

Differentielle Entwicklungsverläufe der Fähigkeitsselbstkonzepte im Grundschulalter in Abhängigkeit von der Bewertung mit und ohne Noten

Corinna Schmude

Empirisch gesicherte Befunde zu spezifischen Effekten von Leistungsrückmeldungen auf die Entwicklung der Fähigkeitsselbstkonzepte finden sich in der Literatur nur im Rahmen gezielter experimenteller Untersuchung (z. B. Lissmann, 1981, 1984) und im Kontext von Studien zu speziellen Rahmenbedingungen des Lernens, wie z. B. zu positiven Sozialisationseffekten in Gesamtschulen (Schwarzer/Royl, 1984) oder zum Einfluss leistungsbezogener Besorgnis und Ängste von Eltern auf die emotionale Entwicklung der Kinder (Gartner/Jirasko, 1999). Sowohl Lissmann (1981, 1984) als auch Schwarzer/Royl (1984) und Gartner/Jirasko (1999) berichteten dabei theoriekonform über persönlichkeitsfördernde Effekte des differenzierenden, individuellen Umgangs mit Leistungen auf die Entwicklung der Fähigkeitsselbstkonzepte.

Die Frage, ob sich verbale Beurteilungen, wie postuliert, günstiger auf die kindliche Persönlichkeitsentwicklung auswirken als die Notengebung, wurde erstmalig im Rahmen des Forschungsprojektes NOVARA (**N**oten- oder **V**erbalbeurteilungen – **A**kzeptanz, **R**ealisierung, **A**uswirkungen) systematisch bearbeitet (Valtin et al., 2002). Das Forschungsprojekt begleitete in einer Längsschnittstudie Klassen mit Notenbewertung bzw. Verbalbeurteilung ab Klasse 2. Neben der Akzeptanz beider Beurteilungsformen aus Sicht der Betroffenen und der Realisierung der Berichtszeugnisse wurde die Entwicklung der schulrelevanten Persönlichkeitsmerkmale, u. a. die der Fähigkeitsselbstkonzepte, sowie die der Schulleistungen untersucht. Im Folgenden soll kurz die theoretische Konzeptualisierung des Konstruktes ‚Fähigkeitsselbstbild‘ dargestellt und über die Hauptergebnisse des Forschungsprojektes NOVARA, unter Bezugnahme zur einschlägigen Literatur, berichtet werden.

Konzeptualisierung ‚Fähigkeitsselbstkonzept‘

Wie Krapp (1997) darlegt, ist das Selbstkonzept als Konstrukt weder theoretisch noch empirisch eindeutig zu charakterisieren. Im Projekt NOVARA wurde die Konzeptualisierung des Konstruktes der LOGIK- und SCHOLASTIK-Studie übernommen. Damit wird vom Selbstkonzept als einer internen, kognitiven Repräsentation des Wissens über sich selbst und die darin eingeschlossenen Überzeugungen ausgegangen (vgl. Krapp, 1997).

Das von Helmke (1992) entwickelte Instrument zur Erfassung der Selbsteinschätzung der eigenen Kompetenz erhebt diese in Bezug auf die Mitschüler des Kindes und wird deshalb als

relatives Fähigkeitsselbstkonzept bezeichnet. Das relative Fähigkeitsselbstkonzept ist ein Aspekt des leistungsbezogenen Selbstvertrauens, welches definiert wird als (bereichsspezifisches) Vertrauen in die eigene Fähigkeit, gegenwärtige wie zukünftige unterrichtliche und fachliche Anforderungen bewältigen zu können. In diese Widerspiegelung der ‚subjektiven Kompetenz‘, wie man das leistungsbezogene Selbstvertrauen auch bezeichnen könnte, fließen sowohl Reflexionen über die eigene Begabung im Allgemeinen und bezogen auf den aktuellen Lerngegenstand im Besonderen als auch diagnostische und prognostische Elemente ein. Das relative Fähigkeitsselbstkonzept bildet dabei den komparativen Aspekt des leistungsbezogenen Selbstkonzeptes ab und zwar über den expliziten sozialen Vergleich mit den Mitschülerinnen und Mitschülern. Es wird erfasst über die Beurteilung der momentanen relativen Leistungsposition in dem entsprechenden Lernbereich. „Die Vorgabe eines expliziten sozialen Vergleichsmaßstabes“, so der Autor (Helmke, 1997, S. 495), „wurde gewählt, um Konfundierungen interindividuell verschiedener Vergleichsmaßstäbe (...) so gering wie möglich zu halten.“

Theoriekonform beurteilen die Kinder am Beginn von Klasse 2 ihre Fähigkeiten in Rechnen und Rechtschreiben sehr positiv (Valtin/Wagner, 2002). Die in der Literatur beschriebene deutliche Abnahme dieser günstigen Fähigkeitsselbstbilder im Verlauf der weiteren Schulzeit zeigte sich im Rahmen des Forschungsprojektes NOVARA nicht in der zu erwartenden Deutlichkeit (Rosenfeld/Valtin, 1997; Wagner/Valtin, 2003). Dagegen bestätigte sich die Geschlechtsspezifität der Entwicklungsverläufe: Mädchen weisen günstigere Fähigkeitsselbsteinschätzungen in Rechtschreiben, Jungen dagegen in Rechnen auf. Des Weiteren zeigte sich, dass leistungsstärkere Kinder positivere Selbsteinschätzungen vornehmen als leistungsschwächere (Rosenfeld/Valtin, 1997, Valtin et al., 1996; Valtin/Wagner, 2002; Wagner/Valtin, 2003).

Der erhoffte positive Effekt verbaler Beurteilung auf die Entwicklung der Fähigkeitsselbstkonzepte konnte jedoch nicht belegt werden (Valtin/Wagner, 2002, Wagner/Valtin, 2003, S. 27). Es deutete sich sogar ein gegenläufiger Effekt an: Kinder, die mit Noten bewertet wurden, nahmen günstigere Fähigkeitsselbsteinschätzungen in Mathematik vor als verbal beurteilte Kinder. Dies galt auch, wenn die zu vergleichenden Gruppen hinsichtlich ihrer Leistungen in Schulleistungstests in zwei Leistungsgruppen gesplittet wurden. Dies widerspricht der hypothetischen Annahme, dass vor allem die Selbstkonzeptentwicklung leistungsschwächerer Kinder unter der Bedingung verbaler Beurteilung günstiger verläuft (Wagner/Valtin, 2003, S. 33).

Dass sich keine entwicklungsfördernde Wirkung der verbalen Beurteilung gegenüber der Notengebung zeigt, erscheint vor dem Hintergrund der Befunde zur Evaluation der Berichtszeugnisse einerseits durchaus schlüssig. Die Realisierung der Berichtszeugnisse kann bisher nicht als zufriedenstellend bezeichnet werden (Schmude, 2001). : Der streng gegliederte formale und inhaltliche Aufbau der evaluierten Berichtszeugnisse wies deutliche Parallelen zum Aufbau des klassischen Notenzeugnisses auf. Es handelte sich mehrheitlich um an traditionellen Beurteilungsaspekten orientierte Statusberichte mit geringem Informationsgehalt. Die Berichtsfunktion des Zeugnisses besitzt Priorität und nicht, wie intendiert, die ermutigende und förderdiagnostische Funktion von Leistungsbeurteilungen. Somit ist anzunehmen, dass die mit der Einführung dieser Beurteilungsform verbundenen Intentionen in der schulischen Praxis auch kaum zum Tragen kommen. Andererseits überrascht der Befund dennoch. Man hätte erwarten können, dass allein der Verzicht auf Noten bereits zu einer allgemeinen Entspannung der Lernsituation führt, die sich in einer stabilen und positiven Einschätzung der eigenen Kompetenzen äußert. Dass sich dies hier nicht zeigt, könnte zwei Gründe haben: (1) Die längere Grundschulzeit mildert die postulierten negativen Effekte der Notengebung durch das spätere Einsetzen des Selektionsdrucks. (2) Das Rückmeldeverhalten der Lehrkräfte unterscheidet sich nicht in Abhängigkeit von der als Zeugnis gewählten Beurteilungsform. Letzteres wird durch die Befunde aus dem Projekt NOVUS (**Noten oder Verbalbeurteilung: Unterrichtsorganisation und Sanktionsverhalten von Lehrkräften in Ost- und Westberliner Grundschulen**; Wagener, 2003) gestützt: Lehrkräfte, die Verbalbeurteilungen oder Noten präferieren, unterscheiden sich nur wenig in Bezug auf die Unterrichtsgestaltung und so gut wie gar nicht in ihrem Sanktions- und Rückmeldeverhalten. Daraus folgt, dass sich die Lernsituation der Kinder in Abhängigkeit von der Beurteilungsform nicht unterscheidet und dementsprechend auch keine unterschiedlichen Entwicklungsverläufe bedingt werden können. Pekrun (1996) wies in einer Stellungnahme zum Projekt NOVARA daraufhin, dass die gewählte Form der Beurteilung – ob mit oder ohne Noten – eben nur ein und dazu ein offensichtliches und oberflächliches Merkmal einer Leistungsbeurteilung darstellt und nichts über die tatsächliche Beurteilungspraxis aussagt (Pekrun, 1996, S. 254). Die im Rahmen des Forschungsprojektes vorgenommene Evaluation von Berichtszeugnissen (Schmude, 2001) und die Unterrichtsbeobachtungen (Wagener, 2003) konnten diese Lücke teilweise schließen und nachhaltig belegen, dass allein der Verzicht auf Noten die pädagogische Praxis nicht wesentlich verändert.

Neben dieser Überlegung, inwieweit die Variable ‚Beurteilungsform‘ inhaltlich für die Definition der Vergleichsgruppen geeignet ist, stellt sich die Frage, ob allein auf Mittelwerten

basierende Vergleiche von über die Designvariablen festgelegten Äquivalenzklassen Entwicklungsverläufe tatsächlich zutreffend abbilden können. Asendorpf (1995) weist daraufhin, dass sich zwar auf der Basis variablenzentrierter Analysen durchschnittliche Verlaufsgestalten identifizieren lassen, ihre Verallgemeinerung jedoch der Prüfung der intraindividuellen Gültigkeit des Verlaufs für alle Mitglieder der betrachteten Stichprobe bedarf. Zweifellos sind das Geschlecht und das Leistungsniveau der Kinder Faktoren, die zu beobachtende Unterschiede erklären, wie die Arbeiten Helmkes (Helmke, 1997, 1998) und die Publikationen aus dem Forschungsprojekt NOVARA belegen (z. B. Wagner/Valtin, 2003). Aber auch Helmke (1996, S. 175) und Valtin (2000, S. 20) verweisen auf die Notwendigkeit, im Rahmen der pädagogisch-entwicklungspsychologisch orientierten Forschung diesen methodischen Überlegungen der persönlichkeitspsychologischen Forschung Rechnung zu tragen. In jeden der vorgenommenen Vergleiche auf der Ebene einer der Designvariablen gehen die Kinder auch mit Merkmalen ein, die ihre Zuordnung zu anderen Äquivalenzklassen ermöglichen. Mittels mehrfaktorieller Varianzanalysen kann natürlich die Möglichkeit bestehender Wechselwirkungen zwischen den Designvariablen überprüft werden. Allerdings verlässt man auch mit diesem Verfahren nicht den variablenzentrierten, an methodisch festgelegte Stichproben gebundenen Analyseansatz, der Gesetzmäßigkeiten durch die Mittelwerte der Stichproben beschreibt und individuelle Besonderheiten als Messfehler betrachtet (Asendorpf, 1995, S. 243). Bestimmte Kombinationen der Merkmale der Äquivalenzklassen können ganz eigene Gesetzmäßigkeiten aufweisen und damit eine eigene Äquivalenzklasse definieren, die dann in Abhängigkeit von der Anzahl ihrer Mitglieder das Analyseergebnis beeinflussen. Vor diesem Hintergrund wird deutlich, dass es äußerst problematisch ist, von der durchschnittlichen Verlaufsgestalt einer Variablen auf der Stichprobenebene auf die intraindividuelle Gültigkeit dieses Verlaufs für alle Mitglieder der betrachteten Äquivalenzklasse zu schließen. Zulässig ist dies eigentlich nur dann, wenn vom variablen- zum personenzentrierten Ansatz gewechselt wird, d. h. wenn für alle Stichprobenmitglieder die ‚funktionelle Homogenität‘ der gefundenen Gesetzmäßigkeit überprüft wird (Asendorpf, 1995, S. 243). In der Regel ist es allerdings nicht leistbar für jede Person einer untersuchten Stichprobe abzugleichen, ob bzw. in welchem Maße ihre persönliche Merkmalsausprägung der der Gesamtstichprobe entspricht.

Einen Kompromiss zwischen variablen- und personenzentrierter Analysestrategie stellt das von Helmke (1996, S. 181) vorgeschlagene methodische Herangehen dar. Helmke (1996) schlägt vor, über die Klassifizierung der Kinder nach der Verlaufsgestalt der Entwicklungen der Variablen mit Hilfe von Clusteranalysen Zeitreihen zu untersuchen. Mit dieser Methode könnte man feststellen, ob sich die Äquivalenzklassen vollständig reproduzieren lassen, oder aber neue

Äquivalenzklassen entstehen und inwieweit diese mit den Designvariablen zusammenhängen. Ersteres würde bedeuten, dass die Designvariablen tatsächlich homogene Vergleichsgruppen definieren, letzteres würde die Überlegungen Asendorpfs (1995) stützen, dass neben den Designvariablen andere bzw. Kombinationen der Designvariablen für die Erklärung der Entwicklungsverläufe herangezogen werden sollten.

Helmke identifizierte auf diesem Wege sechs unterschiedliche Verlaufstypen der Entwicklung mathematischer Fähigkeitsselbstbilder, deren Erklärung allerdings weiterführenden Analysen vorbehalten bleibt (Helmke, 1998, 125).

Fragestellung und Hypothesen

Im Rahmen der qualitativen Inhaltsanalyse von Berliner Berichtszeugnissen hinsichtlich ihrer pädagogischen Qualität wurde anhand einer Teilstichprobe aus dem Projekt NOVARA auch untersucht, ob sich in Abhängigkeit der Beurteilungsform Unterschiede in der Entwicklung der Fähigkeitsselbstbilder von Schülerinnen und Schülern in Rechnen und Rechtschreiben zeigen (Schmude, 2001). Im Mittelpunkt des vorliegenden Beitrags steht die Beantwortung dreier Fragen: (1) Zeigen sich in der betrachteten Teilstichprobe im variablenzentrierten Vergleich der Entwicklungsverläufe Effekte der Beurteilungsform? (2) Verbergen sich hinter diesen variablenzentrierten Entwicklungsverläufen differentielle Entwicklungsverläufe der Fähigkeitsselbstbilder in Rechnen und Rechtschreiben? Und (3) inwieweit lassen sich diese differentiellen Entwicklungsverläufe mit den Designvariablen in Beziehung setzen?

Bezogen auf diese Fragestellung wurden folgende Hypothesen, die sich aus dem bisher dargestellten ableiten lassen, geprüft:

1. Unabhängig von den bisher aus dem Projekt NOVARA vorliegenden Befunden wird zunächst davon ausgegangen, dass sich der entwicklungspsychologisch und erziehungswissenschaftlich gut zu begründende positive Effekt der verbalen Beurteilung zeigt: Die relativen Fähigkeitsselbstkonzepte unter der Bedingung verbaler Beurteilungen weisen eine deutlich positivere und stabilere Entwicklung auf als in den Klassen mit Benotungen.

2. Es existieren neben den durch klassische Mittelwertvergleiche beschreibbaren Unterschieden in den Entwicklungen der Fähigkeitsselbstkonzepte weitere differenzielle Entwicklungsverläufe.

3. Es wird postuliert, dass sich auch bei diesen differenziellen Verlaufstypen die beobachteten geschlechts- und leistungsniveauabhängigen Effekte bestätigen, sich darüber hinaus aber Entwicklungsverläufe zeigen, die (a) auf kombinierte Zusammenhänge zwischen den Designvariablen schließen lassen bzw. (b) die sich nicht mit Hilfe der Designvariablen erklären lassen.

Die hier im Weiteren vorzustellenden Analysen können diesen Fragen jedoch nur explorierend nachgehen. Dies begründet sich zum einen durch die hinsichtlich der differenziellen Fragestellungen kleinen Stichprobe, zum anderen durch die eingesetzten Befragungsinstrumente. In Absprache mit dem Autor des Untersuchungsmaterials (Helmke) konzentrierte sich aus zeitökonomischen Gründen im Projekt NOVARA die Erfassung der Selbsteinschätzungen für die Lernbereiche ‚Mathematik‘ und ‚Deutsch‘ auf die fachlichen Anforderungen ‚Rechnen‘ bzw. ‚Rechtschreiben‘ (vgl. Methode).

Methode

Stichprobe

Die in diesem Artikel vorgestellten Ergebnisse beruhen auf der Analyse von Daten aus dem Forschungsprojekt NOVARA. Die längsschnittlich angelegte Studie umfasste 5 halbjährliche Messzeitpunkte vom Beginn des 2. Schuljahres im Herbst 1994 bis zum Beginn des 4. Schuljahres im Herbst 1996. Am 1. Messzeitpunkt (1. MZP) nahmen insgesamt 434 Kinder (228 Mädchen, 206 Jungen) teil. Diese stammten aus 41 Berliner Grundschulklassen, wobei jeweils 7 bis 12 Kinder je Klasse in Interviews befragt wurden. An den späteren Messzeitpunkten, an denen vor allem Gruppenbefragungen erfolgten, erhöhte sich die Teilnehmerzahl auf 550 bis 650 Kinder (zu weiteren Details vgl. Valtin et al. 2002).

Für den nachfolgenden Beitrag wird eine Teilstichprobe von Kindern betrachtet. Die Untersuchung von Entwicklungsverläufen ist an die Analyse längsschnittlich erhobener Daten gebunden. Somit wurden in die Analysen nur Kinder einbezogen, die an der gesamten Studie teilgenommen haben. Dabei handelt es sich um 185 Kindern, die, mit vereinzelt Ausnahmen, an allen vier zu betrachtenden Messzeitpunkten teilgenommen haben. Vereinzelt fehlende Befragungsergebnisse wurden dabei toleriert, da sich die Stichprobe ansonsten um weitere 39 Kinder verringert hätte. Bei der Auswahl der Kinder wurden, neben dem Geschlecht, folgende Designvariablen berücksichtigt: Beurteilungsform (Noten oder verbale Beurteilung), Einschätzung des Leistungsvermögens am Beginn von Klasse 2 durch die Klassenlehrerin (Zugehörigkeit zu einer von drei Leistungsgruppen) und Schulstandort (West- und Ostberlin) (s.

Tabelle 1). Von den ausgewählten Kindern wurden 121 (50 Mädchen, 71 Jungen) auch in Klasse 2 weiterhin verbal beurteilt, 64 erhielten ab der 2. Klasse Noten (je 32 Mädchen und Jungen). Da für die folgenden Auswertungen die Aspekte ‚Leistungsgruppe‘ und ‚Beurteilungsform‘ besondere Relevanz besitzen, wurde zunächst die diesbezügliche Häufigkeitsverteilung geprüft. Das Prüfergebnis weist für die Verteilung nach Beurteilungsform und Leistungsgruppe für die Gesamtstichprobe ein Überwiegen verbal beurteilter Kinder aus, sofern auch hier die Irrtumswahrscheinlichkeit auf $p = 0.10^1$ festgesetzt wird: $\chi^2(2, 185) = 5.11, p = .07$. Die signifikant werdende Verschiebung der Häufigkeitsverteilung erklärt sich durch den geringen Anteil an mit Noten bewerteten Kindern aus dem Westteil der Stadt am Beginn der Studie, $\chi^2(1, 185) = 10.41, p < .01$. Differenziert man die Verteilungsprüfungen nach Ost und West, so zeigt sich, dass sowohl für den Vergleich nach Beurteilungsform und Leistungsgruppe für Ostberlin, $\chi^2(1, 105) = 2.71, p > .10$, als auch für Westberlin, $\chi^2(1, 80) = 3.47, p > .10$, der Chi²-Test die Verteilung als akzeptabel ausweist. Darüber hinaus kann die Stichprobe hinsichtlich des Geschlechtes als gleichverteilt betrachtet werden, $\chi^2(1, 185) = 1.27, p > .10$.

Tabelle 1. Zusammensetzung der Teilstichprobe (N = 185)

Beurteilungsform	Schulstandort	Leistungs- gruppe	männlich	weiblich	Gesamt	
Noten	Ost	1	8	5	13	
		2	9	12	21	
		3	8	5	13	
	West	1	2	3	5	
		2	2	3	5	
		3	3	4	7	
		Gesamt		32	32	64
	Verbale Beurteilung	Ost	1	14	8	22
			2	16	12	28
3			6	3	9	
West		1	14	11	25	
		2	14	11	25	
		3	7	5	12	
		Gesamt		71	50	121

¹ Für die gesicherte Feststellung vergleichbarer Häufigkeitsverteilungen wurde eine Irrtumswahrscheinlichkeit von $p = 0.10$ festgesetzt. Daraus folgt, dass hinsichtlich der Designvariablen Geschlecht/Herkunft je Zelle der Vierfelder-Tafel die α -adjustierte Irrtumswahrscheinlichkeit $p = 0.025$ beträgt.

Erhebungsinstrumente

Die Erfassung der bereichsspezifischen relativen Fähigkeitsselbstkonzepte im Rahmen des Forschungsprojektes NOVARA erfolgte auf der Basis der in der LOGIK- bzw. SCHOLASTIK-Studie (vgl. Weinert/Helmke, 1997) verwendeten Materialien, die unter Beratung von Helmke projektspezifisch überarbeitet wurden. In der Interviewsituation wurde die leistungsbezogene Rangreihe der Klasse durch fünf Spielfiguren, wie man sie beispielsweise bei ‚Mensch-ärgere-dich-nicht‘ verwendet, symbolisiert. Das Kind wurde aufgefordert, sich (symbolisiert durch eine andersfarbige Spielfigur) in diese Rangreihe einzuordnen, in dem es sich neben ‚den/die Beste‘, ‚die Besten‘, ‚Mittelguten‘, ‚die Schlechteren‘ oder ‚den/die Schlechteste‘ stellt. Für die Befragungen ab dem zweiten Messzeitpunkt wurde ein Fragebogen eingesetzt. Die dort abgebildeten Rangreihen stehen symbolisch für die Klasse, berücksichtigen aber nicht die konkrete Klassengröße. Die Darstellung der Klasse geht auf die Nicholls-Skala (1978) zurück, die aus schematisierten Gesichtern besteht, in denen weder Geschlecht noch unterschiedliche Ausdrucksformen erkennbar sind. Die Position ‚der/des Besten‘ bzw. ‚Schlechten‘ wird durch ein solches Gesicht veranschaulicht, die Gruppen ‚der Besten‘ bzw. ‚Schlechteren‘ durch vier und die der ‚Mittelguten‘ durch 11 Gesichter. Daraus resultieren fünf Antwortstufen: der/die Beste (4), bei den Besten (3), mittelgut (2), bei den Schlechteren (1), der/die Schlechteste (0). Die genannten Konstrukte wurden bereichsspezifisch erfasst. Diese Vorgehensweise folgte dem aktuellen Forschungsstand, der die bereichsspezifische Entwicklung der Einschätzung der eigenen Kompetenzen belegt (vgl. Helmke, 1997). Im Rahmen der LOGIK- und SCHOLASTIK-Studie wurden dazu unterschiedliche fachliche Anforderungen je Lernbereich erfasst und dann in einem bereichsspezifischen Skalenwert zusammengefasst. Der Autor berichtet für diese Skalen zufriedenstellende Reliabilitäten $> .70$ (vgl. Helmke, 1992, S. 92).

Im Rahmen des Projektes NOVARA wurde je Lernbereich eine spezifische fachliche Anforderung ausgewählt, wie die nachstehenden Items veranschaulichen:

Items:

„Wie gut bist du, verglichen mit den anderen in der Klasse, im Rechnen?“

„Wie gut bist du, verglichen mit den anderen in der Klasse, in Rechtschreiben?“

Für die Verlaufsanalyse wurden hinsichtlich ‚Rechnen‘ die Daten der Messzeitpunkte 1 bis 4 und hinsichtlich ‚Rechtschreiben‘ die der Messzeitpunkte 2 bis 4 betrachtet:

1. MZP = Beginn des 2. Schuljahres (Herbst 1994), entspricht 2. Kl. (I);

2. MZP = Beginn des 2. Halbjahres des 2. Schuljahres (Frühjahr 1995), entspricht 2. Kl. (II);

3. MZP = Beginn des 3. Schuljahres (Herbst 1995), entspricht 3. Kl. (I);
4. MZP = Beginn des 2. Halbjahres des 3. Schuljahres (Frühjahr 1996), entspricht 3. Kl. (II).

Das spätere Einsetzen der Befragung zur Rechtschreibung begründet sich durch die Spezifik des Lerngegenstandes und der Gestaltung des Rechtschreibunterrichts in der Praxis. Zahlreiche Forschungsergebnisse der letzten Jahre belegen, dass es sich beim Erlernen der Rechtschreibung nicht um einen mechanischen Prozess des Einprägens, sondern um einen kognitiven Entwicklungsprozess handelt (vgl. Valtin, 2000a, S. 17). Dies bedeutet, dass ‚Rechtschreibung‘ nicht als ein separater Lerngegenstand und damit eigenständiger Lernbereich zu betrachten ist, sondern als ein Aspekt des Sprachunterrichts in der Grundschule, der dementsprechend in den Prozess der Sprachbildung integriert werden muss. Moderne fachdidaktische Konzepte berücksichtigen dies im Anfangsunterricht. Da dies unter Umständen dazu führen kann, dass die von uns befragten Kinder ‚Rechtschreiben‘ erst mit fortschreitendem Schulbesuch als gesonderte fachliche Anforderung erleben, erschien es ratsam, erst zu Beginn des 2. Halbjahres der 2. Klasse ‚Rechtschreibung‘ als spezifischen Lerngegenstand in die Befragung mit aufzunehmen.

Statistische Analysen

Die Verlaufsanalyse zu Unterschieden und Veränderungen in den Fähigkeitsselbsteinschätzungen erfolgte mittels ein- und mehrfaktoriellen Varianzanalysen mit Messwiederholung (mindestens drei MZP). Für die Signifikanzprüfungen zu einem Erhebungstermin wurde der U-Test nach Mann und Whitney eingesetzt. Die unabhängigen Variablen (Designvariablen) waren Beurteilungsform (Noten, verbale Beurteilung), Geschlecht und Leistungsniveau (Zugehörigkeit zum oberen, mittleren bzw. unteren Klassendrittel aus Sicht der Klassenlehrerin am Beginn von Klasse 2). Die Datenanalyse bezog sich auf die Einzelitems Fähigkeitsselbstkonzept in Rechnen bzw. Rechtschreiben.

Etwas ausführlicher soll an dieser Stelle auf das eingesetzte clusteranalytische Verfahren zur Identifikation differenzieller Entwicklungsverläufe eingegangen werden.

Eingesetzt wurde die K-Means-Methode, da sie im Gegensatz zu den hierarchischen Methoden einzelne Messausfälle toleriert (vgl. Bortz, 1993). Die Anzahl der Messausfälle variiert in Abhängigkeit vom Messzeitpunkt zwischen 8 bis zu 17 Kindern.

Die Clusteranzahlen werden bei dem gewählten Verfahren, im Gegensatz zu den hierarchischen Verfahren, vorgegeben und über Schwer- bzw. Mittelpunkte definiert. Die Stärke des Verfahrens liegt eigentlich in der Verfeinerung einer mit hierarchischen Methoden ermittelten Gruppenbildung. Wenn sie aber wie hier zur Exploration des Datensatzes genutzt wird, ist

unbedingt zu beachten, dass die Reihenfolge der Objekte die Clusterlösungen beeinflusst. Dies kann über Variation der Startpositionen kontrolliert werden (Bortz, 1993, S. 536).

Berechnet wurden die Clusterlösungen für die Vorgaben Zwei-, Drei-, Vier-, Fünf- und Sechs-Clusterlösung. Die Sechs-Clusterlösung als obere Grenze wurde aufgrund der Befunde Helmkes gewählt, der, wie dargestellt, über eine Sechs-Clusterlösung zur optimalen Beschreibung differenzieller Entwicklungsverläufe des relativen Fähigkeitsselbstbildes in Mathematik berichtet (vgl. Helmke, 1998, S. 125).

Die Beurteilung der aus den Clusteranalysen resultierenden Klassifizierungen individueller Entwicklungsverläufe erfolgte anhand dreier Kriterien: (1) varianzanalytische Überprüfung der vorgeschlagenen Clusterlösungen, (2) Besetzung der Cluster ($> 10\%$ der Gesamtstichprobe) und (3) Interpretierbarkeit mit Hilfe der Designvariablen (vgl. Schmude, 2001). Zur Beschreibung der Cluster hinsichtlich der Designvariablen wurden mittels Chi²-Tests die Verteilungen der Merkmale Beurteilungsform, Geschlecht und Leistungsniveau über die Cluster analysiert. Es sei an dieser Stelle darauf hingewiesen, dass alle im Folgenden als signifikant ausgewiesenen Abweichungen von den erwarteten Verteilungen auf α -adjustierten Irrtumswahrscheinlichkeiten und entsprechend korrigierten standardisierten Residuen für die einzelnen Zellen der Kreuztabellen basieren.

Ergebnisse

Entwicklung der relativen Fähigkeitsselbstbilder.

In Abhängigkeit von der Beurteilungsform zeigten sich keine signifikanten Unterschiede im Verlauf der Entwicklung der relativen Fähigkeitsselbstkonzepte. Bei keiner der beiden betrachteten Variablen erreichten die Haupteffekte ‚Beurteilungsform‘ und ‚Messzeitpunkt‘ bzw. die Wechselwirkung beider das Signifikanzniveau. Somit stellen sich auf der Basis der Mittelwerte die Entwicklungsverläufe der Fähigkeitsselbstkonzepte als stabil positiv dar (Abb. 1).

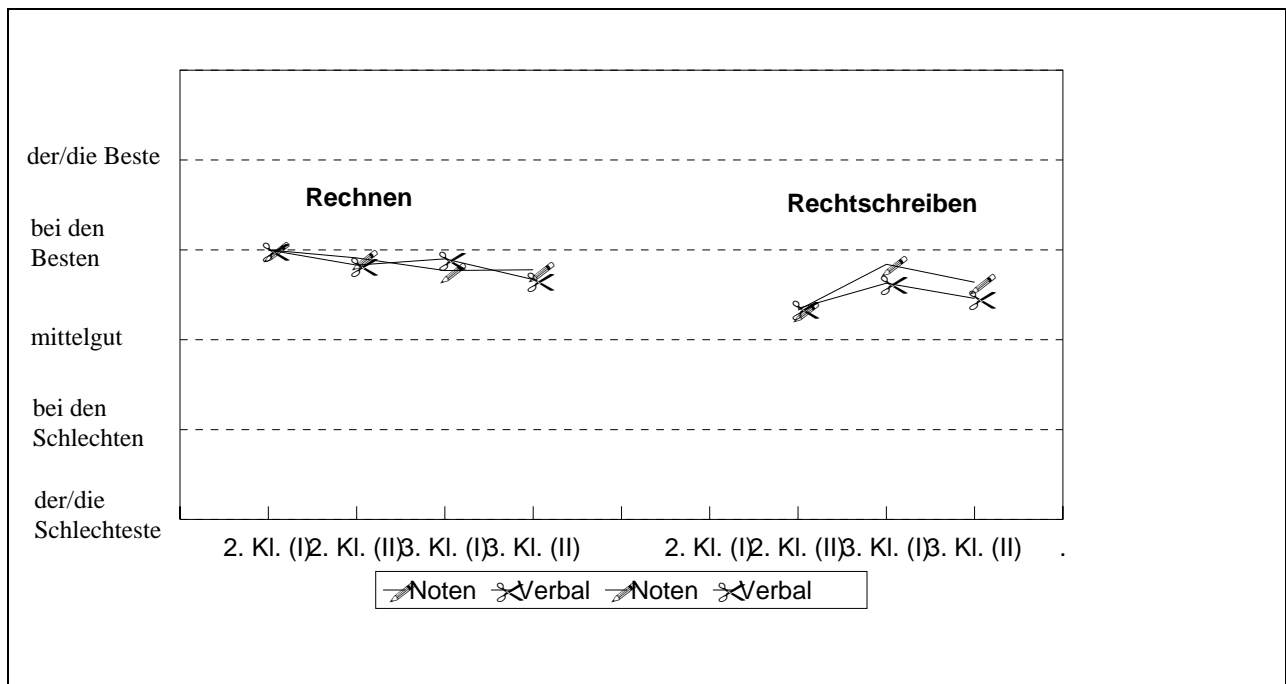


Abb. 1 Entwicklungsverläufe der relativen Fähigkeitsselbstkonzepte in Rechnen und Rechtschreiben in Abhängigkeit von der Beurteilungsform in Klassenstufe 2 ($N_{\text{Noten}} = 64$, $N_{\text{Verbal}} = 121$)

Für das relative Fähigkeitsselbstkonzept in Rechnen bestätigte sich ein signifikanter Effekt der Geschlechtszugehörigkeit ($F_{\text{Geschlecht}}(1) = 8.16$, $p < .01$). Im Einzelvergleich je Messzeitpunkt zeigte sich, dass die Jungen zu allen vier Befragungsterminen über eine positivere Einschätzung ihrer Fähigkeiten berichteten als die Mädchen. Allerdings erwies sich diese Differenz zum Beginn des 2. Halbjahres der 3. Klasse nicht mehr als signifikant. Kinder, die von ihrer Lehrerin zu Beginn von Klasse 2 dem oberen Leistungsdrittel ihrer Klasse zugeordnet worden waren, verfügten ab Mitte der 2. Klasse über ein deutlich positiveres Fähigkeitsselbstbild als die Kinder der beiden anderen Drittel. Die ungünstigsten Selbsteinschätzungen nahmen zunehmend die Kinder des unteren Leistungsdrittels vor ($F_{\text{Leistungsniveau}}(2) = 11.47$, $p < .001$, $F_{\text{Zeit}}(3) = 3.83$, $p < .01$, Wechselwirkung ‚Leistungsniveau‘/‚Zeit‘, $F(6) = 3.97$, $p = .001$.) Eine Wechselwirkung zwischen Leistungsniveau und Beurteilungsform konnte jedoch nicht nachgewiesen werden.

Hinsichtlich des relativen Fähigkeitsselbstbildes in Rechtschreiben zeigte sich, dass die Mädchen zum ersten Befragungstermin (Beginn 2. Halbjahr der 2. Klasse) zunächst über eine signifikant geringere Einschätzung ihrer Fähigkeiten berichten ($M_{\text{Mädchen}} = 2.18$, $SD = .92$ / $M_{\text{Jungen}} = 2.46$, $SD = .87$), $Z = -2.02$, $p < .05$. Zum zweiten Messzeitpunkt (Beginn der 3. Klasse) hin verbesserten sich die Einschätzungen beider Geschlechter. Die Mädchen schätzten sich nun, wenn auch nicht signifikant, positiver als die Jungen ein ($M_{\text{Mädchen}} = 2.78$, $SD =$

.71/ $M_{\text{Jungen}} = 2.64$, $SD = .68$). Zum Ende der 3. Klasse hin war wiederum bei beiden betrachteten Gruppen ein Absinken des Niveaus der Positiveinschätzung zu beobachten. Dabei vergrößerte sich die Differenz zwischen den Geschlechtern. Die Mädchen verfügten nach wie vor über ein positiveres relatives Fähigkeitsselbstbild in Rechtschreiben als die Jungen, allerdings nur in der Tendenz ($M_{\text{Mädchen}} = 2.65$, $SD = .64$ / $M_{\text{Jungen}} = 2.43$, $SD = .74$), $Z = -1.91$, $p = .056$. Wie beim relativen Fähigkeitsselbstkonzept in Rechnen erwies sich auch hinsichtlich Rechtschreiben die Zugehörigkeit der Kinder zu einem der drei Leistungsdritteln der Klasse als für die weitere Entwicklung bedeutsam: Ab dem 2. Halbjahr der 2. Klasse waren drei getrennte Entwicklungslinien zu beobachten, die ab dem Beginn der 3. Klasse durch Einzeltestungen auch als signifikant verschieden ausgewiesen wurden. Es zeigte sich wiederum, dass die Kinder, die durch die Lehrerin am Beginn von Klasse 2 dem oberen Klassendrittel zugeordnet worden waren, die günstigsten und die Kinder des unteren Leistungsdrittels die ungünstigsten Selbsteinschätzungen vornahmen ($F_{\text{Leistungsniveau}}(2) = 11.46$, $p < .001$, $F_{\text{Zeit}}(2) = 8.54$, $p < .001$.) Aber auch hier erreicht die Hinzunahme der ‚Beurteilungsform‘ als zusätzlicher Faktor in die Varianzanalyse mit Messwiederholung weder als Hauptfaktor noch in Wechselwirkung mit den anderen Faktoren das Signifikanzniveau.

Signifikante Effekte der Variablen ‚Schulstandort‘ zeigten sich in keiner der vorgenommenen Analysen.

Differentielle Entwicklungsverläufe der relativen Fähigkeitsselbstbilder.

Für die Abbildung differenzieller Verläufe der Entwicklung des relativen Fähigkeitsselbstbildes in Bezug auf Rechnen erschien eine Dreier-Gruppierung als geeignet (Abb. 2). 48% der Kinder (Cluster 2) sahen sich bei den Besten der Klasse, weitere 32% (Cluster 3) schätzten sich als ‚mittel gut‘ ein. Beide Cluster wiesen zwar leichte Schwankungen in den jeweiligen durchschnittlichen Einschätzungen auf, verliefen aber nahezu parallel. Davon wich der Verlauf von Cluster 1 dramatisch ab, $Z = -5.08$, $p < .001$. 20% der Kinder schätzten ihre Fähigkeiten am Beginn von Klasse 2 äußerst positiv ein, dies relativiert sich bereits zum Beginn des 2. Halbjahres der Klasse 2. Die nun gegebene Einschätzung stabilisierte sich zunächst bis zum Beginn von Klasse 3, um dann nochmals eine negative Entwicklung hin zu den ‚Mittelguten‘ zu nehmen.

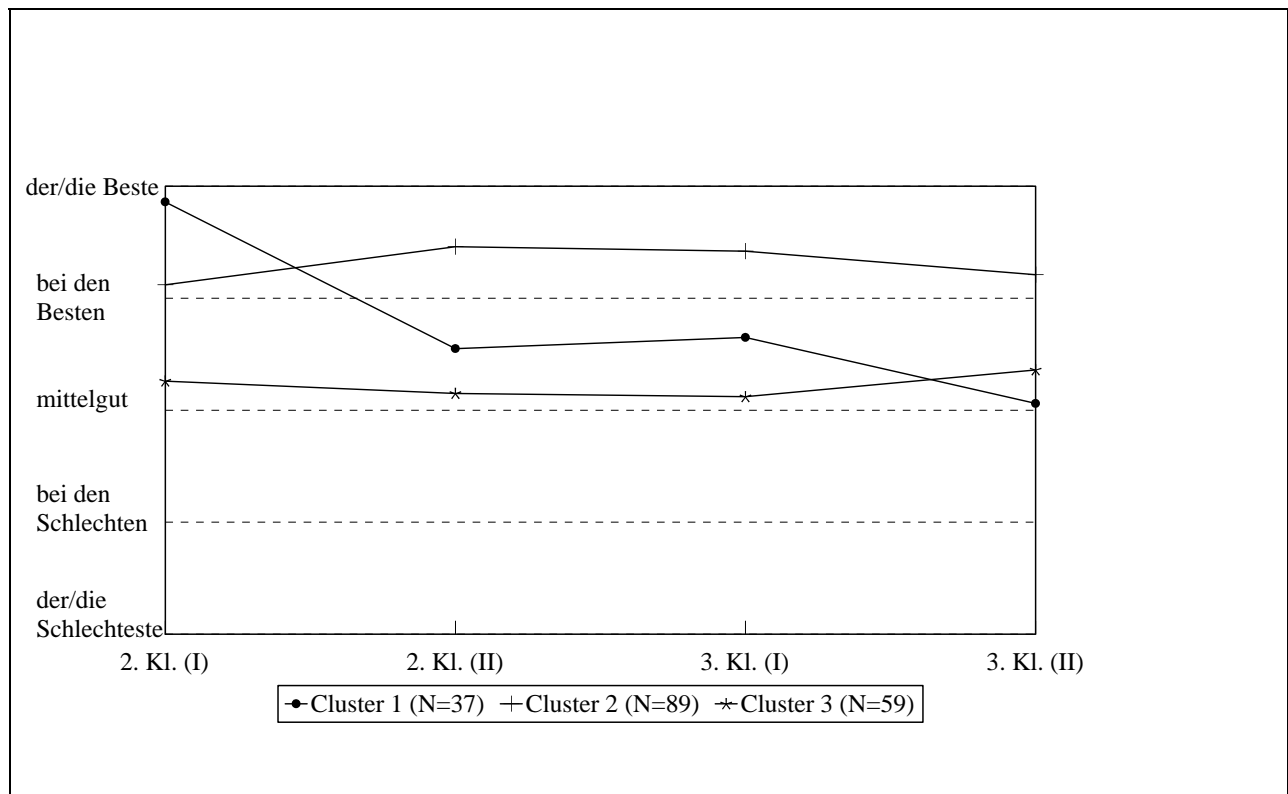


Abb. 2: Clusteranalyse von Entwicklungsverläufen des relativen Fähigkeitsselfkonzeptes in Rechnen (N = 185)

Vor allem die Einschätzung des Leistungsvermögens durch die Lehrkraft am Beginn von Klasse 2 erwies sich als ausschlaggebend für die Zugehörigkeit zu einem der Cluster. So erfolgte unabhängig von der Beurteilungsform eine eindeutige Zuordnung der Kinder aus dem oberen Klassendrittel zu Cluster 2 und die der aus dem unteren zu Cluster 1 und 3, $\chi^2(4, 185) = 30.87, p < .001$. Es ist zu beobachten, dass in Cluster 3 die Mädchen in der Tendenz überrepräsentiert sind, $\chi^2(2, 185) = 5.61, p = .06$.

Des Weiteren fand sich ein interessanter, aber nur tendenzieller Befund zu Cluster 1. Verbal beurteilte Kinder des oberen Klassendrittels fanden sich in diesem Cluster nicht, wohl aber in der Tendenz etwas häufiger mit Noten bewertete Jungen des unteren Klassendrittels. Wenn sich dieser Befund auf der Basis einer größeren Stichprobe nachweisen ließe, würde dies den unter der Bedingung der Benotung hypothetisch angenommenen Entwicklungsverlauf bestätigen.

Hinsichtlich der differentiellen Entwicklungsverläufe der relativen Fähigkeitsselfbilder in Rechtschreiben stellte sich eine Vier-Clusterlösung als sinnvolle Klassifizierung dar (Abb. 3). Für 34% der Kinder (Cluster 1) bestätigte sich die auf der Basis der Gesamtstichprobe beschriebene positive Entwicklung des relativen Fähigkeitskonzeptes nicht, sondern das in der Literatur beschriebene Absinken. Alle anderen Cluster wiesen positive Verläufe auf. 32% der Kinder (Cluster 2) sahen sich im Durchschnitt am Beginn des 2. Halbjahres von Klasse 2 zwischen den Besten und Mittelguten der Klasse. Am Beginn von Klasse 3 verbesserte sich diese

Einschätzung erheblich und veränderte sich nicht nennenswert zum Beginn des 2. Halbjahres von Klasse 3 hin. Cluster 3, der 24% der Kinder vereinte, verlief auf einem deutlich niedrigeren Niveau. Die Kinder sahen sich im Durchschnitt bei den Mittelguten. Ihre Einschätzung verbesserte sich zwar kontinuierlich ($Z = -3.05, p < .01$), blieb aber auch noch am Beginn des 2. Halbjahres von Klasse 3 deutlich unterhalb der Einschätzung, zu den Besten zu gehören. 10% der Kinder (Cluster 4) zeigten einen von den drei anderen Clustern deutlich abweichenden Verlauf. Am Beginn des 2. Halbjahres von Klasse 2 sahen sich diese Kinder bei den Schlechten, wenn nicht sogar als der/die Schlechteste. Zu Beginn von Klasse 3 schätzten sich dieselben Kinder nun plötzlich als Mittelgut bzw. zu den Besten gehörig ein. Diese Einschätzung erwies sich im Verlauf von Klasse 3 als stabil.

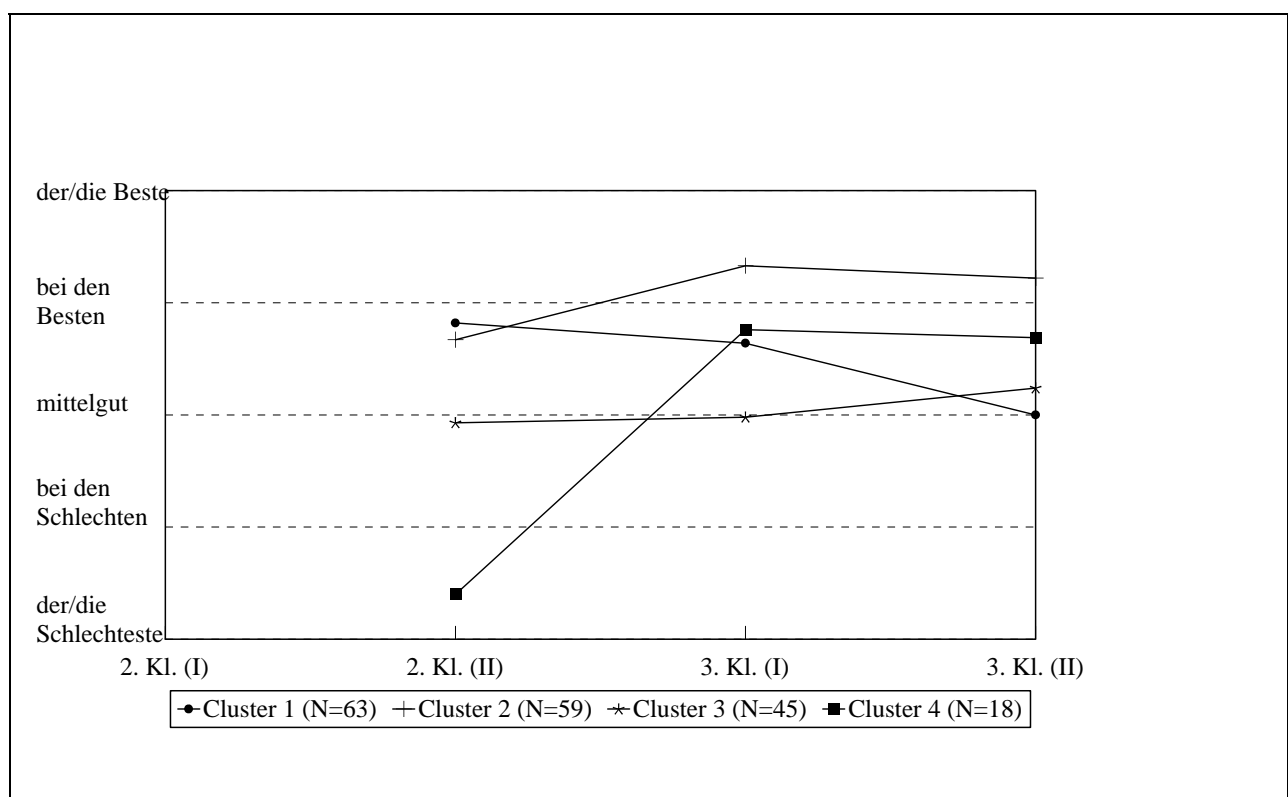


Abb. 3: Clusteranalyse von Entwicklungsverläufen des relativen Fähigkeitsselbstkonzeptes in Rechtschreiben (N = 185)

In gleicher Weise wie hinsichtlich des relativen Fähigkeitsselbstbildes in Rechnen fanden sich bei den Besetzungen der Cluster die Lehrerurteile wieder. So sammelten sich in Cluster 1 die Kinder des oberen Klassendrittels, in Cluster 2 die des mittleren und in Cluster 3 und 4 die des unteren, $\chi^2(2, 185) = 20.99, p < 0.01$. Dabei zeigte sich in Bezug auf einen Cluster (Cluster 3) ein Effekt der Beurteilungsform: Zu beobachten war wiederum ein tendenziell verstärktes Vorkommen von Kindern mit Notenbewertung des unteren Leistungsdrittels und darüber hinaus

ein tendenziell verstärktes Vorkommen von verbal beurteilter Kinder des oberen Klassendrittels, $\chi^2(2, 45) = 5.67, p < 0.058$.

Cluster 4 erfasste unabhängig von der Beurteilungsform in erster Linie Jungen des unteren Klassendrittels $\chi^2(2, 18) = 6.4, p < 0.05$. Eine Interpretation dieses Entwicklungsverlaufs ist ohne Berücksichtigung weiterer Variablen nicht möglich.

Diskussion

Die Analyse der Entwicklungsverläufe in Abhängigkeit von den Designvariablen zeigt, dass die Entwicklung der relativen Fähigkeitsselbstkonzepte den in der Literatur berichteten Geschlechtsstereotypen entspricht und ganz offensichtlich in engem Zusammenhang mit der am Beginn von Klasse 2 gegebenen Einschätzung des Leistungsvermögens des Kindes durch die unterrichtende Lehrkraft steht. So bestätigte sich, dass Mädchen günstigere Fähigkeitsselbsteinschätzungen in Rechtschreiben, Jungen dagegen in Rechnen aufweisen (Helmke, 1997, 1998, Valtin et al., 1996; Valtin/Wagner, 2002; Wagner/Valtin, 2003) sowie, dass leistungstärkere Kinder positivere Selbsteinschätzungen vornehmen als leistungsschwächere. (Valtin et al., 1996; Valtin/Wagner, 2002; Wagner/Valtin, 2003). Die Beurteilungsform hingegen beeinflusst bei keiner der betrachteten Variablen deren Entwicklung. Weder die ersten Erfahrungen mit Noten im Verlauf der 2. Klasse bei den ‚Noten-Kindern‘, noch der Übergang von der Verbalbeurteilung zur Benotung von Klasse 2 zu Klasse 3 führen zu sichtbaren Effekten auf den Verlauf der Entwicklung. Valtin/Wagner (2002) berichteten in Bezug auf die Gesamtstichprobe des Forschungsprojektes darüber, dass Notenkinder bis zur 4. Klasse ein höheres Selbstkonzept in Bezug auf Rechnen haben als Kinder mit verbaler Beurteilung und, ausgehend von einem zunächst ebenfalls höheren Selbstkonzept der Notenkinder in Rechtschreiben Mitte der 2. Klasse, sich erst im weiteren Verlauf die Fähigkeitsselbsteinschätzungen der mit Noten und der mit Verbalbeurteilungen beurteilten Kinder angleichen (Valtin/Wagner, 2002, S. 120). Die Analysen des Effektes der Beurteilungsform auf Basis einer ausbalancierten Teilstichprobe (Wagner/Valtin, 2003) bestätigen die höheren Selbstkonzepte der Notenkinder in Rechnen und weisen darüber hinaus von Mitte der 2. Klasse an vergleichbare Unterschiede hinsichtlich Rechtschreiben aus. Die eigenen Analysen basieren ebenfalls auf einer Teilstichprobe. Hinsichtlich dieser Teilstichprobe zeigte sich weder für Rechnen noch für Rechtschreiben zu einem der Befragungstermine ein signifikanter Effekt der Beurteilungsform auf die Fähigkeitsselbsteinschätzungen. Die voneinander offensichtlich abweichenden Effekte der Variablen ‚Beurteilungsform‘ könnten dahingehend interpretiert werden, dass sich die Analyseergebnisse in Abhängigkeit der

Zusammensetzung der Stichprobe verändern. Dies würde die Kritik Asendorpfs (1995) an der Verallgemeinerbarkeit variablenzentrierter Analyseergebnissen stützen, ebenso seine Überlegung, dass möglicherweise bestimmte Kombinationen von Variablen zu spezifischen Effekten führen, u. U. eigene Äquivalenzklassen definieren, die dann in Abhängigkeit von der Anzahl ihrer Mitglieder das Analyseergebnis beeinflussen können.

Neben den unterschiedlichen Stichprobenumfängen weisen die drei Stichproben wesentliche Unterschiede in der Zusammensetzung bezüglich der betrachteten Designvariablen auf. In der hinsichtlich Geschlecht und Schulstandort ausgewogenen Gesamtstichprobe² überwiegen in Klassenstufe 2 die Kinder mit Verbalbeurteilungen, ab Klasse 3 jedoch deutlich die mit Noten bewerteten. Dagegen können sowohl die ausbalancierte Stichprobe (Wagner/Valtin, 2003) als auch die hier im Mittelpunkt stehende Teilstichprobe hinsichtlich der Variable ‚Beurteilungsform‘³ und ‚Geschlecht‘ als gleichverteilt angesehen werden. Bei der Berücksichtigung unterschiedlicher Leistungsniveaus besteht allerdings zwischen diesen beiden Teilstichproben ein ganz wesentlicher Unterschied: die ausbalancierte Stichprobe (Wagner/Valtin, 2003) wurde über die Werte in den Schulleistungstest der 2., 3. und 4. Klasse nach der Ausbalancierung in zwei Leistungsgruppen aufgeteilt (vgl. Wagner/Valtin, 2003, S. 29). In der hier betrachteten Teilstichprobe basierten die Leistungsgruppen dagegen, wie bereits dargestellt, auf den Einschätzungen der Lehrkräfte am Beginn von Klasse 2. Die drei daraus resultierenden Leistungsgruppen gingen von vornherein in die Verteilungsprüfungen der Teilstichprobe ein. Die Studien Jerusalems und Weinstein's belegen, dass die Fähigkeitseinschätzungen der Lehrkräfte die Fähigkeitsselbstbilder der Kinder im Verlauf der ersten Schuljahre zunehmend prägen (vgl. Jerusalem, 1984; Weinstein et al., 1987) bzw., wie Fend und Stöckli es formulieren, davon auszugehen ist, dass sich die Selbsteinschätzungen der Kinder den Leistungsbeurteilungen der Lehrkräfte zunehmend annähern (1997). Vor diesem Hintergrund erscheint es, neben der auf objektiven Daten basierenden Herangehensweise Wagners/Valtins (2003), durchaus sinnvoll, die Leistungsgruppen über die Lehrerurteile zu definieren, wenn man Effekte der Beurteilungspraxis analysieren möchte. Zwischen den am Beginn von Klasse 2 erhobenen Einschätzungen des Leistungsvermögens des Kindes durch die unterrichtende Lehrkraft und der im weiteren beobachteten Entwicklung der kindlichen Fähigkeitsselbstkonzepte besteht ein nachgewiesener enger Zusammenhang (Valtin et al., 1996, Schmude, 2001). Dies spricht dafür, dass die Kinder auch ohne Benotung einzuschätzen lernen, wie ihre Leistungsfähigkeit von der Lehrkraft wahrgenommen wird (vgl. Valtin et al., 1996, S.

² Auf eine annähernde Gleichverteilung in Bezug auf das Leistungsniveaus wurde mit Hilfe von Einschätzungen durch die unterrichtenden Lehrkräfte bei der Rekrutierung der Stichprobe geachtet.

³ unter Berücksichtigung des Schulstandortes

237). Inwieweit es sich dabei um eine objektive Widerspiegelung der tatsächlichen Leistungsfähigkeit handelt, sei dahin gestellt. Gestützt durch die Befunde zur geschlechtsspezifischen bzw. eigentlich schon geschlechtsstereotypen Entwicklung der Fähigkeitsselbstkonzepte kann vermutet werden, dass in dem betrachteten Zeitraum die Entwicklung der relativen Fähigkeitsselbstkonzepte nicht durch die Form der Leistungsbeurteilung beeinflusst wird, sondern primär durch die mit den Leistungsbeurteilungen transportierten, auf impliziten Persönlichkeitstheorien basierenden Vorstellungs- und Bewertungsmuster der Lehrenden.

Die durchgeführten Clusteranalysen belegen die Existenz deutlich von den Gruppenmittelwerten abweichender Entwicklungsverläufe. So zeigt sich hinsichtlich der Entwicklung des Fähigkeitsselbstkonzeptes in Rechtschreiben, dass sich für ein Drittel der Kinder die auf der Ebene der Gesamtstichprobe beschriebene positive Entwicklung nicht bestätigt, sondern die in der Literatur beschriebene, zunehmend negative Entwicklung. Dies stützt die Forderung Asendorpfs (1995), die intraindividuelle Gültigkeit genereller Entwicklungsverläufe zu prüfen. Die herausgearbeiteten Cluster lassen sich dabei nur teilweise durch die Designvariablen erklären. Die in der Literatur beschriebenen geschlechtsspezifischen Entwicklungsverläufe in Kombination mit der fachlichen Anforderung (Rechnen bzw. Rechtschreiben) spiegeln sich zwar deutlich in den Clustern wider, können diese aber nicht vollständig erklären. Dies bedeutet, dass sich die per Designvariablen definierten Vergleichsgruppen nicht vollständig reproduzieren lassen. Auch dies stützt empirisch die Diskussion der Problematik variablenzentrierter Analyseansätze und die Frage, welche Merkmale der einzelnen Person, neben der Designvariablen, noch in die Bildung der Äquivalenzklassen eingehen (Asendorpfs, 1995). Als bedeutsamer, allerdings ebenfalls nicht erschöpfender Erklärungsfaktor erwies sich die von den Lehrkräften vorgenommene Einschätzung des Leistungsniveaus zu Beginn von Klasse 2. Damit bestätigt sich auch in der Analyse der differenziellen Entwicklungsverläufe, dass die Fähigkeitseinschätzungen der Lehrkräfte die Fähigkeitsselbstbilder der Kinder im Verlauf der ersten Schuljahre zunehmend prägen (vgl. Fend/Stöckli, 1997; Jerusalem, 1984; Weinstein et al., 1987).

Bemerkenswert ist, dass sich mittels Clusteranalysen, wenn auch nur tendenziell, Effekte der Beurteilungsform in Kombination mit dem Leistungsvermögen zeigen. Dieser kombinierte Effekt mittelt sich vermutlich bei den varianzanalytischen Berechnungen heraus, da es sich dabei um einen gegenläufigen Effekt handelt: (1) verbal beurteilte Kinder des oberen Klassendrittels ebenso wie mit Noten bewertete Kinder des unteren Klassendrittels nehmen hinsichtlich ihrer

Kompetenzen in Rechtschreiben eher eine ungünstigere Einschätzung ihrer Fähigkeiten vor, (2) mit Noten bewertete Kinder des unteren Klassendrittels geben in Bezug auf ihre Kompetenzen in Rechnen sogar eine sich deutlich verschlechternde Einschätzung ab.

Vor dem Hintergrund der Befunde der qualitativen Inhaltsanalyse der Berichtszeugnisse hinsichtlich der nur in Ansätzen intentionsgerechten Realisierung (Schmude, 2001) und den aus dem Projekt NOVUS (Wagener, 2003) berichteten Befund, dass sich das Rückmeldeverhalten der Lehrkräfte als von der Beurteilungsform unabhängig erweist, reduziert sich der hier beschriebene Effekt der Beurteilungsform auf ‚mit oder ohne Noten‘.

Es ist anzunehmen, dass ohne Noten und ohne differenzierte, individuelle Rückmeldungen in verbaler Form leistungsstarken Kindern eine Orientierung hinsichtlich der eigenen Kompetenzen in dieser für die Selbstkonzeptentwicklung sensiblen Phase fehlt. Dies kann, wie beobachtet, zu einer Unterschätzung der eigenen Kompetenz führen. Das Erleben des selbstständigen Bewältigenkönnens der schulischen Anforderungen allein reicht für die Entwicklung angemessener Fähigkeitselbsteinschätzungen offensichtlich nicht aus. Daraus lässt sich schlussfolgern, dass auch für leistungsstarke Kinder die intentionsgerechte Umsetzung der ermutigenden und förderdiagnostischen Funktion verbaler Leistungsbeurteilungen als Orientierungshilfe unentbehrlich ist. Hinsichtlich der mit Noten bewerteten Kinder des unteren Leistungsdrittels deutet sich hier eine Bestätigung der in der Literatur beschriebenen negativen Effekte der Benotung an: Sie erfahren frühzeitig, dass sie viele von den an sie gestellten Anforderungen nicht adäquat bewältigen können. Ihnen wird bereits am Beginn ihrer Bildungskarriere über die Noten ihre persönliche Position in der leistungsabhängigen Rangreihenbildung innerhalb der Klasse zugewiesen. Dabei dürfte sich dies vor allem ungünstig auf die Entwicklung der Fähigkeitsselbstbilder der Jungen auswirken, auch wenn sich dies in den vorgestellten Befunden nicht in dieser Deutlichkeit zeigt. Entwisle et al. (1987) weisen auf frühzeitige Geschlechtsunterschiede im Aufbau des schulischen Selbstbildes hin. Im Gegensatz zu den Mädchen orientieren sich die Selbsteinschätzungen der Jungen in den ersten Schuljahren eher autonom an „objektiven“ Daten, nämlich den Schulleistungen. Über diese geben, scheinbar objektiv, die Noten Auskunft und dürften damit bereits sehr früh die weitere Entwicklung der Fähigkeitsselbstbilder prägen.

Auch mit dieser hier vorgestellten Studie ließ sich die Frage, ob sich intentionsgerecht erstellte verbale Beurteilungen günstig auf die Entwicklung schulrelevanter Persönlichkeitsmerkmale auswirken, nicht beantworten. Die Frage, ob und wie die Variationen schulischer

Rahmenbedingungen wirken, lässt sich offensichtlich nur teilweise über Mittelwertsvergleiche von durch unabhängige Variablen konstituierten Vergleichsgruppen beantworten. Die vorgetragenen Befunde bestätigen, dass im Kontext pädagogisch-entwicklungspsychologischer Forschung Abweichungen von den Mittelwerten individuelle Besonderheiten beschreiben, die notwendigerweise in die Erklärung von Entwicklungsverläufen mit einbezogen werden müssen. Neben diesem forschungsmethodischen Fazit belegen beide hier vorgestellten Analyseansätze die Bedeutsamkeit der impliziten Persönlichkeitstheorien derjenigen, die beurteilen. Die sich unabhängig von der Beurteilungsform zeigenden geschlechtsspezifischen Effekte und die Nachhaltigkeit der einmalig erhobenen Fähigkeitseinschätzungen durch die Lehrenden machen deutlich, dass es für eine Reform der Leistungsbeurteilung nicht ausreicht, Noten einfach durch Wörter zu ersetzen.

Literatur

- Asendorpf, J. (1995). Persönlichkeitspsychologie: Das empirische Studium der individuellen Besonderheit aus spezieller und differentieller Perspektive. *Psychologische Rundschau*, 46, 235-247.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. (4. vollständig überarbeitete Aufl.). Berlin: Springer-Verlag.
- Entwisle, D. R., Alexander, K. L., Pallas, A. M. & Cadigan, D. (1987). The emergent academic self-image of first graders: Ist response to social structure. *Child Development*, 58. No. 5, 1190 – 1206.
- Fend, H. & Stöckli, G. (1997). Der Einfluss des Bildungssystems auf die Humanentwicklung: Entwicklungspsychologie der Schulzeit. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule. Enzyklopädie der Psychologie. Serie Pädagogische Psychologie* (Bd. 3), 1-35). Göttingen: Hogrefe.
- Gartner, F. & Jirasko, M. (1999). „Ehrgeizige“ Eltern und schlechte Schulleistungen. Fördern überhöhte elterliche Ausbildungswünsche die kindliche Prüfungsangst? *Report Psychologie*, 4, 272-279.
- Helmke, A. (1992). Selbstvertrauen und schulische Leistungen. Göttingen: Hogrefe.
- Helmke, A. (1996). Statement zum Projektbericht: NOVARA. Leistungsbeurteilung und psychosoziale Entwicklung im Grundschulalter. In D. Benner, H. Merckens, & Th. Gatzemann, (Hrsg.), *Pädagogische Eigenlogiken im Transformationsprozeß von SBZ, DDR und neuen Ländern - Neue Ergebnisse aus der an der Freien Universität und der Humboldt-Universität zu Berlin eingerichteten DFG-Forschergruppe*.
- Helmke, A. (1997). Entwicklung lern- und leistungsbezogener Motive und Einstellungen: Ergebnisse aus dem SCHOLASTIK-Projekt. In F.E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter*. Weinheim: Beltz.
- Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzeptes vom Kindergarten bis zur 6. Klassenstufe. In Weinert, F.E., *Entwicklung im Kindesalter*. Weinheim, München: Psychologie VerlagsUnion.
- Jerusalem, M. (1984). *Selbstbezogene Kognitionen in schulischen Bezugsgruppen. Eine Längsschnittstudie*. (Bd. I). Institut für Psychologie der Freien Universität, Berlin.
- Krapp, A. (1997). Selbstkonzept und Leistung – Dynamik ihres Zusammenspiels: Literaturüberblick. In F.E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter*. Weinheim: Beltz.

- Lissmann, U. (1981). Zur Wirkung verschiedener Rückmeldetechniken auf Lernende. Ein Literaturbericht. In K. Ingenkamp (Hrsg.), *Wert und Wirkung von Beurteilungsverfahren*. Weinheim: Beltz.
- Lissmann, U. (1984). *Leistungsrückmeldungen in der Schule: Formen und Wirkungen*. Schulleitung 4. Grundwerk vom August 1984.
- Nicholls, J. G. (1978). The development of the concepts of effort and ability. Perception of academic attainment, and the understanding that difficult tasks require more ability. *Child development*, 49, 800 – 814.
- Pekrun, R. (1996). Ziffernzensuren oder Berichtszeugnisse? Drei kritische Anmerkungen zur Annahme unterschiedlicher Wirkungen. In D. Benner, H. Merckens & F. Schmidt (Hrsg.), *Bildung und Schule im Transformationsprozeß von SBZ, DDR und neuen Ländern - Untersuchungen zu Kontinuität und Wandel. Erste Ergebnisse aus der an der Freien Universität und der Humboldt-Universität zu Berlin eingerichteten DFG-Forschergruppe*. Berlin: Zentrale Universitäts-Druckerei. S. 253 – 259.
- Rosenfeld, H. & Valtin, R. (1997). Zur Entwicklung schulbezogener Persönlichkeitsmerkmale bei Kindern im Grundschulalter. Erste Ergebnisse aus dem Projekt NOVARA. *Unterrichtswissenschaften*, 25, 316-330.
- Schmude, C. (2001). „Berichtszeugnisse – unnötiger Aufwand oder aufwendige Notenwendigkeit?“. Dissertation, Humboldt-Universität.
- Schwarzer, R. & Roysl, W. (1984). Angst und Schulunlust als Sozialisationseffekte verschiedener Schularten. *Zeitschrift für Pädagogik*, 30(6).
- Valtin, R. & Wagner, C. (2002). Wie wirken sich Notengebung und verbale Beurteilung auf die leistungsbezogene Persönlichkeitsentwicklung aus?. In R. Valtin, C. Schmude u. a. *Was ist ein gutes Zeugnis? Noten und verbale Beurteilungen auf dem Prüfstand*. München: Juventa Verlag.
- Valtin, R. (2000). *Schulische Adaptation und Bildungsaspiration (SABA). Schlussbericht an die Deutsche Forschungsgemeinschaft zum Abschluss der Arbeit der Forschergruppe „Bildung und Schule im Transformationsprozess von SBZ, DDR und neuen Ländern – Untersuchungen zu Kontinuität und Wandel“. Förderzeitraum 01.06.1994 – 30.05. 2000*. Humboldt-Universität. Institut für Schulpädagogik und Pädagogische Psychologie. Abteilung Grundschulpädagogik. Berlin.
- Valtin, R. (2000a). Ein Entwicklungsmodell des Rechtschreibenlernens. In R. Valtin (Hrsg.), *Rechtschreiben lernen in den Klassen 1 – 6, Grundlagen und didaktische Hilfen*. Frankfurt/Main: Grundschulverband – Arbeitskreis Grundschule e. V.
- Valtin, R., Schmude, C. u.a. (2002). *Was ist ein gutes Zeugnis? Noten und verbale Beurteilungen auf dem Prüfstand*. München: Juventa Verlag.
- Valtin, R., Würscher, I., Rosenfeld, H., Schmude, C. & Wisser, Ch. (1996). Der Beitrag der Schule zur Persönlichkeitsentwicklung. Erste Ergebnisse einer Längsschnittstudie. In D. Benner, H. Merckens & F. Schmidt (Hrsg.), *Bildung und Schule im Transformationsprozeß von SBZ, DDR und neuen Ländern - Untersuchungen zu Kontinuität und Wandel. Erste Ergebnisse aus der an der Freien Universität und der Humboldt-Universität zu Berlin eingerichteten DFG-Forschergruppe*. Berlin: Zentrale Universitäts-Druckerei.
- Wagener, M (2003). *Ziffernzensuren oder verbale Beurteilung. Eine empirische Untersuchung zum Zusammenhang von Beurteilungsform und Unterrichtsorganisation, Leistungsrückmeldung und Klassenraumgestaltung von Grundschullehrerinnen aus östlichen und westlichen Bezirken Berlins*. Weinheim: Beltz Verlag.
- Wagner, C. & Valtin R. (2003). Noten oder Verbalbeurteilung? Die Wirkung unterschiedlicher Bewertungsformen auf die schulische Entwicklung von Grundschulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*. 35 (1). S. 27 – 36.
- Weinert, F.E. & Helmke, A. (Hrsg.) (1997) *Entwicklung im Grundschulalter*. Weinheim:

Beltz.

Weinstein, R. S., Marshall, H. H. , Sharp, L. & Botkin, M. (1987). Pygmalion and the student: Age and classroom, differences in children's awareness of teacher expectations. *Child Development*, 58. S. 1079 – 1093.