungen über die Bildungslaufbahn von SchülerInnen zu rechtfertigen.

Ich fürchte sowieso, dass bei der Durchsetzung der internationalen Leistungsvergleiche Hoffnungen geweckt worden sind, die die standardisierte Leistungsmessung nicht erfüllen kann. Ich halte das für sehr problematisch.

Wir müssen lernen: Statistiken sprechen nicht für sich. Auch die produktiven Impulse aus TIMSS, PISA und IGLU leben nicht von den Zahlen allein, sondern von der Lauterkeit und Klugheit der Leute, die sie interpretieren. Das heißt dann auch: Wenn andere Personen dieselben Zahlen interpretieren, können sie zu ganz anderen Schlüssen kommen⁶. Diese Personabhängigkeit der Folgerungen wird bei uns noch nicht zureichend verstanden.

Die Begrenztheit und die Interpretationsbedürftigkeit von Testwerten muss LehrerInnen und Eltern deutlich gemacht werden - dann, aber auch *nur dann* können auch Tests das diagnostische Repertoire sinnvoll ergänzen. Es ist eine bedenkliche Fehlentwicklung, wenn jetzt z. B. Computerprogramme aus dem Boden sprießen, die aufgrund eines einmaligen Tests versprechen, einen Satz von Arbeitsblättern als „individualisiertes Förderprogramm“ für ein ganzes Jahr auszuwerfen.

Systemisch denken: PISA, bleib bei deinem Leisten!

Wenn Menschen etwas Neues erfinden und Erfolg damit haben, neigen sie dazu, es über den ursprünglichen Anwendungsbereich hinaus zu nutzen. Wer einen Hammer hat, dem wird die Welt zum Nagel, wer einen Schraubenzieher in der Hand hat, findet überall etwas, woran man schrauben kann...

PISA & Co haben viel in Bewegung gebracht. *Gedacht* sind die Studien für eine Evaluation des Bildungssystems insgesamt und sie sind - bei allen Mängeln⁷ - hilfreich für die Bildungspolitik. Aber *gedeutet* und *genutzt* werden sie von vielen für die Evaluation von Unterricht. Nicht nur in den Kultusverwaltungen glauben viele, mit landesweiten Testprogrammen auch eine Lösung der Rechenschaftsprobleme auf Schul- und Lehrer-Ebene und sogar ein verlässliches Mittel der Lerndiagnose auf Schüler-Ebene gefunden zu haben. Testprogramme wie VERA sind sogar explizit mit diesem individualdiagnostischen Anspruch verkauft worden und haben ihn erst später revidiert.

⁶ Ich erinnere an meinen alten Vorschlag, dass die Ergebnisse von Studien aus der Perspektive „stellvertretender LeserInnen“ unterschiedlicher Disziplinen und Positionen vorgestellt werden sollten, um diese Personabhängigkeit der Interpretation auch transparent zu machen.

⁷ Vgl. zu den Problemen die aktuelle Diskussion im Anschluss an die Beiträge (vor allem von Joachim Wuttke) in Jahnke, T./ Meyerhöfer, W. (Hrsg.) (2006): PISA & Co - Kritik eines Programms. Franzbecker: Hildesheim.

Dagegen setze ich die Forderung: PISA, bleib bei deinem Leisten!

Landesweite Testprogramme für das System-Monitoring sind grundsätzlich sinnvoll - aber

- es reichen Stichproben,
- die Abstände können länger sein,
- eine Bindung an Sanktionen ist kontraproduktiv (keine „high stakes tests“⁸).

Daneben können Tests die Leistungsfeststellung im Klassenzimmer bereichern - aber eben nur im Rahmen eines breiteren Repertoires.

Wir brauchen ein differenziertes System der Rechenschaftslegung im Bildungswesen, Der Grundschulverband hat bereits 1999 dafür einen konkreten Vorschlag unterbreitet⁹

Wichtig daran erscheint mir die Differenzierung unterschiedlicher Funktionen und Formen der Evaluation¹⁰. Wir brauchen nämlich zwei verschiedene Typen von Evaluation¹¹:
- *fallübergreifend* für Politik und Verwaltung: statistische Aussagen zu wenigen Variablen über viele Fälle, Verdichtung in Mittelwerten und Gruppenaussagen, Betonung von Konstanten;

- *fallbezogen* für Schulen und LehrerInnen: viele Merkmale des Einzelfalls erfassend, interessiert an Streuungen und offen für Veränderungen.

Man kann die Struktur eines solchen Rechenschaftssystems in verschiedenen Dimensionen präzisieren. Aktuell geht es um die Stärkung jeweils des ersten Pols in den folgenden Dimensionen:

- lokal vs. zentral

⁸ Mit diesem Begriff werden in den USA Testprogramme bezeichnet, deren Ergebnisse für SchülerInnen, LehrerInnen und/ oder Schulen bzw. Schulbezirke Folgen haben wie die Zulassung zu Bildungsgängen, Gehaltszulagen oder die Zuweisung von finanziellen Zuschüssen.

⁹ Bartnitzky, H., H., u. a. (1999): Zur Qualität der Leistung - 5 Thesen zu Evaluation und Rechenschaft der Grundschularbeit. Grundschulverband - Arbeitskreis Grundschule e. V.: Frankfurt. Auch in: Schmitt, R. (Hrsg.) (1999): An der Schwelle zum dritten Jahrtausend. BundesGrundschulKongress 1999. Grundschulverband - Arbeitskreis Grundschule: Frankfurt, 165-196.

¹⁰ Vgl. dazu bereits 1980 mein Gutachten für das *Basic Education Project* der OECD - Auszug unter: www.agprim.uni-siegen.de/printbrue/brue.oecd.evaluation.pdf

¹¹ Brügelmann, H. (2005i): Von Fall zu Fall. Plädoyer für einen Miss-Marple-Stil in der Bildungsforschung Miss Marple. Publiziert im: Forum Kritische Pädagogik (4.10.2005) □ <http://forum-kritische-paedagogik.de/start/download.php?view.15>

- kontext-sensitiv vs. kontext-neutral
- begleitend vs. punktuell
- entwicklungsorientiert vs. vergleichend
- person- statt gruppenbezogen.

Der letzte Punkt ist besonders heikel und wichtig. PISA & Co haben einen fatalen Nebeneffekt gehabt: Die Aufmerksamkeit der LehrerInnen wird auf Kinder als *Mitglieder von Gruppen* fokussiert. Was auf Systemebene eine Differenzierung bedeutet - Aufschlüsselung der Daten nach Untergruppen - erweist sich auf Unterrichtsebene als irreführende Pauschalisierung:

- „Jungen brauchen...“
- „Migrantenkinder haben...“
- „Die unteren 25% können nicht...“

PISA & Co haben insofern - wenn auch ungewollt - das Denken in Grobkategorien gestützt: Diese bedeuten für den Alltag Etikettierung. Auch Lehrer reden über ihre „Jungen“ statt über Ben und Marc und über „Kinder mit Migrationshintergrund“ statt über Reyhan und Nassrin. Inhaltlich lenken Aussagen über Mädchen vs. Jungen, Kinder mit vs. ohne Migrationshintergrund, Unter- vs. Oberschicht ab von der Notwendigkeit, das einzelne Kind genauer zu betrachten. Die große Streuung *innerhalb* der Gruppen - viel größer als die Differenzen zwischen ihnen - macht darauf aufmerksam, dass der Blick auf die Besonderheit des einzelnen Kindes gestärkt werden muss.

Unterricht und Erziehung sind nämlich mehr als die bloße *Anwendung* von allgemeinen Erkenntnissen und Programmen. Sie verlangen deren Übersetzung und Anpassung auf die jeweils spezifische Situation

Denn was fängt ein Lehrer mit einer Erkenntnis an wie: 60% der Schüler mit Merkmal X hilft es, wenn man Methode Y anwendet, für 40% ist Methode Z besser? Über seine Laufbahn hinweg mag es für einen Lehrer klug sein, immer Methode Y anzuwenden. Dann hat er bis zum Ende seiner Pensionierung weniger Fehler gemacht als ein Lehrer, der immer Methode Z angewendet hat.

Aber was bedeutet das alles für den Einzelfall: Was also macht der Lehrer konkret, wenn er einen Schüler mit Merkmal X hat? Hier ist zusätzliche Information nötig. Zum Glück

verfügen LehrerInnen nicht nur über allgemeines, explizites Wissen, sondern auch über Können, das in *impliziter* Erfahrung wurzelt¹².

Und: Sie denken nicht nur deduktiv, sondern auch *analog*. Der Schüler ist nicht nur einer, der das Merkmal X hat. Er ist auch „wie Tobias“ und „anders als Ben“.

Dieses Potenzial müssen wir nutzen. Und wir müssen es kultivieren, statt Lehrer abhängig zu machen vom Expertentum von SpezialistInnen, das der Vielfalt des je besonderen Einzelfalls nicht gerecht werden kann.

Und ein zweites: PISA privilegiert die Mächtigen. Gefüttert wird die zentrale Entscheidungskompetenz. Mein Plädoyer: PISA war ein Fortschritt in der Entwicklung des System-Monitoring. Auf einem Bein kann unser Rechenschaftssystem nicht stehen. Wir brauchen jetzt einen vergleichbaren Schub für die Lernbeobachtung im Unterricht - aber *nicht* als Kopie von PISA, sondern *eigenständig* und ausgerichtet an den besonderen Bedürfnissen der Schulentwicklung¹³ und einer individuellen Förderdiagnose. Letzteres ist die Zielsetzung des Projekts „Pädagogische Leistungskultur“ des Grundschulverbands¹⁴.

Stärkung der Evaluationskompetenz vor Ort

Wir haben Konzepte, Verfahren und Materialien entwickelt, die LehrerInnen helfen sollen, die Evaluationsanforderungen des Unterrichtsalltags besser zu bewältigen, d. h. Beobachtung zu systematisieren und stärker auf Fördermöglichkeiten zu beziehen.

Aufgaben zur Lernbeobachtung sollen...

- der Lehrperson Informationen geben...
 - nicht nur über den aktuellen Stand von Einzelleistungen,
 - sondern auch über die angewandten Strategien („Tiefenstrukturen“)
 - und über deren Entwicklung („Lerngeschichte“);

¹² Neuweg, G. H. (2006): Das Schweigen der Könner. Strukturen und Grenzen des Erfahrungswissens. Trauner Verlag: Linz.

¹³ Vgl. dazu die Standards der „Blick-über-den-Zaun“-Schulen → www.blick-ueber-den-zaun.de

¹⁴ Bartnitzky, H., u. a. (Hrsg.) (2005&2006): Pädagogische Leistungskultur: Materialien für Klasse 1/2 und Klasse 3/4. Beiträge zur Reform der Grundschule, Bd. 119 & 121. Grundschulverband: Frankfurt.

- dialogisch angelegt sein als wechselseitige Verständigung über Lernziele, Bewertungskriterien und tatsächliche Leistungen und damit
- die Fähigkeit der Kinder zur Selbsteinschätzung eigener Arbeiten entwickeln; sowie zugleich
- für die Kinder auch inhaltlich eine produktive Lernsituation darstellen.

Beispiele finden sich für die zentralen Lernbereiche und alle vier Jahrgangsstufen in den bereits erwähnten Materialien des Grundschulverbands.

Wir versuchen das zu erreichen durch eine Verbindung von

- Fremdbeurteilung durch die Lehrperson, z. B. über
 - punktuelle Erhebungen (standardisiert & informell)
 - begleitende Beobachtungen (strukturiert & beiläufig anfallend aus produktiven Lernsituationen)

und

- Selbstbeurteilung der Kinder, z.B. über
 - Portfolios: „Meine besten Arbeiten“
 - Selbstzeugnisse: „Das kann ich gut, das muss ich noch besser machen.“
 - Lerntagebücher: „Wie habe ich das gelernt?“
 - Lernverträge: „Das will ich als nächstes lernen.“

Damit stellt sich die Frage nach der zweiten eingangs erwähnten Option:

Verbalgutachten statt Noten?

Wenn nicht Tests - können dann sprachlich ausformulierte Gutachten die Schwächen von Noten überwinden?

In unserem Gutachten haben wir das besondere Potenzial von verbalen Beurteilungen in vier Punkten zusammengefasst¹⁵:

- Differenzierung der Beschreibung (Leistungsprofil statt Ziffer)
- Kontextbezug bei der Erklärung (systemisch statt personalisiert)
- Bewertung nach der Entwicklungsnorm (Lernzuwachs statt Leistungsvergleich)
- Förderorientierung der Folgerungen (nächste Schritte statt bloßem Lernstand)

Die Studien zur aktuellen Praxis dieses Ansatzes sind aber ernüchternd: Auch Verbalbeurteilungen realisieren ihre Ansprüche nur sehr eingeschränkt¹⁶. Insbesondere werden die Leistungen nur selten inhaltlich differenziert und auf die Person bezogen beschrieben und auch der entwicklungsbezogene Bezugsnorm wird nicht oft realisiert

Zudem ist die Frage: Was bringen...

Veränderte Formen bei gleicher Funktion?

Auch ein Wechsel zu Verbalbeurteilungen, also eine Veränderung der Darstellungsform verspricht nur begrenzte Verbesserung der Situation¹⁷. Notwendig erscheinen mir

- ein Mentalitätswandel
und
- eine Strukturveränderung.

Solange die Selektionsfunktion dominiert, helfen auch Veränderungen der Formen von Leistungsbeurteilungen wenig¹⁸. Umgekehrt gilt aber auch: Rangziffern sind nur so lange nötig, wie wir Kinder sortieren...

¹⁵ Arbeitsgruppe Primarstufe (2006, Kap. 2 und 3).

¹⁶ Arbeitsgruppe Primarstufe (2006, Kap. 3.1).

¹⁷ Arbeitsgruppe Primarstufe (2006, Kap. 7).

- bei der Versetzung (vs. Sitzenbleiben)
- bei der Entscheidung über den Besuch von Regel- vs. Sonderschule
- bei den Laufbahneempfehlungen für Gymnasium vs. Real- vs. Hauptschule.

Mit *Mentalitätswandel* meine ich, dass sich Leistungsbeurteilung von der Bewertung der Person („du kannst nichts“, „du bist faul“, „du bist dumm“) zur Sachinformation entwickeln muss: „Das kannst du gut, das hast du dazugelernt, aber da hast du Schwierigkeiten. Hieran müssen wir - gemeinsam - arbeiten.“ Ein solcher Mentalitätswandel ergibt sich nicht schon aus neuen Strukturen und veränderten Beurteilungsformen. Empirische Studien verweisen auf die Notwendigkeit, einen regelmäßigen Erfahrungsaustausch zu institutionalisieren, um die Kriterien und ihre Abwendung anhand konkreter Fälle immer wieder neu abzustimmen¹⁹.

¹⁸ Vgl. auch Schlömerkemper, J. (2006): Qualitätsentwicklung durch Kompetenz- und Prozessorientierung - eine Alternative zu selektionsorientierter Lernorganisation. In: Ferdinand Eder, Angela Gastager, Franz Hofmann (Hrsg.) (2006): Qualität durch Standards? Beiträge zum Schwerpunktthema der 67. Tagung der AEPF. Waxmann: Münster, S. 103-111.

¹⁹ Vgl. Arbeitsgruppe Primarstufe (2006, Kap. 1.2.3 und 7).