

Formen und Inhalte von Rasterzeugnissen: Eine Untersuchung der strukturierten Berichtszeugnisformulare in den Grundschulen des Bundeslandes Bremen

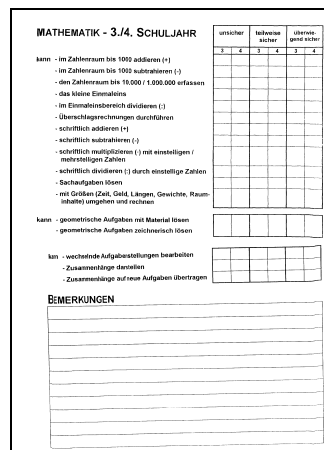
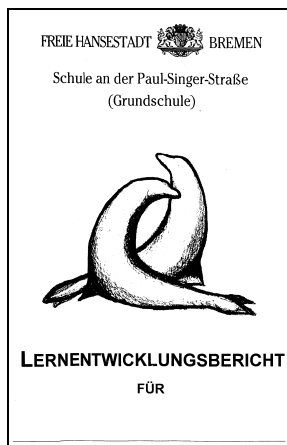
von Dr. Karl-Heinz Arnold, Schulpsychologischer Dienst Bremerhaven, 30.11.1997

Zusammenfassung

Untersucht werden die im Schuljahr 1996/97 an den Grundschulen im Bundesland Bremen verwendeten Rasterzeugnisformulare (insgesamt 54 Formulare, 23 Schulen). Die pädagogisch-diagnostische Bedeutung dieser Zeugnisform wird im Vergleich zu Ziffernzeugnissen und Textzeugnissen erörtert. In der Analyse der gestalterischen und der inhaltlichen Strukturmerkmale der Rasterzeugnisformulare zeigen sich sowohl ein hohes Maß an differenzierter, sachnormorientierter Beurteilung als auch methodische und pädagogische Probleme dieser Zeugnisvariante, für deren Behebung – soweit dies konzeptionell möglich erscheint – Vorschläge gemacht werden.

Inhaltsverzeichnis

1 Pädagogische Funktionen von Grundschulzeugnissen.....	2
1.1 Die informierende Funktion von Grundschulzeugnissen.....	2
1.2 Motivationale Aspekte der Beurteilungsalternativen.....	2
1.3 Diagnostische Probleme der Beurteilungsalternativen.....	3
1.4 Sachnormorientierte Leistungsbewertung, Selbstbewertung und unterrichtliche Differenzierung.....	3
1.5 Sachnormorientierung in Rasterzeugnisformulierungen.....	4
1.6 Lernwegeabbildung und Markierung von Lernniveauunterschieden als besondere Leistung von Lernentwicklungsberichten.....	4
2 Strukturmerkmale von Rasterzeugnissen.....	5
2.1 Merkmale der äußeren Gestaltung.....	6
2.2 Darstellung der Beurteilungsinformationen.....	7
2.2.1 Gliederungsformen und Lesbarkeit.....	7
2.2.2 Anordnung der Beurteilungsbereiche.....	9
2.2.3 Differenzierung in den Beurteilungsbereichen.....	9
2.3 Formen der Itemformulierung.....	12
2.3.1 Iteminhalt.....	12
2.3.2 Skalierung.....	12
2.4 Freie Beschreibungen.....	15
2.5 Abbildung von Lernentwicklung.....	16
2.6 Bezugsnormorientierung.....	17
3 Zeugnisbeispiele.....	17
4 Entwicklungsperspektiven für Zeugnisformen im Grundschulbereich.....	20
5 Literaturverzeichnis.....	21



1 Pädagogische Funktionen von Grundschulzeugnissen

1.1 Die informierende Funktion von Grundschulzeugnissen

Zeugnisse sollen eine *informative Funktion*¹ haben, dies ist eine in der pädagogischen Diskussion gewiß unstrittige Bestimmung. Schon die Folgefrage, *für wen* diese Informationsfunktion bestehe, führt zu divergenten Anforderungen:

- Sollen Zeugnisse primär *Schüler* über deren Lernentwicklung informieren und deshalb für Schüler verständlich formuliert sein,
- oder sollen primär *Eltern* der Schüler informiert werden, die weitaus weniger Gelegenheit zur Kommunikation mit den Lehrern haben als die Schüler und die ein gewisses Interesse an der lehrplanbezogenen Leistungsbeurteilung haben oder an dem Leistungsvergleich innerhalb der Schulklasse?
- Sollen Zeugnisse andere *Lehrer* informieren – mit entsprechend fachlicher Beschreibung und lehrplanbezogener oder fähigkeitsbeschreibender Darstellung – über die vorgängige Lernentwicklung der Schüler, die aus einer anderen Klasse, Schule oder Schulstufe, Schulart oder Schulform kommen?
- Sollen Zeugnisse anonyme »*außerschulische Abnehmer*« informieren wie z.B. Arbeitgeber oder Einrichtungen der tertiären Bildung, wobei schulfachbezogene und fachübergreifende Qualifikationen von Interesse sein können und die Bezugsnorm auf Bestenauslese beruhen kann?

Zugleich wird die Zeugniserteilung auch als ein *genuiner Teil der pädagogischen Arbeit der Schule* aufgefaßt, so daß Wirkungen – und Nebenwirkungen – dieser Berichterstattung untersucht werden können. Da der Grundschulbereich in einem Bildungswesen, das eine schulformunabhängige »Orientierungsstufe« an die Grundschule anschließt, keine Berechtigungsfunktionen vorhält – dies ist in z.B. in Bremen² der Fall –, fehlt der Zeugnisgebung in der Grundschule weitgehend diese entscheidungsorientierte Komponente. Somit geraten *pädagogische Funktionen* stärker in Betracht, die in positiver Formulierung als »lernförderliche Leistungsbeurteilung«³ bezeichnet werden können.

1.2 Motivationale Aspekte der Beurteilungsalternativen

In der grundschulpädagogischen Diskussion wird das im deutschen Schulsystem dominanten Beurteilungsinstrument der »Ziffernzensierung« aufgrund ungünstiger Wirkungen und Rückwirkungen auf die Lernmotivation kritisiert. Zwar ist bislang keine »experimentelle« Untersuchung zur motivationalen Wirkungsqualität unterschiedlicher Zeugnisformen vorgenommen worden, die verfügbaren Berichte und die Befragungsergebnisse von Schülern und Eltern lassen die Variante »Textzeugnis« jedoch als durchaus günstig erscheinen⁴. Zugleich muß allerdings berücksichtigt werden, daß auch die herkömmliche Ziffernzensierung mit beachtlicher Motivationswirkung selbst für eher schwache Grundschüler ausgestattet zu sein scheint.⁵ Die Praktizierung einer Sozialnormorientierung, d.h. eines klassenbezogenen Maßstabes, induziert grundsätzlich eine »Konkurrenz« um »gute Zensuren«, was in pädagogischer Hinsicht als weniger günstige, gleichwohl sehr wirksame Motivationsquelle bewertet wird.

¹ Eine Zusammenfassung der in der Literatur genannten »Zeugnisfunktionen« gibt Jürgens (1992, S. 50ff). Während Dohse (1960) die Mitteilungsfunktion bereits pädagogisch wertet, gliedert Zielinski (1974) »Rückmeldelfunktionen« für Schüler, Eltern und Lehrer als eigenständige Funktionen aus.

² Andere Bundesländer wie z.B. Hamburg haben zwar schulformabhängige Orientierungsstufen, räumen allerdings weder der abgehenden Grundschule noch der Sekundarstufenschule ein Bestimmungsrecht für die Schulformentscheidung ein. Insofern sind auch hier keine »Berechtigungseigenschaften« mit dem Grundschulzeugnis verbunden.

³ s. Olechowski, R. & Persy, E. (Hrsg.). (1987).

⁴ s. Ulbricht (1993); Lübke (1996); Valtin & Rosenfeld (1997).

⁵ Helmke (1992, S. 277) nennt die generelle Tendenz von Schülern zur Überschätzung ihrer schulischen Leistungen einen »Immunitätsfaktor«, der offensichtlich eine gewisse Resistenz gegenüber demotivierenden Wirkungen von »schlechten Zensuren« ermöglicht.

1.3 Diagnostische Probleme der Beurteilungsalternativen

Von seiten der pädagogisch-psychologischen Diagnostik wird die »informative Qualität« der herkömmlichen Ziffernzensierung unter drei entscheidenden Aspekten als »fragwürdig« dargestellt:

- (1) Ziffernzensuren besitzen vergleichbare Güteinformation »nur« innerhalb der Schulklasse und geben insofern nur vage Hinweise auf die lehrplanbezogene Kompetenz der Schüler.
- (2) Ziffernzensierung ist weniger objektiv und zuverlässig als Schulleistungstestverfahren, so daß »eigentlich« erhebliche »meßfehlerbedingte« Ungenauigkeiten in der Interpretation von Zensuren zu berücksichtigen wären.
- (3) Empirisch betrachtet erscheint der schulverlaufsbezogene Prognosewert der Zensuren zwar im Vergleich zu Schulleistungstestverfahren als nicht wesentlich ungünstiger, für Einzelfallentscheidungen allerdings ergeben sich erhebliche Entscheidungsgewißheiten. Zudem erweisen sich zensurenbasierte Schullaufbahnpfehlungen als sozial selektiv.⁶

Für die alternative Form der Leistungsbeurteilung, d.h. für lernprozeßorientierte Beschreibungen und für deren Integration in Lernentwicklungsberichten bzw. Textzeugnissen gelten die Einwände (2) und (3) in zumindest gleicher Weise, da es sich auch hier um »Beurteilungen« und insofern um »Schätzurteile« handelt, die grundsätzlich fehlerbehaftet sind. Der unter (1) genannte Einwand trifft allerdings zumindest konzeptionell nicht auf Lernentwicklungsberichte zu, wenn diese tatsächlich den erreichten Lernstand beschreiben. Damit stehen diese Lernbeschreibungen der »sachnormorientierten« Beurteilung nahe, die grundsätzlich die Beurteilung von »Lehrplanadäquanz« ermöglicht.

Der motivationale Rahmen für sachbezogene Beurteilungsstrukturen kann kaum ein anderer als der »sachliche«, d.h. der auf den Lerngegenstand bezogene sein, was als wünschenswerte, schulische Lernhaltung angesehen wird. Gleichwohl ist aus Schülerbefragungen und Schulerinnerungen von ehemaligen Schülern bekannt, daß »sachliches Interesse« bei Schülern nicht durchweg und vielleicht noch nicht einmal überwiegend besteht.⁷ In der Grundschule dominiert eher eine kindlich-personbezogene Lernmotivation, die in Wertschätzung der Lehrperson und in Respekt ihr gegenüber gründet. Insofern werden auch die Werturteile der Lehrperson über Lernwege und Lernergebnisse gänzlich oder zumindest tendenziell übernommen. Anhand dieser Wertungen entwickeln die Schüler eigene Gütestandards, sofern ihnen die Beurteilungsstrukturen dazu erschlossen werden.

1.4 Sachnormorientierte Leistungsbewertung, Selbstbewertung und unterrichtliche Differenzierung

Das System der herkömmlichen Ziffernzensierung kann zwar von Schülern mit bemerkenswerter Treffersicherheit nachgeahmt werden⁸, dennoch bleibt den Schülern die Art dieser (Selbst-)Beurteilungsprozesse weitgehend verschlossen, da hier eine diffuse und wenig konstante Mischung aus sachlicher Gütebeurteilung und Verteilung der Leistungsqualität in der Klasse praktiziert wird. Die sachnormorientierte Grundstruktur der alternativen Leistungsbeschreibung bietet hingegen eine grundsätzlich klare Struktur der Bewertungsprozesse, da »sachbezogene Güte« intersubjektiv nachvollziehbar ist – bei entsprechender Sachkompetenz. Insofern bieten Textzeugnisse eine in hohem Maße objektivierbare Information über die erreichten und die nicht erreichten Lernfortschritte an. Probleme treten auf, wenn individuelle Standards für die Zielerreichung festgelegt oder verändert werden sollen. Der Informationsbedarf von Eltern und auch von Schülern könnte folgende Fragen umfassen:

⁶ s. Lehmann & Peek (1997), S. 93.

⁷ s. Krapp, A. & Prenzel, M. (Hrsg.). (1992). Interesse, Lernen, Leistung. Neuere Ansätze einer pädagogisch-psychologischen Interessenforschung. Münster: Aschendorff.

⁸ s. Krampen, G. & Lehmann, P. (1981). Zensuren-Erwartungen von Schülern. Antezedensbedingungen und Realitätsgemessenheit. Zeitschrift für empirische Pädagogik, 5, 27-36.

- Was sind individuell angemessene Lernfortschritte?
- Welche lehrplangemäßen Ziele stellen Überforderungen für einen Schüler dar?
- Welche Lerndefizite sind mit vertretbarem Aufwand kurzfristig zu beheben, welche sind nur langfristig korrigierbar?

Die Beantwortung dieser Fragen basiert auf einer Einschätzung der grundlegenden, d.h. der als mittelfristig stabil gedachten Kompetenz in den Lernbereichen des schulischen Curriculum. Diese Kompetenz wird »irgendwie« von Lehrern und auch von Eltern geschätzt und als Maßstab verwendet, um »angemessene Lernanforderungen« im differenzierenden Unterricht zu bestimmen. Im nicht-differenzierenden Unterricht entfällt diese diagnostische Aufgabe; hier sind die Lernanforderungen für Klassenmitglieder gleich. Ziffernzensurierung ist insofern grundsätzlich ungeeignet, lernniveaudifferente Unterrichtsformen und damit niveaudifferente Prüfungsgelegenheiten (»Klassenarbeiten«) zu begleiten, es sei denn, die »Vorabzensuren« werden für bestimmte Aufgabenbereiche bereits mitgeteilt. Kindern fällt es allerdings schwer, sich für »leichte, aber lernstands- bzw. fähigkeitsangemessene Aufgabenbereiche« in einer Klassenarbeit zu engagieren, wenn bereits klar ist, daß deren mühevollen Bewältigung maximal eine »schlechte Zensur« erbringen wird. Binnendifferenzierendes Unterrichten und das damit verbundene leistungsdifferenzierende Prüfen erzwingen unter pädagogischem Gesichtspunkt ein alternatives Bewertungssystem, das die faktisch erreichten Lernfortschritte würdigt und möglicherweise weitere Informationen gibt, z.B. über das Ausmaß der aufgewandten Anstrengung oder über die – ungefähre – Position im Leistungsspektrum der Klasse.

Ein solches Zeugnis integriert multiple Beurteilungen, die aus unterschiedlichen Bezugsnormvergleichen resultieren.⁹ Teil eines solchen integrierten Zeugnisses kann ein »Unterrichtsbericht«¹⁰ sein, der die faktisch unterrichteten Lernbereiche und Lerngegenstände aufzeigt und – sofern Zweifel bestehen könnten – das Ausmaß der Lernplanadäquanz benennt.

1.5 Sachnormorientierung in Rasterzeugnisformulierungen

»Rasterzeugnisse« enthalten zwar keine Unterrichtsberichte, die aufgeführten Beschreibungen von Lernzielen spiegeln hingegen die zentralen Lernanforderungen des Lehrplans wider. Wäre dem nicht so, könnten Rasterzeugnisse behördlich nicht genehmigt werden. Unklar bleibt in Rasterzeugnissen, in welchem Ausmaß der Unterricht »Lerngelegenheiten« für den Erwerb der aufgeführten Lernziele gegeben hat.

Die Beurteilungen in Rasterzeugnissen stellen insofern »lehrplanorientierte Bewertungen« dar, d.h. sie repräsentieren »sachnormorientierte« Leistungsbeurteilungen. Die Genauigkeit der Feststellungen zum Grad der Lernzielerreichung variiert prinzipiell, möglicherweise aber nicht empirisch mit der Art der vorgegebenen Antwortmöglichkeiten, wobei allerdings unklar bleibt – wie auch bei der herkömmlichen Ziffernzensurierung –, ob skalenmäßig angrenzende oder entferntere Beurteilungen geringe oder große Kompetenzunterschiede abbilden, z.B. leichthin erreichbare oder kaum erreichbare Kompetenzstufen. Kurzum: die »Metrik« der rangunterscheidenden Beurteilungsmöglichkeiten pro Lernzielnennung ist unklar.

1.6 Lernwegeabbildung und Markierung von Lernniveaunterschieden als besondere Leistung von Lernentwicklungsberichten

Lernentwicklungsberichte sind grundsätzlich geeignet, diese »inhaltlichen« Unterschiede deutlich zu machen, indem der Lernaufwand beschrieben wird, der mit dem Erreichen höherer Kompetenzstufen verbunden ist. Zwar wird damit keine »formale Metrik« angeboten, jedoch

⁹ s. Schulbegleitforschungsprojekt »Integrierte Leistungsbeurteilung in der Sekundarstufe I« (1997).

¹⁰ s. von Hentig, 1982, S. 250.

können noch nicht erreichte Lernstände bzw. überschrittene Lernniveaus mit dem erreichten Lernniveau so in Beziehung gesetzt werden, daß die zwischengeschalteten »Lernprozesse« sichtbar werden. Damit erreichen Lernentwicklungsberichte ein Informationsniveau, das grundsätzlich dem Niveau der Rasterzeugnisse überlegen ist, wobei Rasterzeugnisse bereits das Informationsniveau von Ziffernzeugnissen in bedeutsamem Ausmaß übertreffen.

Ob Zeugnisinformationen über »Lernwege«¹¹, die zu einem bestimmten Lernstand geführt haben, für Adressaten wie Eltern oder Schüler von größerem Interesse sind, ist nicht leicht zu entscheiden. Wenn mit der Beschreibung von »Prozeßeigenschaften des Lernens« allerdings Qualifikationen benannt werden, die mit kognitiven Strategien und mit Lernstilen verbunden sind, so geben diese Informationen Hinweise zu den transferstärksten Elementen schulischen Lernens, die zumindest im Kindesalter nicht »Gegenstand« des Lernens sein können, sondern die in sachbezogenen Lernverläufen mehr oder minder entwickelt werden. Grundsätzlich können Rasterzeugnisse auch Angaben zu diesen Qualifikationen enthalten. Als Aussagen wirken diese oftmals wenig »überzeugend«, da diese »gegenstandsfern« berichtet werden müssen. Lernentwicklungsberichte können diesen erhellenden Gegenstandsbezug herstellen. Rasterzeugnisse erscheinen somit eher einem eher »produktorientierten« Leistungsbegriff verpflichtet als einem »erweiterten Verständnis« von Lernleistung¹².

2 Strukturmerkmale von Rasterzeugnissen

Definition von Rasterzeugnissen

»Rasterzeugnisse« werden gelegentlich auch als »Lernentwicklungsberichte in gebundener Form« bezeichnet. Rasterzeugnisse bestehen aus Beurteilungen und z.T. Beschreibungen über *Lernergebnisse*; sie können auch Angaben zur *Lernentwicklung* enthalten. Während in Ziffernzeugnissen Beurteilungen gegeben werden, die aus einer einzigen (gelegentlich auch aus mehreren), dimensionslosen Beurteilungsvariablen bestehen, werden in Rasterzeugnissen *mehrfache und inhaltlich dimensionierte Beurteilungen zu den Bereichen schulischer Unterrichtung und schulischen Verhaltens* mitgeteilt.

Curriculare Normorientierung von Rasterzeugnissen

Herkömmliche Ziffernzensurierung besteht aus einer eher diffusen Mischung von sach- und sozialnormorientierter Beurteilung (vgl. Kap. 1.2, S. 3); die bislang verbindliche KMK-Vereinbarung von 1968 zur Zensurengebung in Zeugnissen steht dieser Praxis jedoch entgegen und fordert eine Sachnormorientierung¹³. Die Aussageinhalte, d.h. die Inhalte der einzelnen Formulierungen von Rasterzeugnissen spiegeln insgesamt – jedenfalls konzeptionell betrachtet – den curricularen Anforderungsrahmen und damit den kulturministeriell geforderten, allerdings spezifisch skalierten Sachnormbezug wider.

Rasterzeugnisse und Zeugnisordnung

Die »Zeugnisordnung« des Bundeslandes Bremen trägt die Bezeichnung »Verordnung für Zeugnisse und Lernentwicklungsberichte« und benennt damit die beiden Varianten, denen allerdings kein Oberbegriff zugeordnet wird. »Rasterzeugnisse« werden in der Verordnung nicht erwähnt. Da Rasterzeugnisse keine »Noten« enthalten, ist diese Zeugnisform der Variante »Lernentwicklungsberichte« zuzuordnen. Lernentwicklungsberichte und Zeugnisse haben den

¹¹ s. Senator für Bildung und Wissenschaft, 1993, S. 79.

¹² s. Kommission »Zukunft der Bildung – Schule der Zukunft« beim Ministerpräsidenten des Landes Nordrhein-Westfalen/ Bildungskommission NRW, 1995, S. 87f. – s.a. Klafki, 1994.

¹³ s. Sacher, 1994, S. 56.

übergreifenden Zweck, »eine zusammenfassende Beurteilung der Lernentwicklung ...in einem bestimmten Zeitraum«¹⁴ zu geben; sie enthalten »die Ausweisung der Lernentwicklung und der Leistung«¹⁵ und beziehen »sowohl die fachlichen Fähigkeiten und Kenntnisse als auch die Arbeitsweisen, die Bestandteil der fachlichen Leistungen sind«¹⁶, ein. Ob die herkömmliche Ziffernzensur diesen Anforderungen genügen kann und zugleich ein hinreichend überschaubares Urteilsmodell bleibt, kann aus diagnostischer Sicht durchaus bezweifelt werden. Ob Rasterzeugnisse diese Leistung erbringen können, soll die vorliegende Untersuchung klären. Daß Textzeugnisse diese Funktion erfüllen können, erscheint evident; ob die vorfindliche Textzeugnispraxis allerdings bereits so beschaffen ist, stellt eine kritische Frage dar.

Untersuchungsstichprobe

In die Untersuchung sind 54 Rasterzeugnisformulare, die aus insgesamt 23 (18 stadtbremischen und 5 bremerhavener) Schulen stammen, einbezogen worden. Die Zeugnisformulare hatten Geltung im ersten Schulhalbjahr 1996/97; von einzelnen Schulen ist seinerzeit angekündigt worden, daß Überarbeitungen anstünden, so daß die berichteten Ergebnisse *im Detail* nicht mehr aktuell sein könnten. Im Erfassungszeitraum wurde mir bekannt, daß weitere Schulen an der Entwicklung von Zeugnisformularen arbeiten, so daß die aktuelle Anzahl von Schulen, die mit Rasterzeugnissen arbeiten, etwas höher liegen wird als 23. Diese Angabe stellt bereits für den Untersuchungszeitraum eine Unterschätzung dar, da mehrere Schulen aufgrund von seinerzeit laufenden Umarbeitungen des Rasterformulars noch nicht an der Untersuchung teilnehmen mochten.

2.1 Merkmale der äußeren Gestaltung

Die untersuchten Rasterzeugnisse zeigen ein breites Spektrum an Text- und Layoutgestaltung. Alle Zeugnisformulare sind mit den Möglichkeiten elektronischen Satzsetzes erstellt und nutzen die üblichen Mittel der Textanordnung in tabellierter Form mit unterschiedlichen Hervorhebungsarten des Schriftschnittes, der Schriftgröße und -art. Da die so erstellten Druckvorlagen gespeichert werden können, sind Änderungen der Zeugnisform und der Zeugnisinhalte leichthin möglich. Die Zulassung von Zeugnisformularen und deren inhaltliche Veränderungen regeln »Zeugniserlasse«¹⁷.

Format, graphisches Deckblatt, Seitenzahl, Anzahl der Formulare pro Schule, Kosten

Bei der überwiegenden Mehrzahl der Rasterzeugnisformulare wird das DINA4-Format (20 Schulen) verwendet; 3 Schulen verwenden Heftformen im A5-Format. Teilweise werden die A4-Bögen als Zeugnishefte hergestellt, teilweise als Doppelbögen (A3) gedruckt und gefaltet oder als Einzelblätter mit Klammerung verwendet.

Graphische Deckblätter zeigen die Rasterzeugnisse von 21 Schulen.

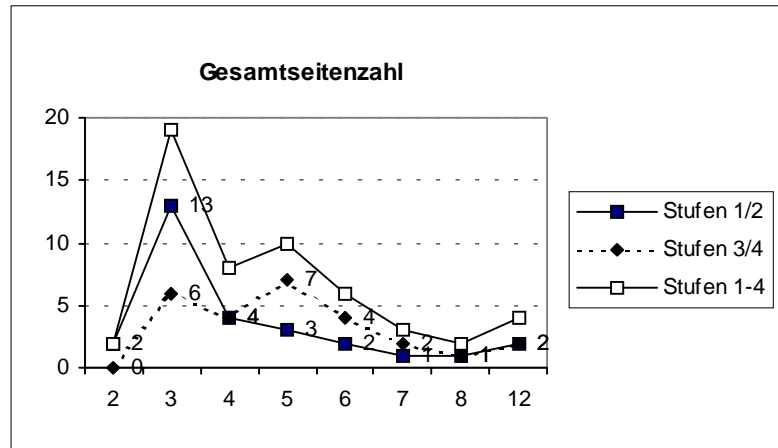
¹⁴ s. Senator für Bildung und Wissenschaft, 1997, § 2.

¹⁵ s. op. cit., § 3, Abs. 1.

¹⁶ s. op. cit., § 3, Abs. 2.

¹⁷ s. § 13, Abs. 3 der Bremischen Verordnung für Zeugnisse und Lernentwicklungsberichte.

Der *Seitenumfang* beträgt durchschnittlich ca. 4 und variiert zwischen 2 und 12 Textseiten. 3 Textseiten (plus 1 Deckblatt) bilden das mit Abstand häufigste Layout.



Für die oberen Klassenstufen werden Rasterzeugnisformulare mit etwas größeren Seitenumfängen verwendet als für die unteren Klassenstufen, was weniger auf die größere Fächerzahl als auf die zunehmende Differenziertheit der Beschreibungen, d.h. auf die pro Bereich verwendete Itemzahl zurückzuführen ist (s. Kap. 2.2.3, S. 9).

Die meisten Schulen (n=12) verwenden 2 *Rasterzeugnisformulare*, die zumeist für die Klassenstufen 1/2 und 3/4 ausgearbeitet sind. Eine beträchtliche Anzahl von Schulen hat für jede Klassenstufe ein eigenständiges Rasterzeugnisformular erstellt (n=6). In 5 Schulen wird nur ein Zeugnisformular verwendet, tw. nur für bestimmte Klassenstufen (z.B. Stufen 3 und 4) oder aber als klassenstufenübergreifendes Formular.

Die *Kosten* für die Erstellung der Zeugnisdrucke können durchaus beträchtlich sein. Da zumindest 2 Exemplare erstellt werden müssen (ein Exemplar für die Schülerakte), kann folgende Modellrechnung gelten: Bei durchschnittlich 6 Druckseiten pro Zeugnis ergeben sich bei beidseitigem Druck in einer dreizügigen Grundschule mit einer Klassenfrequenz von 25 Schülern insgesamt 3600 Druckseiten und ein Papierbedarf von 1800 Seiten. Die Verwendung eines schuleigenen Kopierers hat einen Kostenfaktor von ca. 0.08 DM pro Seite, so daß Druckkosten von 288 DM anfallen und Papierkosten von ca. 20 DM. So ergeben sich ca. 300 DM an Minimalkosten bei einem einfach gehefteten Zeugnisformular. Etliche Schulen verwenden stärkeres Papier oder gar maschinell gedruckte und geheftete Zeugnisse, die eine beträchtliche Vorausinvestition erfordern. Wenn für Zeugnis Konferenzen gut lesbare Vorlagen der Rasterzeugnisbeurteilungen erstellt werden sollen und diese Vorlagen z.B. aufgrund von beschlossenen Änderungen nicht als Exemplar weiterverwendet werden können, entsteht weiterer Kopierbedarf von mindestens der og. Größenordnung. Somit beläuft sich eine mittlere Materialkostenschätzung auf ungefähr 600 DM pro Grundschule, d.h. auf 2 DM pro Schüler und Schuljahr. Daß »ziffernzensierende« Schulen auch in diesem Bereich der Beurteilungspraxis erheblich weniger Aufwand betreiben, könnte ein Thema schulvergleichender Diskussion sein.

2.2 Darstellung der Beurteilungsinformationen

2.2.1 Gliederungsformen und Lesbarkeit

Gliederungsstruktur von Rasterzeugnisformularen

Die meisten Rasterzeugnisse sind in die folgenden *Bereiche* gliederbar:

1. Deckblatt
 - a) Angaben zur Schule, tw. mit Zeichnung des Gebäudes
 - b) Abbildung mit kindgemäßem Inhalt
 - c) Name und Geburtsdatum des Schülers
 - d) Schuljahr(e), auf das (die) sich die Zeugnisinformation bezieht

- e) Klasse
- e) Ausstellungsdatum des Zeugnisses
- f) Versetzungsvermerk
- 2. Beurteilungen zum Arbeits- und Sozialverhalten
- 3. Beurteilungen zu den Leistungen in den Unterrichtsfächern
- 4. freie Bemerkungen
- 5. Bemerkungen zur Teilnahme an Arbeitsgemeinschaften und Projekten
- 6. Bemerkungen zur Teilnahme an Förderunterricht
- 7. Bemerkungen zur Teilnahme an muttersprachlichem Unterricht in einer Fremdsprache

Lesbarkeit von Rasterzeugnissen

Die fachbezogenen Beurteilungen bestehen aus einzelnen vorgegebenen Feststellungen (»Items«), denen eine Beurteilung per Markierung auf vorgegebener Stufung (»Skala, Skalierung«) zugeordnet wird. Die Items werden in listenartiger Form optisch zusammengefaßt und mit *Bereichsbezeichnungen* (Fächernamen, tw. fächerspezifischen Lernbereichen, Verhaltensbereichen) versehen.

Diese Gruppierungen von Items erscheinen zumeist als Tabellen, die unterschiedliche Zeilen- und Spaltenzahlen aufweisen. Einzeltabellen können durchaus mehr als 10 Zeilen und fünf Spalten umfassen, so daß Felder von mehr als 50 Zellen entstehen. Tabellen einer solchen Größenordnung stellen erhebliche Anforderungen an den Leser; für Kinder im Grundschulalter werden Informationsverdichtungen dieser Art kaum als didaktisch angemessen eingestuft werden können. Die geringe Variabilität der Skalierungsinformation wirkt sich bezogen auf eine dauerhafte Informationsübernahme aus wie der aus der Gedächtnispsychologie bekannte »Ranschburg-Effekt«, d.h. differentielle Bewertungsspeicherungen sind bei Kindern – und vermutlich auch bei vielen Erwachsenen – nach Lektüre eines Rasterzeugnisses bzw. einer größeren tabellarischen Darstellung innerhalb eines Rasterzeugnisses (z.B. die Angaben zum Mathematikunterricht) nicht zu erwarten. Kognitionspsychologische Untersuchungen fänden hier eine interessante und anwendungsbezogene Fragestellung, deren hypothetische Beantwortung mit den vorgängigen Bemerkungen versucht worden ist.

Das Verständnis der Rasterzeugnisinformation wird Kindern auch dadurch erschwert, daß ihnen die didaktische Struktur eines Unterrichtsfaches noch wenig erschlossen ist. Insofern bietet ihnen die Auflistung von Items eine zum Teil neue und abstrahierende Sichtweise auf die eigenen Lernvollzüge, die darin mehr oder minder wiedererkannt werden können.

Grundschulern einen »Leseanlaß« zu bieten, können Rasterzeugnisse jedenfalls kaum beanspruchen, da diese »Textgattung« von strenger Informationsorientierung geprägt wird, die auf alle Mittel »prosaischer Satzverknüpfung« verzichtet. Insofern begegnen Kinder in den Rasterzeugnissen zwar einer auf sie und ihre schulischen Verhaltensweisen, Lerntätigkeiten und Lernergebnisse bezogenen Mitteilung, d.h. einer höchst individuellen »Lernbiographie«, deren Anmutungsqualität entbehrt allerdings jedweder persönlichen Verbindlichkeit. Eine interessante Frage wäre, welche Gefühle und Assoziationen sich bei Kindern, die mit ihren Grundschullehrern nur sehr selten oder vielleicht gar nicht diese Form der unpersönlichen Rückmeldung persönlicher Information erleben, einstellen, wenn sie die vom Klassenlehrer ausgehändigten und unterzeichneten Rasterzeugnistexte »lesen«. Jedenfalls werden jüngere Grundschulkinder annehmen müssen, daß ihr Lehrer sie mit diesem ihnen zugeordneten Text bei weitem überfordert, was dem vom ihm gewohnten didaktischen Verhalten auffallend wenig entspricht. Als Schlußfolgerung wird sich so vermutlich einstellen, daß der Lehrer diese Art der »Undidaktik« aus anderen als didaktischen oder pädagogischen Gründen praktiziert, was ihm hoffentlich nachgesehen wird und nachgesehen werden kann, wenn folgerichtig unterstellt wird, das nicht die Schüler, sondern die Eltern Adressaten der Rasterzeugnisse seien.

Manche Grundschulen, die Rasterzeugnisse verwenden, händigen den Kindern lesbare Texte zusätzlich aus, die »Lernbriefe« oder »Lernentwicklungsberichte« oder »Kinderzeugnisse«

genannt werden können. Das Vorhandensein dieser Doppelpraxis der Berichterstattung könnte als Beleg für die kindbezogenen Schwächen der Rasterzeugnispraxis angesehen werden.

2.2.2 Anordnung der Beurteilungsbereiche

Die og. Gliederungsstruktur wird in fast allen Rasterzeugnisvarianten so eingehalten. Die beiden sog. Hauptfächer treten in etwa gleich häufig in der *Reihung* »Deutsch–Mathematik« wie in deren Umkehrung »Mathematik–Deutsch« auf.

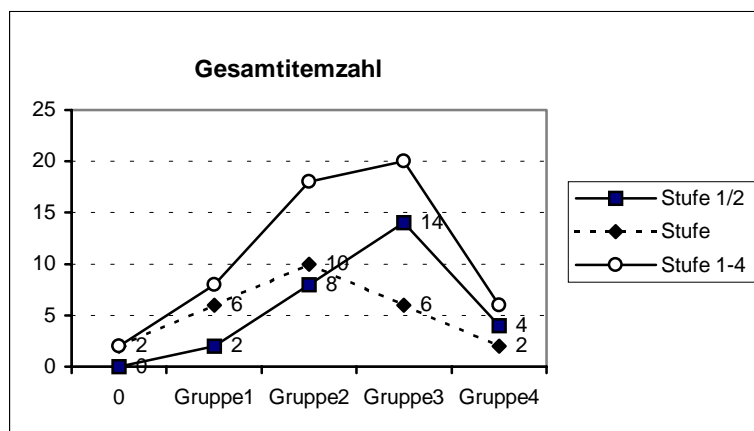
Die Berichtsreihenfolge für die fachübergreifenden Bereiche hingegen ist deutlicher akzentuiert: ungefähr 2/3 der Rasterzeugnisvarianten stellen die *Beschreibung des Sozialverhaltens voran* und ordnen dieser die Beschreibung des Arbeitsverhaltens nach.

Innerhalb der Beurteilungsbereiche wird keine *Gewichtung* der berichteten Informationen vorgenommen: so stehen Items von großem Aussagebereich und hoher kognitionspsychologischer Bedeutung (Beispiel: »kann Zusammenhänge erkennen«) an weder gestalterisch noch inhaltlich herausgehobener Position im Ensemble der übrigen Items. Aus diesem Umstand könnte fälschlicherweise geschlossen werden, daß alle Items gleichrangige Lernziele benennen, was gewiß nicht der Fall ist. Welche pädagogischen Wertigkeitsunterschiede allerdings zwischen den aufgeführten Items bestehen, wird der Beurteilung des Lesers überlassen, die nicht unbedingt identisch sein muß mit derjenigen der Lehrer.

2.2.3 Differenzierung in den Beurteilungsbereichen¹⁸

Gesamtitemzahl

Durchschnittlich werden ca. 64 Items in den Rasterzeugnisformularen aufgeführt; die Streuung¹⁹ beträgt ca. 21. Drei Rasterzeugnisformulare verwenden mehr als 100 Rasteritems. Auf der Basis der durchschnittlich verwendeten Itemmenge kann errechnet werden, daß ein Rasterzeugnis in etwa 3^{64} Kombinationsmöglichkeiten mit Einzelangaben über das Lernverhalten aufweisen.



Für die beiden Stufenbereiche der Grundschule ergeben sich die erwartbaren Unterschiede: die Itemzahl steigt mit der Schulstufe an.

¹⁸ Itemzahlen werden – falls hinreichend große Streuung vorliegt – so gruppiert, daß beidseits des Mittelwertes der Verteilungen zwei Häufigkeitskategorien erscheinen, deren Breite jeweils eine Standardabweichung beträgt, so daß sich in diesen beiden Kategorien der »Durchschnittsbereich« zeigt, d.h. ca. 68% der Werte liegen darin – unter der allerdings kritischen Annahme von Normalverteilung. Andernfalls bildet die mittlere Kategorie diesen »Durchschnittsbereich«. Die Gruppierungskennzeichnung in den Tabellen verwendet »größer« im Sinne von »größer oder gleich« dem angegebenen Punktwert.

¹⁹ Die Standardabweichung der Punktwerte wird als »Streuung« bezeichnet.

Obige Beispielrechnung macht deutlich, daß die Itemmenge der Rasterzeugnisse eine andere Funktion haben muß als die *direkte* Abbildung von individuellem Schülerverhalten. Als *mittelbare* Funktion kommt in Betracht, daß die mehr oder minder ähnliche Aspekte erfassenden Items gruppierbar sein können und daß innerhalb dieser Itemgruppen Schätzungen von »itemübergreifenden Merkmalen« vorgenommen werden, d.h. die Itemangaben könnten zu *Skalenwerten* zusammengefaßt werden. Items ähnlichen Meßanspruchs sind dann trotz der durch sie beigetragenen »Erhöhung von Redundanz« (s. Modellrechnung) dadurch zu rechtfertigen, daß sie die Meßgenauigkeit übergeordneter Skalen bedingen, die kaum auf isolierten Items basieren können.

Kein Rasterzeugnis aus der untersuchten Stichprobe unternimmt den hier skizzierten Versuch, Einzelitemwerte numerisch zusammenzufassen. In methodischer Hinsicht ist ein solches Vorhaben dadurch zu realisieren, daß Modelle der Testtheorie auf größere Stichproben von individuellen Rasterzeugnissen angewandt werden in der Hoffnung, entweder bereichsbezogene Skalen verifizieren oder empirisch vorfindliche Itemgruppen pädagogisch-didaktisch sinnvoll interpretieren zu können. Eine solche Vorgehensweise ließe auch Fragen der Effizienz beantworten: Möglicherweise könnte ohne gravierenden Informationsverlust auf eine gewisse Menge von Items verzichtet werden.

Itemzahl zum Arbeitsverhalten

Das Arbeitsverhalten wird mit durchschnittlich 10 Items beschrieben (Streuung: 4). Damit zeigt sich für diesen fächerübergreifenden Aspekt schulischen Lernens ein Maß an differenzierter Berichterstattung, das über demjenigen der sog. Nebenfächer liegt und gleichwohl deutlich unter demjenigen der beiden sog. Hauptfächer rangiert.

Itemzahl zum Sozialverhalten

Das Sozialverhalten wird mit durchschnittlich 5 Items beschrieben. Der Differenziertheitsgrad liegt somit unter demjenigen, der für das gleichfalls fachübergreifende Arbeitsverhalten anzutreffen ist.

Itemzahl zum Fach Deutsch

Das Fach Deutsch wird durchschnittlich mit 16 Items beschrieben (Streuung: 4). Die bereits beschriebene Jahrgangsstufenabhängigkeit zeigt sich auch hier. Das Fach Deutsch hat im Vergleich zu den anderen Bereichen die höchste Itemzahl.

Da die Fächer Deutsch (und Mathematik) mit relativ vielen Items beschrieben werden und zudem gestalterische Itemgliederungen in vielen Rasterformularen verwendet werden, können Unterbereiche definiert und in ihrer Häufigkeit untersucht werden. Die für das Fach Deutsch verwendeten Items werden 7 Kategorien zugeordnet, die mit den vorangestellten Kürzeln nicht gänzlich treffend bezeichnet werden. Die achte Kategorie dient als Restkategorie und wird hier nicht ausgewiesen.

- a) »Sprechen«: Fähigkeiten und Leistungen in der Hervorbringung und Darstellung von freien und vorgegebenen Unterrichtsbeiträgen.
- b) »Lesen«: Lesequalität (ohne sprachgestalterische Aspekte, s. dazu a).
- c) »Texterstellung«: Fähigkeit zur Erstellung unterschiedlicher Textarten und zum informatorischen Umgang mit Texten.
- d) »Grammatik«: Kenntnisse und Anwendung grammatischer Formen.
- e) »Orthographie«: Rechtschreibleistung und Regelwissen.
- f) »Schriftbild«: Lesbarkeit, Ästhetik und graphische Normentsprechung von Geschriebenem.

»Sprechen«

Mit durchschnittlich ca. drei Items wird dieser Bereich beschrieben, der für die untere und obere Stufe der Grundschule in etwa ähnlich differenziert wird.

»Lesen«

Die Lesefähigkeit wird durchschnittlich mit ca. drei Items beschrieben. Für die Unterstufe der Grundschule wird dieser Bereich deutlich differenzierter berichtet, was die curriculare Platzierung des Leselehrgangs widerspiegelt.

»**Texterstellung**«

Die Angabe eines stufenübergreifenden Durchschnittswertes erscheint angesichts der schwerpunkthaften Differenzierung für die obere Stufe der Grundschule wenig sinnvoll. Für diese Teilgruppe werden durchschnittlich drei Items zur Beschreibung verwendet.

»**Grammatik**«

Für diesen Leistungsbereich gelten die gleichen Feststellungen wie für den Bereich »Texterstellung.

»**Orthographie**«

Die Rechtschreibfähigkeit wird mit durchschnittlich vier Items beschrieben. Für die Oberstufe der Grundschule wird nicht unter einer Minstdifferenzierung nach 3 Merkmalen der Leistungsstand berichtet.

»**Schriftbild**«

Das Schriftbild wird mit wenigen Merkmalen beschrieben (häufig mit einer freien Antwortkategorie). Der Stufenunterschied in der Differenzierung ist unbedeutend.

Insgesamt kann festgestellt werden, daß die zentralen Leistungsbereiche bezogen auf die Anforderungen von unterer und oberer Stufe der Grundschule in etwa gleich differenziert dargestellt werden. Eher formale Qualifikationen werden mit geringerer Differenziertheit berichtet; die Rechtschreibleistung zählt zu den höher differenziert berichteten Leistungsbereichen.

Itemzahl zum Fach Mathematik

Das Fach Mathematik weist mit durchschnittlich 14 (Streuung: 5) eine etwas geringere Itemzahl auf als das Fach Deutsch.

Den für das Fach Mathematik verwendeten Items werden 4 Kategorien zugeordnet, die mit den vorangestellten Kürzeln nicht gänzlich treffend bezeichnet werden. Die fünfte Kategorie dient als Restkategorie und wird hier nicht ausgewiesen.

- a) »Zahlbegriff« : Orientierung in den unterschiedlichen Zahlräumen.
- b) »Rechenoperationen« : Kenntnis und Beherrschung der Rechenarten.
- c) »Sach-/Größenrechnen« : Anwendung der Rechenarten auf Sachaufgaben bzw. Größenberechnungen.
- d) »Geometrie« .

»**Zahlbegriff**«

Mit durchschnittlich ca. einem Item wird diese Qualifikation beschrieben, zumeist durch Benennung des entsprechenden Zahlbereichs, der unterrichtlich erschlossen worden ist und für den Orientierungsfähigkeit beurteilt wird.

»**Rechenoperationen**«

Mit ungefähr sechs Items wird der Bereich der Rechenoperationen beschrieben. Für die obere Stufe der Grundschule ist curricular der Erwerb von mehr Rechenoperationen vorgesehen, insofern zeigt sich hier eine größere Itemmenge.

»**Sach-/Größenrechnen**«

Dieser Bereich wird mit durchschnittlich ca. drei Items beschrieben. Zwischen unterer und oberer Stufe der Grundschule besteht kein nennenswerter Unterschied.

»**Geometrie**«

Durchschnittlich zwei Items werden zur Beschreibung der erworbenen Qualifikationen verwendet. Auffallend erscheint, daß in einer beträchtlichen Menge von Rasterzeugnissen der unteren wie der oberen Grundschulstufe dieser Bereich keine Erwähnung findet.

Insgesamt zeigt sich zur Beurteilungsdifferenzierung im Fach Mathematik ein ähnliches Bild wie für das Fach Deutsch. Die Anzahl der hier gewählten zusammenfassenden Itemkategorien erbringt diese Ähnlichkeit: offensichtlich führt diese fachlich-didaktische Strukturierung zu einer mittleren Differenzierungsqualität von ca. drei Items pro Beschreibungsbereich, die unabhängig von der fachlichen Bindung das Ausmaß der Bereichsdifferenzierung charakterisiert.

Itemzahl zum Fach Sachunterricht

Über das Fach Sachunterricht wird in den Rasterzeugnissen durchschnittlich mit ca. 5 Items berichtet. Damit steht dieses Fach deutlich in der Berichtsdimensionierung hinter den beiden sog. »Hauptfächern«.

Itemzahl zu den Fächern Kunst und Musik

Dem Fach *Kunst* werden durchschnittlich 3 Items zugeordnet; für das Fach *Musik* gilt ähnliches. Die Angaben zum Fach Kunst werden in einigen Rasterzeugnissen tabellarisch gemeinsam verknüpft berichtet mit den Fächern Werken und Textiles Gestalten.

Itemzahl zu den Fächern Werken und Textiles Gestalten

Die Fächer werden durchschnittlich mit 5 Items beschrieben. Häufig wird über beide Fächer, die zumeist epochal unterrichtet werden, in einem gemeinsamen tabellarischen Bereich berichtet.

Itemzahl zum Fach Sport

Dem Fach Sport werden durchschnittlich 4 Items zugeordnet.

Zusammenfassend kann festgestellt werden, daß Rasterzeugnisse eine stimmige Differenzierung aufweisen: Die zur Bereichsbeschreibung verwendeten Itemmengen spiegeln die Wertigkeit der Grundschulfächer überwiegend so wider, wie dies in der recht groben, gleichwohl aber unterrichtszeitbezogenen Stufung nach »Haupt- und Nebenfächern« der Fall ist. Wie allerdings *fachübergreifender Unterricht* und dessen Lernergebnisse in diesem eher »traditionellen« Schema abgebildet werden können, stellt eine kritische Frage dar, die kaum anders beantwortbar ist als durch den Hinweis, daß diese Unterrichtsgestaltung entweder »aus dem Hauptraster« fällt und im »Nebenraster« der Bemerkungen beschrieben wird oder daß z.B. Projektarbeit hinsichtlich deren Beurteilung in die fächerspezifische Betrachtung eingebunden wird. Daß die Grundschule das Lernverhalten selbst zu prägen und zu strukturieren beabsichtigt, wird über den relativ großen Itembereich des »Arbeitsverhaltens« verdeutlicht. Die sozialerzieherische Orientierung der Grundschule zeigt sich in einer itembezogenen Differenziertheit, die unter dem Kennwert für das Arbeitsverhalten liegt und in etwa auf dem Niveau der sog. Nebenfächer rangiert.

2.3 Formen der Itemformulierung

2.3.1 Iteminhalt

Die Grundstruktur der Iteminhalte entspricht der Operationalisierung von Lernzielen: bestimmte *Verhaltensweisen* bzw. *-merkmale* werden angeführt, deren Ausprägung durch die Art der Itemskalierung differenziert wird. Insofern enthalten die meisten Itemformulierungen sprachliche »Anzeiger« für die gemeinten Schülerverhaltensweisen. Selten werden Items verwendet, die »*Personeigenschaften*« benennen – eine Beurteilungspraxis, die als »Beeigenschaftung«²⁰ kritisierbar ist, z.B. auch für die Abfassung von Lernentwicklungsberichten.²¹

Zumeist ergibt sich trotz der grundsätzlich angelegten Verhaltensorientierung eine recht monotone Art der Itemformulierung, die durch die permanente Verwendung des Hilfsverbs »können« induziert wird. Teilweise werden in Rasterzeugnissen nebeneinander zwei Arten von Itemformulierungen verwendet: So wird über »*Können*« berichtet und über »*Tun*«. Diese

²⁰ s. Kleber, 1992, S. 140.

²¹ s. Arnold (1997).

offenbar gewünschte Unterscheidung ist jedoch nicht als systematische Strategie nachvollziehbar. Sie gründet vermutlich in der diagnostischen Annahme, daß der Unterrichtsablauf nur sehr beschränkt den Schülern die Möglichkeit gibt, ihr fachliches Können dem beurteilenden Lehrer faktisch zu zeigen. Hingegen sind z.B. unangepaßte unterrichtliche Verhaltensweisen leichthin in die »fast« vollständige Beobachtung durch den Lehrer gestellt.

Der zentrale Bedeutungsinhalt der meisten Rasterzeugnisitems wird durch die Benennung eines *thematischen Elementes* gestiftet (Beispiel: »mit Längenmaßen rechnen«; »Rechtschreibregeln anwenden«) oder durch die *qualifikationsbezogene Präzisierung allgemeiner Lernhandlungen* (»sinnerfassend lesen«).

Gelegentlich sind Itemformulierungen anzutreffen, die *Gütemerkmale* für Lernhandlungen bzw. Verhaltensweisen enthalten (Beispiel: »kann ›verständlich‹ erzählen«; »kann Texte ›selbstständig‹ erlesen«). Die Kombination der güteanzeigenden Itemaussage mit der Skalierung führt dann zu sprachlichen und logischen Problemen, wenn die Skalierung selbst güteorientiert (»sicher vs. unsicher«) erfolgt. Beispiele dazu werden in Kap. 3 angeführt.

Das *methodische Grundproblem der Rasterzeugnisse* wird an dieser Stelle deutlich: Die Qualität von Lernergebnissen soll ebenso abgebildet werden wie das Ausmaß des Zeigens von Kompetenzen. Beide Aspekte sind in jedweder Theorie schulischen Lernens enthalten; schwierig hingegen ist es, beide Aspekte in ein einheitliches Beurteilungsschema (»Itemformat«) zu integrieren. Insofern sind unterschiedliche Itemformate kaum zu vermeiden, was allerdings der gewünschten Einfachstruktur für Rasterzeugnisse zuwiderläuft. Als Grundregel sollte jedoch beachtet werden, daß Güteabstufungen möglichst nur auf Sachverhalte angewendet werden, die hinreichend klar formuliert werden können und die selbst nicht zu einer Verdoppelung der Gütespezifizierung auf gleicher Dimension beitragen (s. Kap. 2.3.2).

Die überwiegende Zahl der Rasterzeugnisse (4/5) verwendet Items, die in der *Zeitform* des Präsens berichten. Das Präsens signalisiert als sprachliche Eigenschaft, daß das Berichtete grundsätzlich als andauerndes Faktum gekennzeichnet wird. Schulische Unterrichtung hat gewiß den Anspruch, Lernergebnisse von einer gewissen Dauerhaftigkeit hervorzubringen; insofern besteht hier eine Korrespondenz. Kinder neigen vermutlich auch zu der Ansicht, daß schulisch erreichter Kompetenzzuwachs eine erhebliche zeitliche Stabilität aufweist. Einige Rasterzeugnisvarianten berichteten durchweg in der Vergangenheitsform; dies geschieht überwiegend unter Verwendung des Imperfekt und in wenigen Aussagebereichen darüber hinaus im Perfekt. Diese Aussageform erscheint diagnostisch korrekt, da Zeugnisse grundsätzlich »Vergangenes« beschreiben. Pädagogisch gesehen erscheint diese Zeitform dann jedenfalls irreführend, wenn »Kompetenzen« berichtet werden. Für die Schilderung von möglicherweise kritischen, jedoch überwundenen Episoden des zurückliegenden Schuljahrs kann hingegen das Tempus Imperfekt die passende Perspektive erschließen.

2.3.2 Skalierung

Bipolare Items

Die Verwendung bipolarer Items setzt voraus, daß die beiden sprachlichen Anker des Items tatsächliche »Pole« bilden, die über eine bestimmte »Metrik« voneinander geschieden sind. Zumeist muß angenommen werden, daß diese Metrik nur Aussagen darüber zuläßt, daß Rangunterschiede zwischen den Markierungen bestehen. Das Ausmaß der Unterschiede zwischen den Markierungsstufen ist nicht beurteilbar, da kein »Intervallskalenniveau« für die Beurteilung beansprucht werden kann. Bestehen zwischen den beiden sprachlichen Polen dimensionale Unterschiede, d.h. wird neben einem graduellen Güteunterschied zwischen den beiden Polen eine Polung um eine zusätzliche Gütedimension erweitert, so ist grundsätzlich unklar, wo auf der Markierungsskala diese »Qualitätserweiterung« hinzutritt. Weder Itembeantwortungen noch

Interpretationen von Itemmarkierungen sind damit systematisch möglich. In dem Beispiel »trug mit Ausdruck vor vs. hatte Mühe beim Vortrag« bezieht sich der rechte Pol auf jedwede Art des Vortragens, während der linke Pol auf »ausdrucksvolles Vortragen« eingeschränkt ist; die Itemdimension »ausdrucksvolles Vortragen« wird also erweitert um eine nicht genannte Vielzahl von Anforderungsdimensionen für »Vortragen« (z.B. hinreichende Lautstärke, thematische Qualität, emotionale Stabilität bzw. Ängstlichkeit etc.).

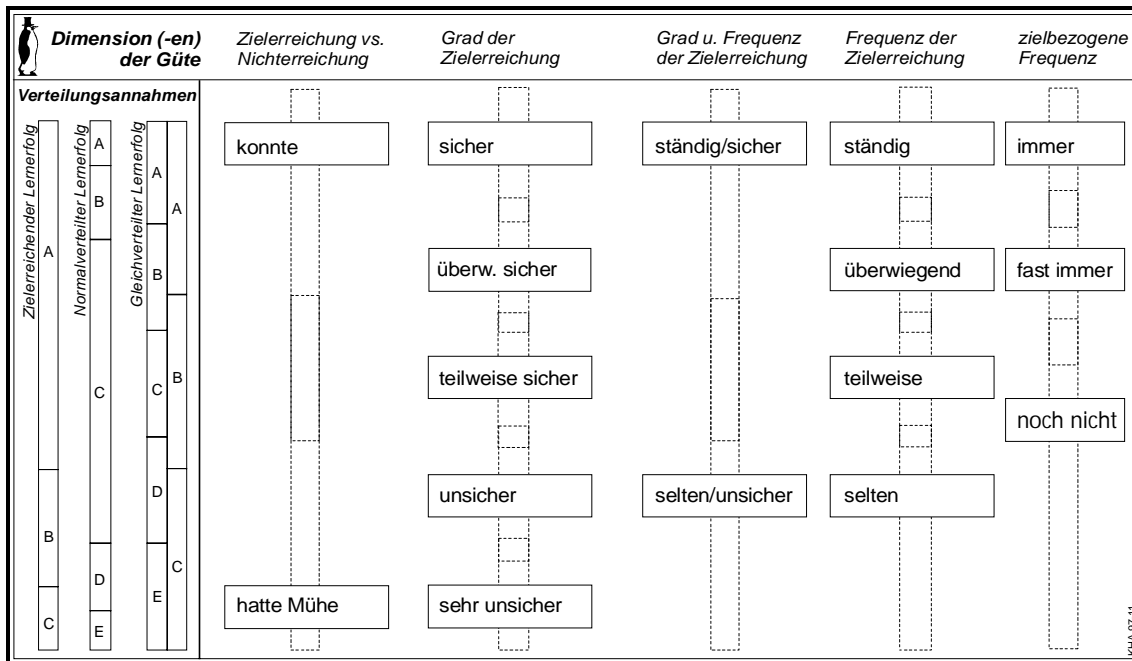
Gestufte Items

Für die Benennung der Skalenstufen werden zwei Dimensionen verwendet: das *Ausmaß der Zielerreichung* (Grad, Güte) und die *Häufigkeit der Zielerreichung*. Mischformen beider Dimensionen werden auch verwendet, was zu den oben erwähnten methodischen und sprachlichen Problemen führt. In Rasterzeugnisse werden Mischformen vermutlich deshalb eingearbeitet, weil damit ein äußerlich einheitliches Skalenformat »sichergestellt« wird für allerdings unterschiedliche Beurteilungsbereiche. Daß schulisches Lernen Zielerreichung anstrebt und dieser graduell sich nähert, ist evident. Eine pädagogische, didaktische und diagnostische Frage ist allerdings noch nicht geklärt: Welchen Sinn macht es, unter den üblichen Bedingungen von Schulunterricht über die »Häufigkeit« von Zielerreichung zu berichten?

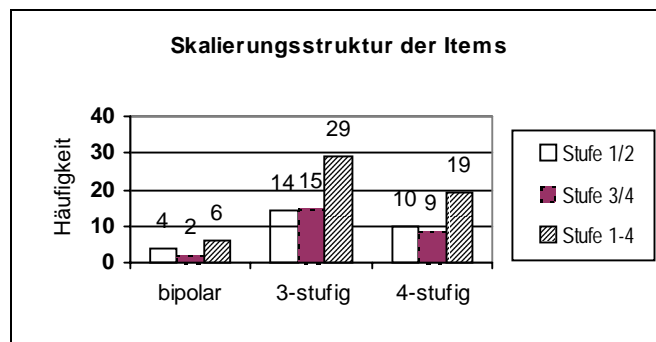
In Abbildung 1 werden die verwendeten Skalierungsbezeichnungen dargestellt und als Urteilsbereiche mit »Überlappungen« gekennzeichnet, wobei empirisch nicht bekannt ist, wie groß diese Überlappungsbereiche ausfallen. Für Ratingskalen ist jedoch mit einem beträchtlichen Fehleranteil immer dann zu rechnen, wenn keine Skalierungsregeln kodifiziert sind und gleichfalls kein Beurteilertraining stattgefunden hat, was in der Grundschule so der Fall sein wird.

Um wenigstens anzudeuten, daß alle Beurteilungen von Lernerfolg von *Vorannahmen* über die erwartete Verteilung der Resultate abhängig sind, werden drei typische und unterschiedliche Verteilungsannahmen den Skalierungen zugeordnet. Zumeist wird allerdings im schulischen Bereich Leistungsbeurteilung so praktiziert, daß die *vorgefundenen* Verteilungen von Lern- oder Prüfungsergebnissen als »Verteilungsmodell« angesehen werden. Eine bekannte Konsequenz dieser Praxis zeigt sich in der geringen Vergleichbarkeit von Ziffernzensuren zwischen Schulklassen und Schulen.

Abbildung 1. Dimensionen der Skalierung in Rasterzeugnissen



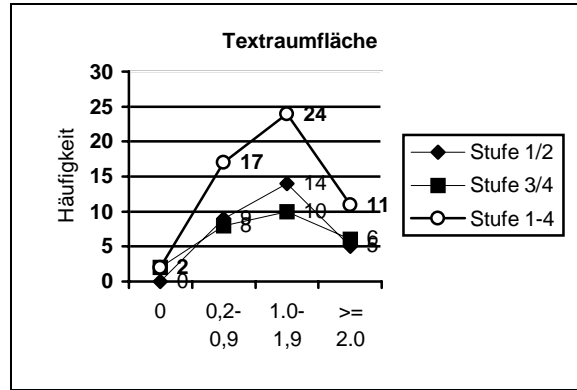
Rasterzeugnisse nutzen alternativ entweder, wie dies für die Minderzahl, d.h. für 6 der Varianten der Fall ist, *bipolare, mit sprachlichen Ankern versehene Items* verwendet (Bsp.: »beherrschte die schriftliche Addition vs. hatte Mühe damit«), oder aber, wie dies für die überwiegende Mehrzahl der Fall ist (n=48), Itemformulierungen werden vorgegeben, die mit einer sprachlich benannten Skala verbunden sind (*gestufte Items*).



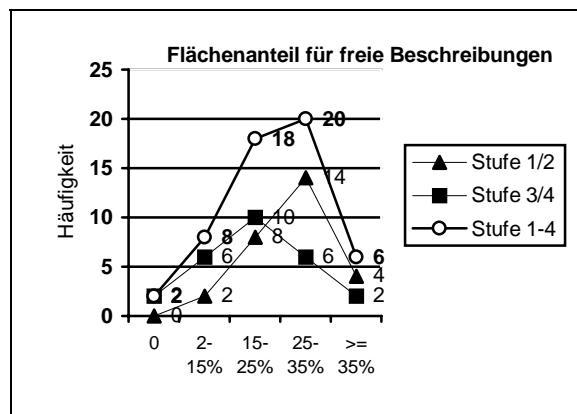
2.4 Freie Beschreibungen

In fast allen Rasterzeugnissen werden Rubriken oder Textflächen ausgewiesen, die der freien Beschreibung dienen. Welche Aspekte des Unterrichtsverhaltens, der Lernentwicklung oder des Lernstandes hier Erwähnung finden sollen, wird durchweg nicht festgelegt. An eine Verdopplung der sachnormorientierten Angaben in den Rasteritems wird hier gewiß nicht gedacht sein; vermutlich sollen jene Aspekte beschrieben werden, die in den Itemformulierungen nicht abgedeckt sind (vgl. Kap. 2.5).

Die angebotene *Textraumfläche* variiert zwischen knapp fünf Seiten (hier: DIN A5-Format) und keiner Textraumfläche. In einem Fall steht die Hälfte der insgesamt 18 Heftseiten für weitere Bemerkungen zur Verfügung, zusätzlich zu weiterhin auf den Rasterseiten ausgewiesenen Bemerkungsflächen.



Der für freie Beschreibungen vorgesehene Zeugnisflächenanteil beträgt durchschnittlich ca. 25% mit einer Streubreite von ca. 10; og. Extremfall ist dabei nicht berücksichtigt. Für die untere Stufe der Grundschule wird tendenziell mehr Bemerkungsanteil zur Verfügung gestellt.



Ob die Bemerkungsflächen allerdings faktisch genutzt werden, stellt eine offene Frage dar. Die Größe einzelner Bemerkungsflächen variiert innerhalb der Rasterzeugnisse. Häufigste Variationsquelle, so entsteht der Eindruck, scheint die Seitenauffüllung zu sein. Oder aber es werden über alle Beurteilungsbereiche hinweg gleich große Bemerkungsflächen angeboten, was wiederum anzeigt, daß keine inhaltliche Gewichtung vorgenommen wird.

2.5 Abbildung von Lernentwicklung

Die Itemstruktur der Rasterzeugnisvarianten ist durchweg nicht so angelegt, daß »Lernentwicklung« dargestellt wird, wenn damit die zeitlich ordnende Beschreibung von Lernprozessen gleicher oder unterschiedlicher Art gemeint ist. So fehlen jedwede Informationen über thematische Folgen der unterrichtlichen Arbeit; ebenso wenig sind Hinweise über die zeitliche Erstreckung von sozialen Lernprozessen oder z.B. über die auch zeitlich gestufte Folge in der Entwicklung bestimmter Arbeitsweisen anzutreffen.

Individuelle Verlaufsformen fachbezogener Lernprozesse können durch die Rasteritems nicht abgebildet werden. Diese Funktion wird auch durch »technische« Grenzen erschwert: Rasterzeugnisse müssen Items beinhalten, die möglichst für *alle* beurteilten Schüler Geltung besitzen. Wenn zu viele individuelle Variationen vorliegen bzw. einzelne Variationen zu selten anzutreffen sind, muß das Instrument versagen. Gleichwohl ist es »technisch« durchaus möglich, bestimmte Entwicklungsverläufe auch in Rasteritems abzubilden, die allerdings dann anders skaliert werden müssen (keine Gütebeurteilung, sondern dichotome Feststellung von Vorhandensein vs. Nichtvorhandensein). Unter dieser Voraussetzung ist es auch möglich, Qualitäten ge-

meinschaftlicher Lernvollzüge abzubilden und per Item zu berichten. Rasterzeugnisse sehen explizit jedenfalls nicht vor, daß Verläufe von Gruppenarbeit dargestellt werden.

Ein zentrales Problem der Rasterform wird damit angedeutet: Durch die Skalierungsstruktur, die immer eine Güteabstufung induziert, werden jene sachbezogenen, aber nicht güteorientiert vergleichbaren Lernvollzüge der Grundschule weitgehend von einer »Mitteilung« ausgeschlossen. Hier wäre die kritische Frage anzuschließen, ob damit nicht zwei wichtige Aspekte von »Lernen« unterrepräsentiert werden: Erlebnisse zu haben und Handlungen zu begehen, die bedeutsam sind und nicht vergleichbar; Abfolgen von Lernhandlungen zu zeigen, die bedeutsam und nicht leichthin typisierbar sind. Letztlich wird damit eine grundlegende Frage berührt: Welchen Stellenwert haben Zeugnisse und damit schulische Beurteilungen für die Feststellung (und Würdigung) von Unverwechselbarkeit, d.h. Individualität? Daß Rasterzeugnisse jene Individualität in der Markierungsstruktur für durchschnittlich 64 Items anzeigen, läßt folgern, daß og. Mitteilungsabsicht den Kindern gegenüber kaum bestehen kann.

2.6 Bezugsnormorientierung

Rasterzeugnisse bilden Lernverhalten und -ergebnisse ab durch sachnormorientierte Vergleiche. Manche Itemformulierungen lassen vermuten, daß auch gewisse sozialnormorientierte Beurteilungen vorgenommen werden. Dies ist insbesondere bei den frequenzbezogen skalierten Items zu unterstellen, da unklar ist, wieviel »Gelegenheit« der Unterrichtsverlauf dem einzelnen Schüler gibt, z.B. »ständig« eine bestimmte Kompetenz vorzuweisen oder zu signalisieren. Möglich sind die Frequenzurteile *relativer* Art: »ständig« kann auch bezogen sein auf die in der jeweiligen Klasse maximal gezeigte Frequenz eines bestimmten Verhaltens.

Individualnormorientierte Beurteilungen sind in den vorliegenden Rasterzeugnissen nicht vorgesehen, da – wie oben bereits erörtert – die Verlaufsperspektive und damit die Feststellung individueller Lernzuwächse nicht realisiert wird. Einige Itemformulierungen zum Arbeitsverhalten deuten auf die motivationale Komponente des Lernens hin und berichten über »Anstrengung«, deren individuelle Variation jedoch nicht kenntlich gemacht werden kann.

3 Zeugnisbeispiele

Nachfolgend werden einige Darstellungsformen von Rasterzeugnissen beispielhaft aufgeführt. Dazu werden Anmerkungen in der Absicht formuliert, Veränderungs- und Verbesserungsmöglichkeiten aufzuzeigen bzw. die Grenzen von Zeugnisberichten deutlich zu machen, die für Ziffernzeugnisse – wie eingangs erwähnt – erheblich enger ausfallen und die für Lernentwicklungsberichte nur dann weniger eng ausfallen, wenn diese die dem Konzept grundsätzlich eigenen Möglichkeiten auch tatsächlich ausschöpfen.

- a) Nicht selten werden Items so formuliert, daß in den verwendeten Polungen die Dimensionierung wechselt und somit unklar bleiben muß, auf welche Dimension die Markierung bezogen ist; die Spezifikationen der Polung sind entweder nicht konsistent, oder aber diese werden verändert. Zudem wird nicht angegeben, welche Eigenschaft die vorgegebene, dreifache Markierungsart hat: Vermutlich ist hier unterstellt, daß eine dreifache Intensitätsunterscheidung gemeint ist; Leser – und vielleicht auch Beurteilungspersonen könnten jedoch auch andere Bezüge herstellen.

Beispiel: *konnte Wörter und Texte fehlerfrei abschreiben vs. machte häufig Fehler*

- b) Negativpole folgen keinem einheitlichen Schema: Teilweise werden Verneinungen der Positivpolungen verwendet, teilweise generalisierende Formulierungen für Fähigkeitsminderung. Ein Zusammenhang mit der Verwendung von »Quantoren« besteht nicht durchweg: So wird der Quantor »sicher« z.T. verneint durch »keine sichere ..«, teilweise aber auch durch die Formulierung »Mühe haben damit«.

Beispiel: *hatte eine sichere Zahlenvorstellung bis 20 vs. hatte keine sichere Zahlenvorstellung*

- g) Der Rückgriff auf Hilfestellung durch die Lehrperson wird tw. als Kennzeichnung des Negativpols verwendet, wobei unklar bleibt, ob Hilfestellung nur für die so bezeichneten Lernfähigkeiten angeboten und Inanspruchnahme von Hilfe als Indikator für Kompetenzdefizite eingestuft wird, was pädagogisch schwierige Schlußfolgerungen für jene Schüler aufkommen läßt, die mit Hilfestellung Kompetenzprobleme längerfristig erfolgreich überwinden. Für diese Gruppierung ist die Itempolung vermutlich in beiden Polen zutreffend.

Beispiel: *konnte kleine Geschichten schreiben vs. brauchte Hilfe dabei*

- h) Teilweise werden logisch nicht oppositionelle Begriffe als Polung gegeneinandergestellt.

Beispiel: *Texte waren sinnvoll gegliedert vs. Texte waren wenig gegliedert*

- i) Die Negativpolungen verwenden überwiegend die Bezugsform »hatte Mühe damit«. Zugleich werden als Elemente von Negativpolungen folgende »Quantoren« benutzt, die wiederum durch die vorgegebene Dreifachskalierung »quantifiziert« werden sollen, was recht komplizierte Abwägungen verlangt:

*kaum, häufig (Fehler), selten, weniger;
nicht, nur (nachvollziehen),
wenig, viel (braucht ..Hilfe), keine (sichere),
einfach (vs. einfallsreich).*

- j) Der Bezug mancher Negativpolungen könnte unterschiedlich verstanden werden, wenn das Bezugswort »damit« nicht dem gesamten Iteminhalt (d.h. dem »Sachverhalt«) zugeordnet wird, sondern dem grammatikalisch primären Bezugswort, das ein Substantiv ist (im Beispiel: Bedürfnisse; ev. auch das Pronomen: andere).

Beispiel: *respektierte die Bedürfnisse anderer vs. hatte Schwierigkeiten damit*

- k) Die Polung eines Items verwendet sowohl eine Verhaltens- als auch eine Persönlichkeitsbeschreibung, wobei letztere aus diagnostischen und pädagogischen Gründen in Lernentwicklungsberichten keine Verwendung finden sollte. Mischformen sind zudem kaum dimensional interpretierbar.

Beispiel: *konnte konzentriert arbeiten vs. war unkonzentriert*

- l) Manche Itemformulierungen verwenden kritische Polungen. Die Markierung muß z.T. Abstufungen vornehmen, die ein sehr hohes Maß an Genauigkeit verlangen und die damit zugleich einen kaum nachvollziehbaren Bezugsrahmen erhalten. Die verwendete und unbekannte Dreifachskalierung kann sich kaum auf etwas anderes beziehen als auf das Unterscheidselement in den Polungen, d.h. auf »nicht immer«. Wie aber wäre eine dreifache Unterscheidung der Frequenzangabe »nicht immer« sprachlich zu fassen und wie operational, d.h. als zusammenfassende Beurteilung? Hier werden vermutlich sowohl die Grenzen hinreichender Informationsgabe als auch diagnostischer Differenzierung überschritten.

Beispiel: *Du hast selbständig Lösungswege gefunden vs. Du hast nicht immer selbständig Lösungswege gefunden.*

- m) Die Verwendung des Imperfekt kann irreführende Assoziationen begünstigen, da der Fortbestand der berichteten Kompetenzen über den Berichtszeitpunkt hinaus als fraglich erscheinen könnte. Die verwendete Skalierung soll diesen Mißverständnissen vorbeugen, was wiederum zu nicht unproblematischen Einzelbeurteilungen führt. Die Qualifizierung des Quantors »richtig« mit z.B. »sehr unsicher« vs. »unsicher« läßt die Grenzen sprachlich vermittelter Unterscheidungen rasch erkennen. Kompetenzunterscheidungen wären vermutlich günstiger mit einer Differenzierung nach dem Grad der Richtigkeit, d.h. nach dem Sachnormbezug, herstellbar, was aber dem verwendeten Skalierungsschema tendenziell zuwider liefe. Grundsätzlich zeigt sich hier die Problematik einer hierarchisch unklaren, mehrfachen Güteabstufung: Was ist die pädagogisch bedeutsamere Information – a) das Ausmaß der Richtigkeit oder b) die Vorhersagbarkeit und damit die Konsistenz des Leistungsniveaus (sicher vs. unsicher)? Zu bedenken wäre auch, ob die formal mögliche Aussagekombination »konnte etwas unsicher« nicht doch störende, logische Assoziationen nahelegt.

Beispiel: *konnte Sätze richtig formulieren sehr unsicher, unsicher, überw. sicher, sicher*

- n) Die Verwendung von Items, deren Inhalt bis auf die Bezeichnung der Anforderungshöhe gleich ist, hat erhebliche Voraussetzungen.

Beispiel *kann einfache Sachaufgaben lösen selten, tw., überwiegend*
kann schwierige Sachaufgaben lösen selten, tw., überwiegend

Einerseits müßte vorausgesetzt werden, daß die für beide Items möglichen, insgesamt 6 Antwortmuster in hierarchischer Beziehung zueinander stehen (d.h. das erste Item fungierte lediglich als Differenzierung für Stufe 1 des zweiten Items), andernfalls würde unterstellt, daß »Aufgabenschwierigkeit« und »Aufgabenlösungshäufigkeit« von in erheblicher Weise unterschiedlichen Faktoren beeinflußt werden, d.h. das Ausmaß an Kompetenz stellte nicht die primäre Varianzquelle dar. Damit wären Kompetenzbeurteilungen jedoch an situative und/oder an sachstrukturelle Faktoren geknüpft, die hohe Bedeutsamkeit haben und deshalb mitgeteilt werden müßten. Sollte die og. Itemunterscheidung allerdings einen curricularen Schulstufenunterschied anzeigen, wie dies z.B. bei der Kennzeichnung der Rechenfähigkeit in den unterschiedlichen Zahlräumen oder bei der Kennzeichnung der orthographi-

schen Leistung in unterschiedlich umfangreichen Diktaten der Fall ist, dann wären die Items wiederum einheitlich beantwortbar.

Aus testtheoretischer Sicht hingegen erscheint die logische Trennung von »Aufgabenschwierigkeit« und »individueller Lösungswahrscheinlichkeit« durchaus sinnvoll. Die Modelle der Item-Response-Theorie gründen auf dieser Unterscheidung, wohingegen die »Klassische Testtheorie« keine Trennung von Aufgabenschwierigkeit und Fähigkeit vornimmt. Konsequenz des erstgenannten Ansatzes ist allerdings, bei der Fähigkeitsbestimmung *alle* Lösungsmuster zu berücksichtigen, was schulische Lernerfolgsmessung vor sehr erhebliche Anforderungen stellte – so fungierte die übliche Addition von Richtiglösungen zu einem Gesamtwert, der Grundlage der Beurteilung sein soll, dann lediglich als ein Spezialfall einer Vielzahl alternativer Punkteverrechnungen, deren Angemessenheit zu prüfen wäre.

- o) Das dem Mathematikbereich zugeordnete Item »benötigt Anschauungsmaterial« nimmt eine Sonderstellung ein, da die Polung des Iteminhaltes invertiert ist.

Beispiel: *benötigt Anschauungsmaterial* *selten, tw., überwiegend*

Mit diesem Item wird auch von der pädagogischen Grundstruktur der Zeugnisausagen abgewichen, da hier »Defizitgrade« und nicht »Kompetenzgrade« berichtet werden. Das Item soll auf die lernpsychologisch wichtige Qualitätsstufe hinweisen, die von handlungsgebundenen Denkvorgängen hin zu deren rein kognitiver Repräsentierung führt. Eine kompetenzorientierte Itemformulierung könnte diesen Aspekt aufnehmen: »kann auch ohne Anschauungsmaterial Rechenaufgaben lösen«.

4 Entwicklungsperspektiven für Zeugnisformen im Grundschulbereich

Die Untersuchung der Varianten von Rasterzeugnisformularen zeigt die vielfältigen Gestaltungsmöglichkeiten dieser Zeugnisform auf. Rasterzeugnisse besitzen eine besondere Informationsqualität aufgrund ihrer grundsätzlich sachnormorientierten Beurteilungsstruktur, die sich allerdings bei genauerer Analyse als teilweise fraglich herausstellt. In methodischer Hinsicht können zahlreiche Rasterzeugnisvarianten erheblich verbessert werden und damit an diagnostischer Präzisierung und Verständlichkeit gewinnen. Vielleicht wird in der erziehungswissenschaftlichen Diskussion und in der schulischen Praxis von verbaler Lernbeschreibung noch zu wenig beachtet, daß hier immer auch Grundfragen von empirischer Datenerhebung tangiert werden. Dies gilt in besonderer Weise für Rasterzeugnisse, die sich unterschiedlicher formaler Skalierungen bedienen. Durch eine stärker methodisch orientierte Gestaltung und Nutzung dieser Berichtsform könnte auch eine sowohl schulpädagogische als auch wissenschaftlich diagnostische Fragerichtung entstehen, die nach Möglichkeiten einer Informationszusammenfassung für die Vielzahl scheinbar gleichrangiger Items sucht. Weitere praktisch wie theoretisch interessante Untersuchungsfragen betreffen die Rezeption von Rasterzeugnisinformation durch deren potentielle Adressaten.

In pädagogischer Hinsicht lassen Rasterzeugnisse einen deutlichen Unterschied zur Beurteilungsbreite von Textzeugnissen, die in freier Form berichten, erkennen, da zwei grundlegende Merkmale der Rastervariante kaum korrigiert werden können. Vermutlich wird es weder gelingen, die diagnostisch durchaus wichtige Differenziertheit der Berichterstattung, d.h. die Itemmenge, *erheblich* zu reduzieren, noch wird das grundsätzlich tabellarische Format der Zeugnisvariante vermeidbar sein. Damit ist allerdings auch die wenig kindgemäße Eigenschaft dieser Zeugnisinformation festgelegt und deren tendenzielle Fremdheit gegenüber einer schülerbezo-

genen und insofern lernförderlichen Diagnostik. Die eingangs dieser Untersuchung gestellte Frage kann somit in gewissem Umfang beantwortet werden: Rasterzeugnisse informieren primär nicht die Grundschüler. Ob sich Eltern mit dieser Beurteilungsform verständlich informiert fühlen, kann aufgrund der hohen Strukturiertheit und Komplexität sowie der gleichfalls hohen Informationsdichte eher bezweifelt werden. Gewiß aber werden Eltern zu schätzen wissen, daß die ihnen angebotenen Beurteilungskategorien lehrplanvalide sind und insofern ein sachnormorientierter Bezugsrahmen der Leistungsbeurteilung zumindest angestrebt wird. Insofern sind Rasterzeugnisse der Variante Ziffernzeugnisse deutlich überlegen.

Daß Rasterzeugnisse kaum dazu dienlich sind, individuelle Lernentwicklung nachzuzeichnen, daß sie – jedenfalls in den hier untersuchten Formen – wenig über Entwicklungsverläufe im Lernen berichten und daß die didaktischen Absichten von Projektunterricht und von gruppenbezogenen Lernsequenzen nicht bzw. nur vage abgebildet werden, stellt diese Zeugnisform in eine pädagogische Ausrichtung von Grundschularbeit, die hinter den aktuellen bildungsreformerischen Zielen zurückbleibt.

Hingegen sind Rasterzeugnisse besonders geeignet für fachliche Kommunikation unter Lehrern, für die diese Berichtsform eine direkte Abbildung der curricularen Grundlagen ihrer Arbeit bildet. »Wissen« über die Lernleistungen von Schülern kann so unter Fachpersonen »sparsam«, d.h. in hoher Informationsverdichtung übermittelt werden, sofern die Voraussetzung akzeptiert wird, daß ein überwiegend »produktorientiertes« Leistungsverständnis das berichtete Leistungsspektrum einschränkt und daß bestimmte didaktische Formen und Sozialformen des Lernens kaum wiedergegeben werden. Auf dieser »Expertenebene« ergibt sich eine weitere Nutzungsmöglichkeit: die unterrichtsbegleitende und zur gelegentlichen Auswertung zu nutzende Dokumentation von Lernergebnissen wird in klarer Weise durch die Rasteritems strukturiert, wobei zu beachten ist, daß hier zusammenfassende Beurteilungen und keine Beobachtungsprotokolle erstellt werden.

So gesehen könnte eine Zeugnisform avisiert werden, die die Vorteile der alternativen Varianten gleichermaßen nutzt und deren Nachteile kompensiert. Die große Chance der Textzeugnisse liegt unter diesem Gesichtspunkt noch in der Zukunft: Denkbar wäre eine Verknüpfung von verständlicher und lernförderlicher Beschreibung der Lernentwicklung mit den Aspekten der lehrplanbezogenen Berichterstattung, die drei Aspekte umfaßt: (1) den »Unterrichtsbericht« (und dessen lehrplanbezogene Beurteilung) sowie (2) die individuelle Entwicklung in diesen »sachnormorientierten« Bereichen und (3) die individuelle Entfaltung von Lernaktivitäten und die Nutzung von Lernwegen, die in dieser Struktur nicht abbildbar sind. »Integrierte Zeugnisse« könnten hier entstehen, die einen vierten Beurteilungsaspekt in pädagogisch kultivierter Form durchaus einschließen könnten: die Position im Leistungsspektrum der Klasse, deren Angabe nicht notwendig eine Sechserkala erforderlich macht, sondern mit weniger Kategorien pädagogisch vermutlich günstiger umgehen könnte.

5 Literaturverzeichnis

- Arnold, K.-H. (1997). Wege der Zeugnisreform. *Grundschule*, 29 (6), 43-45.
- Dohse, W. (1960). *Das Schulzeugnis: sein Wesen und seine Problematik*. Weinheim: Beltz.
- Helmke, A. (1992). *Selbstvertrauen und schulische Leistungen*. Göttingen: Hogrefe.
- Hentig, H. v. (1982). Das Beurteilungssystem der Bielefelder Laborschule. *Neue Sammlung*, 22, 238-260; 391-405; 481-502.
- Jürgens, E. (1992). Leistung und Beurteilung in der Schule. Eine Einführung in Leistungs- und Bewertungsfragen aus pädagogischer Sicht. Sankt Augustin: academia Richarz.
- Klafki, W. (1994). Sinn und Unsinn des Leistungsprinzips in der Erziehung. In ders. (Hrsg.), *Neuere Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik* (S. 209-249). 4., durchges. Aufl. Weinheim: Beltz.
- Kleber, E. (1992). *Diagnostik in pädagogischen Handlungsfeldern*. München: Juventa.

- Kommission »Zukunft der Bildung – Schule der Zukunft« beim Ministerpräsidenten des Landes Nordrhein-Westfalen/Bildungskommission NRW (1995). Zukunft der Bildung – Schule der Zukunft. Denkschrift der Kommission. Neuwied: Luchterhand
- Lehmann, R. & Peek, R. (1997). Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im Herbst 1996. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lübke, S. (1996). Schule ohne Noten. Lernberichte in der Praxis der Laborschule. Opladen: Leske + Budrich.
- Olechowski, R. & Persy, E. (Hrsg.). (1987). Fördernde Leistungsbeurteilung. Ein Symposium. Wien: Jugend und Volk.
- Sacher, W. (1994). Prüfen – beurteilen – benoten. Theoretische Grundlagen und praktische Hilfestellungen für den Primar- und Sekundarbereich. Bad Heilbrunn: Klinkhardt.
- Schulbegleitforschungsprojekt »Integrierte Leistungsbeurteilung in der Sekundarstufe I« (1997). Das Problem – die Fragestellung – das Ziel – Projektschulen und Projektklassen – Projektanbindung. Beschreibung und wissenschaftliche Begründung des Forschungsprojektes. Bremerhaven: Schulpsychologischer Dienst (Projekt Nr. 87 beim Senator für Bildung, Bremen).
- Senator für Bildung und Wissenschaft (1993). Innovation und Kontinuität. Empfehlungen zur Schulentwicklung in Bremen. Bericht der Kommission zur Weiterführung der Schulreform in Bremen. Bremen: Senator für Bildung und Wissenschaft.
- Senator für Bildung und Wissenschaft (1997). Verordnung für Zeugnisse und Lernentwicklungsberichte und über die Abschlüsse an öffentlichen Schulen (Zeugnisordnung) vom 14.7.1997. Gesetzblatt der Freien Hansestadt Bremen (S. 247-254).
- Rosenfeld, H. & Valtin, R. (1997). Zur Entwicklung schulbezogener Persönlichkeitsmerkmale bei Kindern im Grundschulalter. Erste Ergebnisse aus dem Projekt NOVARA. Unterrichtswissenschaft, 25, 316-330.
- Zielinski, W. (1974). Die Beurteilung von Schülerleistungen. In F.E. Weinert, C.F. Graumann & M. Hofer (Hrsg.), Funk-Kolleg Pädagogische Psychologie (S. 877-900). Frankfurt a. M.: Fischer Taschenbuch Verlag.