

## **Rechtschreibdiagnose in der Grundschule – Theoretische und empirische Analysen zur spezifischen Problematik**

Wolfgang Schneider

**Zusammenfassung:** Theoretische Analysen zum Problem des Rechtschreibens weisen darauf hin, daß weniger Intelligenzmerkmale als vielmehr Gedächtnisleistungen bei dem Erwerb der Schriftsprache von Bedeutung sind. Daraus folgt, daß für die Prüfwörter in normierten Rechtschreibtests zumindest hinreichende Vorkommensfrequenz gewährleistet sein sollte. Diese Frage wird in der vorliegenden Untersuchung am Beispiel des Allgemeinen Schulleistungstests überprüft: für die ausgewählten Klassenstufen kann gezeigt werden, daß die Vorkommenshäufigkeit der Prüfwörter nicht ausreicht und damit wenig geeignet ist, um die tatsächliche Rechtschreibfertigkeit zu erfassen. Der Vergleich mit mehreren eng am Curriculum orientierten Diktatproben kann gleichzeitig die Schwierigkeiten verdeutlichen, die dann entstehen, wenn zuverlässige Bestimmungen der individuellen Rechtschreibkompetenz vorgenommen werden sollen.

### **Specific difficulties in assessing spelling competencies in elementary schools**

**Summary:** While there seems to be a long tradition in educational literature postulating a strict connection between spelling ability and test intelligence, recent theoretical analyses of the spelling process have emphasized the importance of basic memory skills. As a consequence word frequency should be controlled in order to test individual spelling competence.

For a specific norm-oriented spelling test the empirical analysis can show that the importance of sufficient word frequency has been underestimated. The comparison with four other criterion-oriented spelling tests proves that the norm-oriented procedure cannot assess the true spelling competence; but it shows, too, the general difficulty to come to a reliable estimation of individual spelling skills.

### **1. Einleitung**

Wählt man die reine Instruktionsquantität als Bezugsmaßstab, so nimmt der Orthographieunterricht im Grundschulcurriculum nach wie vor zweifellos eine herausragende Stellung ein. Mit dieser Gewichtung geht vielfach die Grundüberzeugung einher, daß die Qualität der Rechtschreibleistung als Indikator für die generelle kognitive Leistungsfähigkeit aufzufassen ist.

Empirische Belege für die Durchschlagskraft solcher Annahmen fanden sich beispielsweise in der Studie von Kemmler (1967), die Unterschiede zwischen den Extremgruppen erfolgreicher Schüler und von Schulversagern aufdeckte. So wie innerhalb der Stichprobe schlechter Schüler weitgehend der

Ausprägungsgrad des Rechtschreibdefizits darüber entschied, ob man zu den Sitzenbleibern gehörte, war auch bei den guten Schülern im wesentlichen die Qualität der Rechtschreibfertigkeit für das Fortkommen in der Grundschule relevant. Fragt man nun nach den empirischen Belegen für den hier implizit unterstellten Zusammenhang zwischen Rechtschreibleistung und Intelligenz, so fallen die Befunde wenig überzeugend aus.

In den Untersuchungen von Glogauer (1977) sowie Pfeiffer und Zielinski (1975) ergaben sich relativ übereinstimmend lediglich mittelhohe Beziehungen zwischen beiden Variablen<sup>1</sup>. Da die Intelligenz demnach maximal 20–30% der Rechtschreibvarianz aufklärte, machten Pfeiffer und Zielinski nachdrücklich auf die Problematik aufmerksam, von Rechtschreibleistungen auf Intelligenzmerkmale schließen zu wollen.

Diese Befunde scheinen insbesondere dann plausibel, wenn man sich – wie bisher nur selten geschehen – auf die Charakteristika der Schriftsprache und ihre spezifischen Anforderungen an Lerner-Fertigkeiten konzentriert und damit aufgabenanalytische Aspekte stärker in den Vordergrund stellt.

Schon die Beschäftigung mit der Geschichte der deutschen Orthographie führt dem Betrachter die Reihe von Zufälligkeiten vor Augen, durch die aus einem ehemals vorwiegend phonetischen Schriftsystem über den Einfluß von Dialekten, Kanzleipflogenheiten und zunehmenden Willkürlichkeiten im Bereich der Doppel- und Großschreibung die auch heute noch weitgehend gültigen Normierungsbestimmungen der Orthographischen Konferenz von 1901 entstehen konnten. Es darf aufgrund dieser Entwicklung kaum verwundern, wenn intelligente Leistungen wie das konsequente und bewußte Anwenden von Regeln nur mäßigen Erfolg bei der Rechtschreibung zeigen, da „nicht die Gleich-, sondern die Andersschreibung den Grundzug der Rechtschreibung ausmacht“ (Kern und Kern, 1954, S. 11).

Weitaus mehr Beachtung haben in jüngerer Zeit deshalb Untersuchungsergebnisse gefunden, die speziell den Gedächtnisleistungen der Schüler einen höheren Stellenwert zuschrieben. Aus Studien zur verbalen Informationsverarbeitung (z. B. Schneider, 1977, Schneider et al., 1977, Tarver et al., 1976, Torgesen und Goldman, 1977) ging hervor, daß hauptsächlich die Verfügbarkeit von effizienten Enkodier- bzw. Speicherstrategien zwischen normalen und schwachen Rechtschreibern differenzierte. Die Relevanz von Gedächtnisprozessen ließ sich auch bei der Überprüfung sog. Funktionsmodelle der Rechtschreibung (s. Simon und Simon, 1973, Simon, 1976, Zur Oeveste, 1977) eindrucksvoll nachweisen, wobei sowohl die Speicherung von einzelnen Phonem-Graphem-Assoziationen als auch die Erfassung von Wortbildern besondere Bedeutung gewann.

Aus diesen Ergebnissen geht also hervor, daß für den Erwerb von Orthographiekenntnissen kognitive Fähigkeiten wie Regelerkennen bzw. -anwenden oder schlußfolgerndes Denken von untergeordneter, elementare Gedächtnisprozesse dagegen von spezifischer Bedeutung sind. Damit wird gleichzeitig der Aspekt des Unterrichtsangebots relevant: es sollte bei der Evaluation der individuellen Rechtschreibfertigkeit in jedem Fall gesichert sein, daß die Schüler im Unterricht zumindest ausreichend Gelegenheit hatten, die zur

Leistungsfeststellung verwendeten Testwörter kennenzulernen, da andernfalls von „Testfairness“ kaum die Rede sein kann.

Die Überprüfung dieses scheinbar simplen Tatbestandes steht im Vordergrund der im folgenden vorgestellten Untersuchung, in der die Schwierigkeiten einer adäquaten Beurteilung der Rechtschreibkompetenz aufgezeigt werden sollen. Es wird dabei zunächst einmal der Frage nachgegangen, inwieweit bei standardisierten Rechtschreibtests die Forderung nach hinreichender Vorkommenshäufigkeit des zugrundegelegten Wortmaterials als erfüllt angesehen werden kann. Da demnach das Problem der curricularen Validität besonders interessiert, sollen Resultate aus normorientierten Prüftexten mit Ergebnissen aus solchen unstandardisierten Diktaten verglichen werden, die sowohl speziell am Lehrplan orientiert als auch in ihrer Zusammenstellung im Hinblick auf Vorkommenshäufigkeit und (theoretische) Wortschwierigkeit systematisch variiert worden sind.

Da die meisten Studien zur Rechtschreibproblematik bisher vorwiegend an eher fortgeschrittenen Klassen der Grund- und Hauptschule durchgeführt wurden und es fraglich ist, ob die dort gewonnenen Befunde auch für beginnende Rechtschreiber Geltung besitzen, schien es sinnvoll, die gewählte Fragestellung zusätzlich an jüngeren Kindern zu untersuchen.

## 2. Methode

### 2.1. Selektion der Stichprobe und Gewinnung der Testinstrumente

Zur Überprüfung dieser Annahmen wurde auf Daten zurückgegriffen, die im Rahmen einer früheren Untersuchung (Schneider, 1980) erhoben worden waren. Es lagen Ergebnisse aus den Rechtschreib-Subtests der Allgemeinen Schulleistungstests (AST2 von Rieder bzw. AST4 von Fippinger) für neun 2. Klassen (N = 226) und zehn 4. Klassen (N = 280) aus Schulen des Rhein-Neckar-Kreises vor. Praktische Relevanz kommt dem AST vor allem wegen seiner häufigen Verwendung bei der Übergangsauslese für weiterführende Schulen zu, woran auch kritische Anmerkungen zu seiner begrenzten Durchführungsobjektivität (Fingerhut und Langfeldt, 1973) sowie seiner eingeschränkten Stabilität und faktoriellen Gültigkeit (Masendorf und Tücke, 1975) bisher kaum etwas geändert haben. Für die vorliegende Untersuchung war insbesondere der von den Testautoren im Testmanual geäußerte Anspruch von Interesse, nach dem für alle Subtests des Verfahrens Lehrplangültigkeit (curriculare Validität) bestehen sollte.

Zur Kontrolle der Worthäufigkeit bzw. -schwierigkeit waren eigene Instrumente entwickelt worden. Da nicht davon ausgegangen werden konnte, daß der Grundschulwortschatz in den verbreiteten Frequenzwörterbüchern (Kæding, 1898 bzw. Meier, 1964) adäquat abgebildet sein würde (vgl. hierzu Ortman, 1975), wurde über das Textverarbeitungsprogramm COBAPH (vgl. Hitzenberger et al., 1977) ein spezielles Frequenzwörterbuch erstellt, das die

in den untersuchten Klassen am häufigsten verwendeten Unterrichtsmaterialien berücksichtigte<sup>2</sup>.

Zur Operationalisierung der Wortschwierigkeit diente ein Computerprogramm zur Laut-Schrift-Transkription. Nachdem etwas mehr als 200 Laut-Schrift-Zuordnungsregeln gespeichert worden waren, generierte der Computer als perfekter Regelanwender für jede vorgegebene Lautsequenz diejenigen Schreibweisen, die aufgrund seines Korrespondenzregelverzeichnisses als legitim einzustufen waren. Als Schwierigkeitsindex wurde nun die Anzahl der unterschiedlichen Schreibungen („Fehler“) gewählt, die sich für ein vorgegebenes Wort aus dem Regelwissen korrekt ableiten ließen<sup>3</sup>.

Die für den Vergleich mit den standardisierten Prüftexten vorgesehenen Diktate wurden über die systematische Kombination von Worthäufigkeit und -schwierigkeit konstruiert, wobei lediglich solche Items berücksichtigt wurden, die nach den Modellannahmen von Mokken bzw. Rasch zu (meßtheoretisch) eindimensionalen Skalen zusammengefaßt werden konnten (vgl. Schneider, 1980). Es entstanden somit vier verschiedene Diktatproben mit häufig/leichtem (HL), häufig/schwierigem (HS), selten/leichtem (SL) und selten/schwierigem (SS) Wortmaterial.

### 3. Ergebnisse

#### 3.1. Deskriptiv- und Korrelationsbefunde

**Tabelle 1:** Mittelwerte (M) und Standardabweichungen (s) der einzelnen Diktatproben, nach Klassenstufe und Geschlecht aufgeschlüsselt (Fallzahlen in Klammern)

##### a) 2. Klassen

	AST		HL		SS		SL		HS		N
	M	s	M	s	M	s	M	s	M	s	
Jung.	45.4	8.7	11.8	6.4	5.0	3.7	6.7	2.5	5.3	4.5	(126)
Mäd.	46.5	7.3	12.6	6.2	5.3	3.8	7.0	2.3	5.6	4.3	(100)

##### b) 4. Klassen

	AST		HL		SS		SL		HS		N
	M	s	M	s	M	s	M	s	M	s	
Jung.	47.6	10.5	14.4	5.0	10.9	4.6	9.5	4.0	11.4	4.9	(131)
Mäd.	49.8	9.9	14.6	5.2	11.4	4.8	13.0	4.4	13.2	5.7	(142)

##### Legende:

AST = Allgemeiner Schulleistungstest

HL = Häufige und leichte Wörter

SS = Seltene und leichte Wörter

SL = Seltene und leichte Wörter

HS = Häufige und schwierige Wörter

(die AST-Befunde sind in T-Werten wiedergegeben; bei den übrigen Texten ist die Anzahl richtig geschriebener Zielwörter wiedergegeben)

In Tabelle 1 sind für die fünf verschiedenen Rechtschreibproben Mittelwerte und Standardabweichungen nach Schulklassen- und Geschlechtszugehörigkeit getrennt wiedergegeben.

Zum AST-Rechtschreibtest liegen für beide Klassenstufen T-Werte vor, die z. T. erheblich von den in den Testnorm-Tabellen enthaltenen Durchschnittswerten abweichen. Man kann dieses Ergebnis dadurch verdeutlichen, daß man die Befunde in Prozentränge umrechnet: es ergibt sich bei den Zweitkläßlern ein mittlerer Prozentrang von 33 (Jungen) bzw. 35 (Mädchen), während für die Viertkläßler Prozentränge von 38 (Jungen) bzw. 50 (Mädchen) resultieren. Der AST-Rechtschreibtest scheint damit zumindest bei den Zweitkläßlern etwas zu schwierig zu sein. Für die vier systematisch nach Häufigkeit und Schwierigkeit zusammengestellten Diktatkomponenten ist in Tabelle 1 die mittlere Anzahl der korrekt geschriebenen Zielwörter wiedergegeben<sup>4</sup>. Vergleiche sind hier lediglich pro Spalte (also für die Geschlechter) möglich, da sich die Anzahl der Zielwörter in den einzelnen Textvarianten unterscheidet. Leistungsdiskrepanzen zwischen den Geschlechtern sind lediglich für die Viertkläßler feststellbar, während die Jungen und Mädchen der 2. Klassen in etwa vergleichbar abschneiden (vgl. dazu auch die Befunde bei Schneider, 1980). Dies gilt auch (und zwar konträr zu den Ergebnissen für die Eichstichproben) für die Differenzen zwischen den AST-Mittelwerten, die bei den Zweitkläßlern statistisch unbedeutsam ausfallen ( $t = -1.00$ ,  $p = .05$ ), bei den Viertkläßlern dagegen das erforderliche Signifikanzniveau erreichen ( $t = -1.77$ ,  $p = .05$ ).

**Tabelle 2:** Produkt-Mengen-Korrelationen zwischen den einzelnen Rechtschreib-Maßen, nach Klassenstufe und Geschlecht aufgeschlüsselt

		AST		HL		SS		SL	
		J	M	J	M	J	M	J	M
HL	2. Kl.	.26	.06						
	4. Kl.	.29	.13						
SS	2. Kl.	.40	.31	.80	.77				
	4. Kl.	.54	.37	.87	.88				
SL	2. Kl.	.40	.33	.37	.32	.39	.38		
	4. Kl.	.43	.50	.68	.69	.83	.84		
HS	2. Kl.	.30	.34	.11	.04	.10	.10	.85	.80
	4. Kl.	.50	.30	.37	.39	.51	.56	.02	.77

Tabelle 2 verdeutlicht die korrelativen Zusammenhänge für die einzelnen Rechtschreibproben, die wiederum nach Geschlecht und Klassenstufe getrennt wiedergegeben sind. Im Hinblick auf die hier interessierende Fragestellung sind vor allem die Beziehungen zwischen den AST-Werten und den Diktat-

varianten von Interesse. Es zeigt sich, daß die Korrelationen insgesamt allenfalls mittelhoch ausfallen: das normierte Verfahren weist in beiden Altersstufen den schwächsten Zusammenhang mit den Ergebnissen zu den häufig-leichten Wortmaterialien auf, wobei er für die Mädchen konsistent am niedrigsten liegt. Die höchsten Korrelationen ergeben sich bei den Viertkläßlern zu selten-schwierigen (Jungen) bzw. selten-leichten Wörtern (Mädchen), bei den Schülern der 2. Klasse zu den seltenen Diktatwörtern (Jungen) bzw. zum häufig-schwierigen Text (Mädchen). Insgesamt gesehen fallen die Zusammenhänge für die Viertkläßler geringfügig höher aus.

Betrachtet man die Anteile der erklärten gemeinsamen Varianz, so können die hierfür erzielten Maximalwerte (29 % bei den Viertkläßlern, 16 % bei den Zweitkläßlern) absolut nicht befriedigen, wenn man die Tests als Indikatoren für eine einzige hypothetische Größe, die Rechtschreibkompetenz auffaßt. In dieser Hinsicht müssen aber auch die Interkorrelationen zwischen den vier systematisch variierten Diktattexten überraschen. Nur in drei Fällen (beim Zusammenhang zwischen SS und HL für beide Klassenstufen, zwischen SL und SS für Viertkläßler bzw. zwischen SL und HS für Zweitkläßler) treten Korrelationen auf, die auf eine gemeinsame zugrundeliegende Größe schließen lassen. Ansonsten bewegen sich die Koeffizienten gerade für die Zweitkläßler eher in Bereichen, die statt (zumindest bei den schwierigen Diktatproben zu erwartenden) hohen Paralleltest-Reliabilitäten einer annähernden Unabhängigkeit der Messungen entsprechen. Da trotz der im Durchschnitt höheren Korrelationen Gleiches auch für die Viertkläßler behauptet werden kann, scheint die Inhomogenität der Resultate zuverlässige Aussagen über die jeweilige Rechtschreibfertigkeit kaum zuzulassen.

### *3.2. Multivariate Analysen*

Diese Vermutung soll im folgenden über ein multivariates Verfahren genauer überprüft werden, wobei insbesondere gefragt wird, wie gut die Leistungsvarianz im normorientierten Testverfahren durch das kombinierte Ergebnis der nicht standardisierten, dafür aber eng am Curriculum orientierten Diktatproben prognostiziert werden kann.

Tabelle 3 gibt die Befunde der multiplen Regressionsanalysen für Zweit- und Viertkläßler getrennt wieder. Wie schon aufgrund der Interkorrelationsmuster zu erwarten war, klären die vier Diktatvarianten bei den Zweitkläßlern weniger AST-Varianz auf: während hier der multiple Korrelationskoeffizient .49 und der Determinationskoeffizient ( $R^2$  als Index der aufgedeckten gemeinsamen Varianz) damit .24 beträgt, fallen die entsprechenden Werte bei den Viertkläßlern mit .60 bzw. .36 etwas höher aus. Über das multiple Regressionsverfahren lassen sich demnach nur 24 % (Zweitkläßler) bzw. 36 % (Viertkläßler) der AST-Leistungsvarianz aus der Prädiktor-Kombination vorhersagen. Da das gewählte statistische Verfahren die wahren Zusammenhänge aufgrund seiner Vorgehensweise mit Sicherheit überschätzt, können die erzielten Ergebnisse nicht befriedigen.

**Tabelle 3:** Ergebnisse der multiplen Regression des AST-Rechtschreibtests auf die vier ausgewählten Diktatkomponenten*a) Zweitkläßler*

	mult. R.	R-Quadr.	Anstieg	beta	RFSK
HS	.32	.10	.10	.20	.20
HL	.35	.13	.03	-.30	.06
SS	.49	.24	.11	.54	.22
SL	.49	.24	.00	.11	.00

*b) Viertkläßler*

	mult. R.	R-Quadr.	Anstieg	beta	RFSK
HS	.41	.17	.17	.13	.29
HL	.42	.18	.01	-.68	.00
SS	.60	.36	.18	.98	.30
SL	.60	.36	.00	-.01	.00

*Legende*

RFSK = Regressions-Faktor-Struktur-Koeffizient

Analysiert man die Befunde nun im Hinblick auf die effizientesten Einzelprädiktoren (vgl. dazu den in der 3. Spalte wiedergegebenen Anstieg des Determinationskoeffizienten), so kann bei beiden Klassenstufen von einer strukturell ähnlichen Konstellation ausgegangen werden: die Leistung im AST-Rechtschreibtest läßt sich jeweils am besten durch die Diktatvarianten HS (häufige und schwierige Wörter) bzw. SS (selten-schwere Wörter) vorher-sagen. Dies kommt nicht nur im Anstieg des Determinationskoeffizienten, sondern auch in den (meist allerdings eher instabilen) Beta-Gewichten sowie den Regressions-Faktor-Struktur-Koeffizienten (RFSK) zum Ausdruck, die das individuelle Prädiktorgewicht adäquater wiedergeben (vgl. Cooley und Lohnes, 1971).

Aus diesen Befunden kann also gefolgert werden, daß die AST-Maße die größten Gemeinsamkeiten mit denjenigen Diktatproben aufweisen, die vorwiegend schwieriges Wortmaterial im oben definierten Sinne enthalten. Im Hinblick auf die Forderung nach hinreichender Vorkommenshäufigkeit scheint insbesondere interessant, daß der Diktatvarianten SS (selten-schwierige Wörter) in beiden Klassenstufen das größte Vorhersagegewicht zukommt.

Wie schlecht es um die Prognosemöglichkeiten bestellt ist, kann schließlich an einer Aufgabe demonstriert werden, bei der auf den vollen Informationsgehalt der abhängigen Variablen verzichtet wird. Es geht hierbei um die Frage, wie gut die nach dem AST-Mittelwert dichotomisierten Gruppen guter und schlechter Rechtschreiber über die kombinierte Analyse der vier Diktatproben rekonstruiert werden können. In Tabelle 4 sind die wesentlichen Befunde der hierzu durchgeführten linearen Diskriminanzanalyse aufgeführt.

**Tabelle 4:** Kreuzvalidierte Re-Klassifikation für die nach den AST-Ergebnissen als gut (1) bzw. schlecht (2) eingestuften Rechtschreiber (Befunde der linearen Diskriminanzanalysen)

a) 2. Klassen			b) 4. Klassen			
	1	2	Treff.-%	1	2	Treff.-%
1	48	37	56.5	70	26	72.9
2	52	89	63.1	48	136	73.9
	100	126	60.6	118	162	73.6

Ähnlich wie bei der Regressionsanalyse liegt auch diesmal die (jackknife-korrigierte) Trefferquote für die Zweitkläßler (60,6 %) unter dem für die Viertkläßler errechneten Wert (73,6 %). Wenn man sich vergegenwärtigt, daß auch Zufallsklassifikationen eine Wahrscheinlichkeit von 50 % besitzen, können die Werte für die Zweitkläßler absolut nicht überzeugen, während sie für die Viertkläßler allenfalls hinreichende Klassifikationsstabilität anzeigen. Es stellt sich demnach heraus, daß die nach ihren AST-Rechtschreibleistungen separierten Gruppen guter und schlechter Rechtschreiber über die kombinierten Befunde der übrigen vier Rechtschreibsituationen nur unbefriedigend rekonstruiert werden können. Dies überrascht um so mehr, als lediglich eine Grobklassifikation gefordert war.

#### 4. Diskussion

Nachdem die theoretischen Erörterungen zur Rechtschreibproblematik den Schluß nahegelegt hatten, daß weniger die allgemeine Intelligenz als vielmehr basale Gedächtnismerkmale für das Erlernen der Schriftsprache von Bedeutung sind, ließ sich daraus für die Rechtschreibdiagnose ableiten, daß hier ein repräsentativer Ausschnitt des Unterrichtsangebots, m. a. W. Wortmaterial mit hinreichender Vorkommensfrequenz Verwendung finden müßte. Die empirische Analyse sollte zunächst Klarheit darüber verschaffen, ob die zur Diagnose üblicherweise verwendeten normierten Rechtschreibtests diesem Aspekt überhaupt Rechnung tragen. Von einer sinnvollen Anwendung dieser Tests kann also nur dann gesprochen werden, wenn die Prüfwörter auch hinreichend häufig im Unterricht vorgekommen sind.

Wie jedoch schon der einfache korrelative Vergleich mit den vier systematisch nach Vorkommenshäufigkeit und (theoretischer) Wortschwierigkeit variierten Diktatproben ergab, waren ausgeprägtere Beziehungen lediglich zu den Befunden für die seltenen Items zu verzeichnen. Bestätigt wurde dieses Ergebnis auch durch die Resultate der multiplen Regressionsanalysen, denen zufolge die Rechtschreibleistung im normierten Test für beide untersuchten Klassenstufen am besten durch das Abschneiden im Diktat der (lehrplanbezo-



gen) seltenen und schwierigen Wörter vorhergesagt werden konnte (dies verwundert kaum, wenn man die Vorkommenshäufigkeit der AST-Zielwörter direkt im Frequenzwörterbuch nachschlägt: so tauchen z. B. bei den Viertkläßlern 18 der 20 herangezogenen Testwörter der Testform A nur viermal oder weniger, fünf davon überhaupt nicht auf).

Diese Befunde verdeutlichen in jedem Fall, daß von einer Lehrplangültigkeit des normierten Rechtschreibtests nicht ausgegangen werden kann. Welche Auswirkungen dies auf die Möglichkeit hat, die individuelle Rechtschreibkompetenz näher zu bestimmen, zeigt der multivariate Vergleich mit den vier systematisch selektierten Rechtschreibproben: da die gemeinsamen Varianzanteile für die Zweitkläßler lediglich 24 %, für die Viertkläßler allenfalls 36 % ausmachen, müssen die Aussagemöglichkeiten aufgrund des standardisierten Verfahrens äußerst begrenzt erscheinen. Dies verdeutlicht auch der Befund der linearen Diskriminanzanalyse, anhand derer die über den AST-Rechtschreibtest erfolgte Klassifikation von Gruppen guter und schwacher Rechtschreiber durch die Ergebnisse der vier Diktierproben rekonstruiert werden sollte. Die bei den Zweitkläßlern absolut unbefriedigenden, bei den Viertkläßlern gerade genügenden Trefferquoten weisen eindringlich auf das Problem hin, daß mit dem normierten Verfahren nur wenig valide Beurteilungen der individuellen Rechtschreibkompetenz vorgenommen werden können.

Einwände gegen eine Verallgemeinerung dieses Befundes auf standardisierte Rechtschreibtests schlechthin lassen sich aus einer theoretischen Position heraus erheben, die die mangelnde Invarianz der abhängigen Variablen (Rechtschreibperformanz) mit den unterschiedlichen Erhebungsweisen (z. B. multiple-choice-Verfahren, Lückentests, Diktate von Einzelwörtern bzw. zusammenhängenden Texten etc.) in Verbindung bringt. Wenn diese Vermutung zutreffen sollte, wäre es durchaus möglich, die wenig überzeugenden Befunde für die AST-Subtests ihren nicht unproblematischen Präsentationsmodi<sup>5</sup> zuzuschreiben. Dieses Argument wird allerdings durch die (laut Testmanual) generell hohen Korrelationen des Verfahrens mit anderen standardisierten Rechtschreibtests (z. B. dem Diagnostischen Rechtschreibtest) sowie traditionellen Diktierverfahren weitgehend entkräftet<sup>6</sup>. Es liegt demnach viel näher, die Gültigkeit der hier vorgefundenen Ergebnisse auch für andere normierte Verfahren zu postulieren.

Ein zweiter wesentlicher Befund bezog sich auf den Zusammenhang zwischen den Ergebnissen in den vier verschiedenen Diktaten. Die Erwartung einer generell konsistenten Leistungsrangfolge ließ sich in beiden Klassenstufen nur für die Beziehung zwischen den objektiv leichtesten und schwierigsten Textkomponenten (HL und SS), sowie noch für den Zusammenhang SL–HS (Zweitkläßler) und SL–SS (Viertkläßler) bestätigen. Demgegenüber fielen die übrigen Koeffizienten für die Viertkläßler überraschend niedrig aus und legten bei den Zweitkläßlern die Annahme eher unabhängiger Messungen nahe.

Es läßt sich daran zum einen demonstrieren, welche enorme Bedeutung die jeweilige Zusammensetzung der Prüfwörter gerade bei jüngeren Kindern be-

sitzt, zum anderen aber auch die Schwierigkeit erkennen, einen zuverlässigen Indikator für die Rechtschreibkompetenz von Grundschulern zu gewinnen. Dies mag bei Viertkläßlern innerhalb gewisser Grenzen noch gelingen, stellt bei den Zweitkläßlern angesichts der unzulänglichen „Paralleltest-Reliabilität“ aber ein gravierendes Problem dar. Speziell für beginnende Rechtschreiber kann also *die Rechtschreibfertigkeit* kaum objektiv bestimmt werden: die Leistungs-Rangfolge in inhaltsvaliden Übungsdiktaten variiert sehr stark in Abhängigkeit davon, welche Ausprägungen der Ordnungsparameter (Worthäufigkeit bzw. -schwierigkeit) herangezogen wurden.

Es ist demnach jedenfalls immer dann besondere Zurückhaltung angebracht, wenn für die Einschätzung der Rechtschreibleistung lediglich singuläre (und/oder normorientierte) Messungen vorliegen. Angesichts der eingangs erörterten Probleme bei der Explikation orthographischer Lernprozesse sollte (etwa im Kontext der Frühdiagnose von Rechtschreibschwierigkeiten) schließlich stärker als bisher darauf geachtet werden, die faktische Vorkommenshäufigkeit zu berücksichtigen, da damit eine wesentliche Fehlerquelle bei der Leistungseinschätzung eliminiert werden kann.

### Anmerkungen

- 1 Während Glogauer für Schüler des 4., 5. und 8. Schuljahrs Korrelationskoeffizienten zwischen .37 und .48 registrierte, variierten die entsprechenden Werte bei Pfeiffer und Zielinski für Viertkläßler zwischen .38 und .54.
- 2 Genauere Informationen zur Erhebungsstrategie bei Schneider (1979).
- 3 Zur detaillierten Programmbeschreibung vgl. Hauenstein et al. (1978).
- 4 Die Maximal-Anzahl der Zielwörter in den einzelnen Diktatproben variierte zwischen 16 und 20 Items.
- 5 Bei den Zweitkläßlern multiple-choice-Ankreuzverfahren, bei den Viertkläßlern Selektion und Korrektur falsch geschriebener Wörter.
- 6 Allein aufgrund dieser engen Beziehungen zu anderen standardisierten Rechtschreibverfahren schien es im übrigen gerechtfertigt, Daten aus diesem Subtest für den Vergleich zu benutzen.

### Literatur

- Cooley, W. W., Lohnes, P. R.: *Multivariate data analysis*. New York: Wiley, 1971
- Fingerhut, W., Langfeldt, H.-P.: Erfahrungen mit dem Allgemeinen Schulleistungstest für 4. Klassen (AST 4). *Psychologie in Erziehung und Unterricht*, 20, 249–257 (1973)
- Glogauer, W.: Rechtschreibleistung und Intelligenz. *Psychologie in Erziehung und Unterricht*, 24, 287–292 (1977)
- Hauenschild, Ch., Huckert, E., Maier, R.: SALAT: Entwurf eines automatischen Übersetzungssystems. *Sprache und Datenverarbeitung*, 2, 126–152 (1978)
- Hitzenberger, L., Konrad, W., Krause, J., Schneider, Ch.: Das Regensburger Textverarbeitungssystem COBAPH. *Sprache und Datenverarbeitung*, 1, 30–32 (1977)
- Kaeding, F. W.: *Häufigkeitwörterbuch der deutschen Sprache*. Steglitz, 1898
- Kemmler, L.: *Erfolg und Versagen in der Grundschule*. Göttingen: Hogrefe, 1967
- Kern, A., Kern, E.: *Der neue Weg im Rechtschreiben*. Freiburg: Herder, 1954

- Masendorf, F., Tücke, M.: Zur Stabilität und faktoriellen Gültigkeit des AST 2 und des , AST 3. *Psychologie in Erziehung und Unterricht*, 22, 314–317 (1975)
- Meier, H.: *Deutsche Sprachenstatistik*, Bde. 1 u. 2. Hildesheim, 1964
- Ortmann, W.D.: *Beispielwörter für deutsche Ausspracheübungen*. München: Goethe-Institut, 1975
- Pfeiffer, G., Zielinski, W.: Über den Zusammenhang zwischen Rechtschreibung und Intelligenzleistung. *Psychologie in Erziehung und Unterricht*, 22, 1–8 (1975)
- Rieder, O.: *Allgemeiner Schulleistungstest für 2. Klassen (AST 2)*. Weinheim: Beltz, 1971
- Schneider, W.: *Taxonomie der Gedächtnisleistungen schwacher und normaler Rechtschreiber*. In: H. Späth (Hrsg.), *Fallstudien Cluster-Analyse*. München: Oldenbourg, 1977
- Schneider, W.: *Materialien zur deutschen Rechtschreibung, Textsammlung von Probearbeiten, Diktaten und Schulbüchern (alphabetische Indices, Frequenzranglisten, Konkordanzen)*. Nürnberg: *Regensburger Microfiche Materialien*, 1979
- Schneider, W.: *Bedingungsanalysen des Recht-Schreibens*. Bern: Huber, 1980
- Schneider, W., Springer, A., Petermann, F.: *Verschlüsselung von verbaler Information bei schwachen und normalen Rechtschreibern*. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 9, 113–118 (1977)
- Simon, D. P.: *Spelling — A task analysis*. *Instructional Science*, 5, 277–302 (1976)
- Simon, D. P., Simon, H. A.: *Alternative uses of phonemic information in spelling*. *Review of Educational Research*, 43, 115–137 (1973)
- Tarver, S. G., Hallhan, D. P., Kaufman, J. M.: *Verbal rehearsal and selective attention in children with learning disabilities: A developmental lag*. *Journal of Experimental Child Psychology*, 22, 375–385 (1976)
- Torgesen, J., Goldman, T.: *Verbal rehearsal and short-term memory in reading-disabled children*. *Child Development*, 48, 56–60 (1977)
- Zur Oeveste, H.: *Untersuchung zur Rechtschreibung von Schülern*. In: K. J. Klauer, H.-H. Kornadt (Hrsg.), *Jahrbuch für empirische Erziehungswissenschaft*. Düsseldorf: Schwann, 1977

Dr. Wolfgang Schneider  
Psychologisches Institut der Universität  
Hauptstr. 47–51  
D-6900 Heidelberg  
Tel.: 0 62 21 / 54 73 76