

Rechtschreibtests im Vergleich: Wie stellt man deren Güte fest und wie besser nicht?

Anmerkungen zur Kritik von Tacke, Völker und Lohmüller an der HSP*

Comparing spelling tests: How to determine the quality the right way

Comments on the critical notes of Tacke, Völker und Lohmüller to the HSP

Peter May, Volkmar Malitzky, Ulrich Vieluf

Institut für Lehrerfortbildung, Hamburg

Im Beitrag von Tacke et al. werden Ergebnisse einer Erhebung von Rechtschreibleistungen in Schwäbisch-Hall berichtet und auf dieser Grundlage die Normierung der HSP 3, die Auswertung von Strategieprofilen sowie die Bestimmung der „Alphabetischen Strategie“ kritisiert. Der erste Teil der folgenden Ausführungen setzt sich mit dieser Kritik auseinander. Im zweiten Teil werden erste Ergebnisse einer aktuellen Untersuchung zum Vergleich von HSP und DRT berichtet, die zeigen, dass die HSP nach Einschätzung der Lehrkräfte besser zur Diagnose von Rechtschreibleistungen geeignet ist.

Erwiderung zur Kritik von Tacke, Völker und Lohmüller

Die Autoren kommen aufgrund der Ergebnisse einer Erhebung von Rechtschreibleistungen in Schwäbisch-Hall zu dem Schluss, dass die Normierung der HSP 3 schlechter mit der Werteverteilung der Stichprobe übereinstimme als die Werte des Diagnostischen Rechtschreibtests (DRT 3) und des Weingartener Grundwortschatz Rechtschreibtests (WeiGrT 3+).¹ Diese Kritik an der Normenverteilung der HSP 3 ist zum Teil berechtigt, sie wird jedoch aufgrund von Untersuchungsfehlern grob überzeichnet. Tatsächlich weisen die Normen der HSP 3 einige Inkonsistenzen auf, die sich in einer umfangreichen Längsschnittuntersuchung im Rahmen der Evaluation des Hamburger Projekts „Lesen und Schreiben für alle“ (PLUS) gezeigt haben (May 2000). Abbildung 1 zeigt die Verteilung der T-Werte in den HSP 1 bis 4 von Hamburger Kindern, die im Laufe der Grundschulzeit alle Versionen der HSP bearbeitet haben. Darin ist zu erkennen, dass die Verteilung der Normen für die HSP 3 nach rechts verschoben ist; dies belegt auch der höhere Mittelwert.

* Der Beitrag ist veröffentlicht in „Psychologie in Erziehung und Unterricht“, 48. Jg., Nr 2/2001, S. 146-152.

¹ Die Autoren bezeichnen den Weingartener Grundwortschatz Rechtschreibtest missverständlich als WRT, obgleich dies das eingeführte Kürzel des bekannten und noch immer verbreiteten Westermann Rechtschreibtests ist.

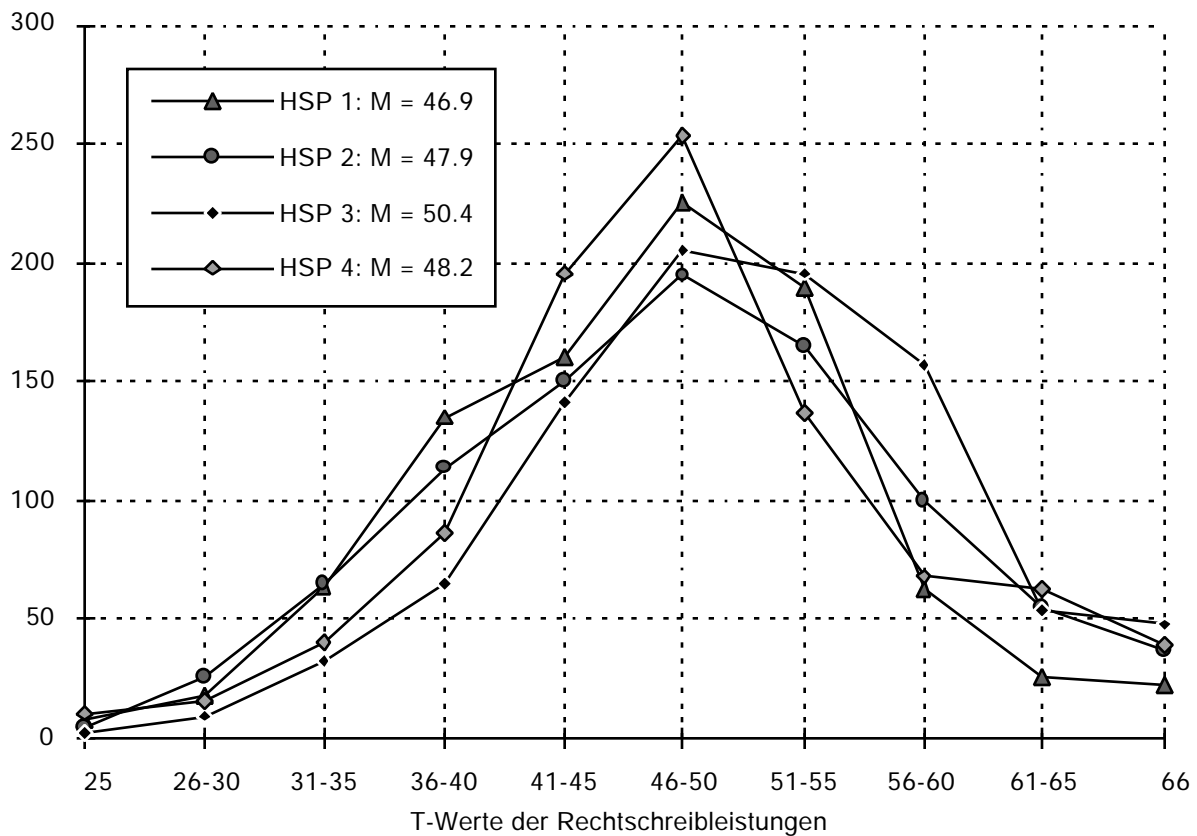


Abb. 1: Normen der HSP 1 bis 4 in einer Längsschnittstichprobe (n = 911)

Da nicht anzunehmen ist, dass die Hamburger Kinder in der dritten Klassenstufe im Vergleich zu ihren bundesdeutschen Altersgenossen einen plötzlichen Leistungsanstieg aufweisen, muss für die HSP 3 tatsächlich eine divergierende Normierung angenommen werden. Die Gründe für diese Normabweichung liegen vor allem darin, dass bei der Eichung der HSP als eines bundesweiten Rechtschreibtests in den Jahren 1989 bis 1994 die Eichstichprobe offenbar unzureichend repräsentativ war; dies gilt insbesondere für die erstmals einbezogenen neuen Bundesländer. Zur Überprüfung und Korrektur dieser Inkonsistenzen wurden im Schuljahr 1999/2000 umfangreiche Erhebungen durchgeführt, deren Ergebnisse in der erweiterten Neuauflage der HSP (Herbst/Winter 2000) auch zu einer Aktualisierung der Vergleichsnormen führen werden.

Allerdings ist das Ausmaß der Verschiebung der HSP 3-Normen keineswegs so hoch, wie dies von Tacke et al. behauptet wird. Die von ihnen berichteten Ergebnisse sind verzerrt, da ihre Untersuchung einige „handwerkliche Fehler“ (Stichprobenauswahl, Auswertungskriterien, Gruppeneinteilung) aufweist. Daraus ergeben sich auch verfälschende Zahlenverhältnisse bei der Analyse der Auswertung von Strategieprofilen mit der HSP.

- Die Stichprobe in Schwäbisch-Hall entspricht nicht der bundesdeutschen Stichprobe, sondern liegt in den Durchschnittswerten deutlich darüber. Bezogen auf die Werte der HSP beträgt der Mittelwertunterschied etwa eine halbe Standardabweichung bei gleichzeitig deutlich geringerer Streuung. Die Tatsache, dass die Werteverteilung der Schwäbisch-Haller Stichprobe

den Normenverteilungen des DRT und WeiGrT eher entspricht, wird von den Autoren zu Unrecht als Indiz für eine geringe Tauglichkeit der HSP-Normen interpretiert. Wie im zweiten Teil unserer Stellungnahme gezeigt wird, spricht dies eher für die Annahme, dass die beiden anderen Tests zu strenge Normen zugrunde legen. Erklärbar ist dies u.a. durch die Tatsache, dass in der Eichstichprobe des WeiGrT das Flächenland Baden-Württemberg deutlich überrepräsentiert ist, während die sozial stärker belasteten Stadtstaaten Berlin und Hamburg fehlen. Zusätzlich wird das Leistungsspektrum der untersuchten Stichprobe durch das Ausklammern der Kinder nicht-deutscher Muttersprache nach oben verschoben. Diese Nichtrepräsentativität ihrer Stichprobe gilt es im Bewusstsein zu halten, wenn wir die Folgerungen verstehen wollen, die die Autoren für die Bewertung der HSP ziehen.

Die Tatsache, dass selbst im WeiGrT nur 18 von 230 Kindern (also 7,8 Prozent) im WeiGrT ein Ergebnis erzielen, das nach der bundesdeutschen Stichprobe dem Leistungsbereich der Prozenträge 0 bis 10 entspricht, erklärt sich dadurch, dass Schwäbisch-Hall soziokulturell eher wenig belastet ist. Da der mit Hilfe des WeiGrT ermittelte Anteil schwacher Rechtschreiber jedoch am höchsten unter den drei verglichenen Tests ist, spricht dies nach Meinung von Tacke et al. für seine besondere Eignung. Die Autoren scheinen sich demnach bei ihrer Beurteilung vor allem daran zu orientieren, ob die verwendeten Tests möglichst viele Kinder als förderbedürftige Legastheniker ausweisen.

- Eine weitere Einschränkung der Aussagefähigkeit der Ergebnisse bildet die Gruppeneinteilung für die Vergleiche zwischen den Testwerteverteilungen: Sie erfolgt bei Tacke et al. nach unterschiedlichen Kriterien, so dass die Stichproben der jeweils einbezogenen Kinder zwischen 88 und 119 schwanken. So werden Vergleiche anhand unterschiedlicher Stichproben gezogen. Ein aussagefähiger Vergleich müsste jedoch an der identischen Stichprobe erfolgen (siehe Teil 2). Wenn zudem anstelle der T-Werte verteilungsabhängige Prozenranggrenzen gewählt werden, erscheint die Zuordnung der Kinder in die verschiedenen Leistungsgruppen relativ willkürlich.

- Die Schreibungen werden nach dem Gesichtspunkt der richtigen Wortschreibungen ausgewertet. Dies ist bei den Tests DRT und WeiGrT das übliche Auswertungskriterium. Demgegenüber bildet die Graphemtreffermethode das entscheidende Auswertungsverfahren der HSP. Im Handbuch der HSP wird vor der Beschränkung auf die wortweise Auswertung für die Diagnose bei rechtschreibschwächeren Kindern ausdrücklich gewarnt. Dies lässt sich am Beispiel ausgewählter Schreibungen des Wortes „Fahrrad“ zeigen (Tabelle 1). Da die HSP im Gegensatz zu den beiden anderen Tests gerade nicht den geläufigen Schreibwortschatz erhebt, sondern auf die orthographische Kompetenz zielt, enthält sie einen größeren Anteil von Wörtern mit relativ hoher Komplexität. Die Beschränkung auf die Auswertung der Schreibwörter nach dem „Alles-oder-Nichts“-Prinzip führt insbesondere bei leistungsschwächeren Schülern zu Fehleinschätzungen.

Tabelle 1: Unterschiedliche orthographische Bewertungsmaßstäbe am Beispiel verschiedener Schreibungen des Wortes "Fahrrad"²

Schreibung	Bewertungsmaßstab		
	Wort richtig/falsch	realisierte Kompetenz	Graphem- treffer
FT	falsch	alphabetisches Schreiben: verkürzte Lautfolge	1
Fart	falsch	alphabetisches Schreiben entfaltet	2
Farat	falsch	vollständiges alphabetisches Schreiben	3
Farad	falsch	orthographisches Regelement bezeichnet	4
Fahrad	falsch	orthographisches Merkelement bezeichnet	5
Fahrrad	richtig	Kompositum morphematisch durchdrungen	6

Obwohl für Profilanalyse, die von den Autoren ebenfalls thematisiert wird, die graphemweise Auswertung vorlag, ziehen die Autoren auch bei der HSP ausschließlich die wortweise Auswertung heran. Dies schränkt die Aussagekraft der Ergebnisse ein und blendet gerade die Dimension der HSP aus, die differenzierte Ergebnisse ermöglicht. Insofern trifft die Kritik nicht die HSP, sondern das gewählte Auswertungsverfahren der Autoren.

- In einem zweiten Kritikpunkt wird von den Autoren die Gültigkeit der Strategieprofilanalyse angezweifelt. Diese stellt neben der Graphemtreffermethode ein Herzstück der HSP dar und wird im Handbuch ausführlich erläutert. Für die Klassenstufen 2 und 4 werden konkrete Zahlen zum Vorkommen verschiedener Profiltypen angeführt³ und die Gültigkeit der Profilanalyse empirisch belegt. Leider nehmen die Autoren darauf keinen Bezug.

Die Kritik an der Profilauswertung von Tacke et al. wird auf der Basis ihrer unzureichenden Stichprobenauswahl geführt: Anstatt die empfohlenen Werte aus dem Handbuch der HSP, die sich auf die bundesdeutsche Verteilung beziehen, zu nutzen, konstruieren die Autoren für ihre Teilstichprobe eigene Kriteriumswerte. Dies ist schon deshalb fragwürdig, weil dadurch aus einem bundesweiten Vergleichstest ein regionales Verfahren abgeleitet wird. Durch die Auswahl einer nicht repräsentativen Stichprobe verändern sich natürlich auch die Kriteriumswerte: Infolge der geringeren Streuung sinken die Werte für die Reliabilität und steigen die kritischen T-Werte-Differenzen auf fast den doppelten (!) Wert. Dass auf diese Weise der Wert der Profilanalyse eingeschränkt ist und kaum noch relevante Profildifferenzen erscheinen, liegt auf der Hand. Man fragt sich, warum die Autoren an die Stelle der eingeführten Handbuchwerte ein

² Vgl. May (1998), S. 15. Die Frage der Großschreibung des Wortes bleibt hier außer Betracht.

³ Im neuen Handbuch der HSP werden die entsprechenden Werte für alle Jahrgangsformen aufgeführt.

nicht geeignetes Regionalkriterium setzen, um anschließend dem Test vorzuwerfen, dass er so nicht funktioniert.

Darüber hinaus legen die Autoren ihrer Analyse der Profilauswertung ein unzulässiges Auswahlkriterium zugrunde: Während das im Handbuch ausführlich beschriebene Konzept der HSP ausdrücklich empfiehlt, alle Profildifferenzen, die über einem kritischen Wert liegen, als Anzeichen für eine mögliche Störung des Aneignungsprozesses zu werten, betrachten die Kritiker nur solche Profildifferenzen als relevant, bei denen mindestens eine Strategie einen Wert oberhalb PR 50 erzielt. Das Konzept der Profilanalyse, das der HSP zugrunde liegt, geht jedoch ganz bewusst davon aus, dass die Ausprägung der verschiedenen Zugriffsweisen auf die Schreibungen in ihrer jeweiligen Relation zueinander, also weitgehend *unabhängig* von der Höhe der Gesamtleistung, interpretiert werden kann.

Damit wird eine zusätzliche Analyse der Schreibungen über die Ausgewogenheit der Strategien ermöglicht. Die

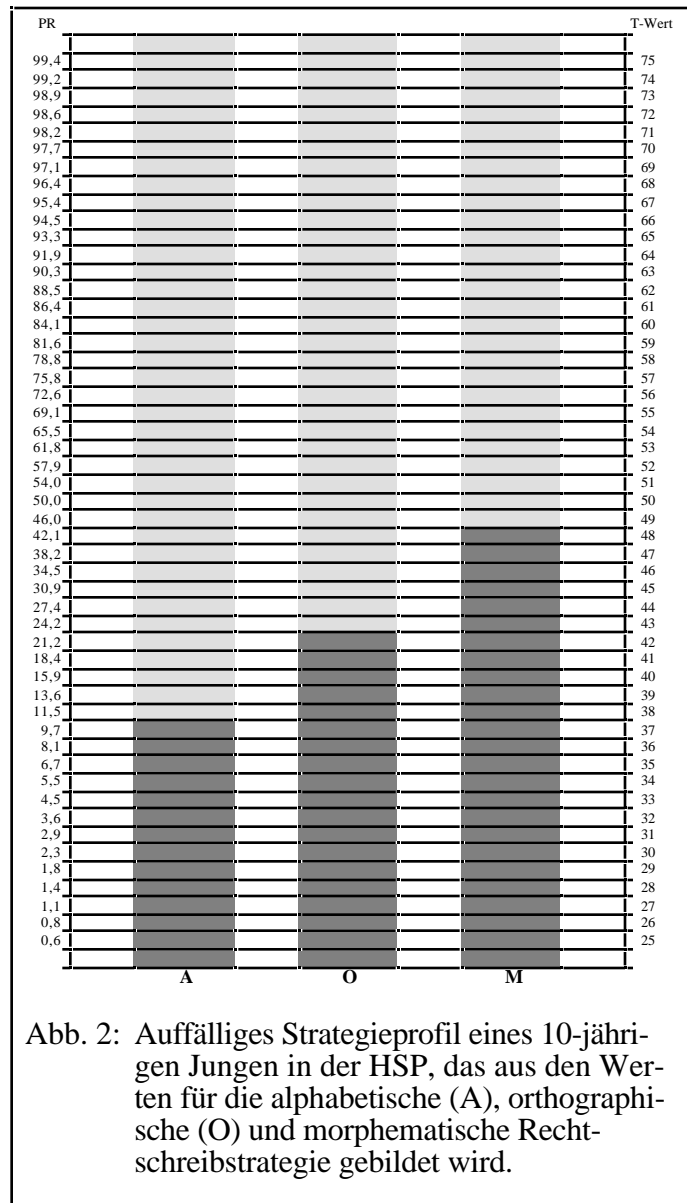


Abb. 2: Auffälliges Strategieprofil eines 10-jährigen Jungen in der HSP, das aus den Werten für die alphabetische (A), orthographische (O) und morphematische Rechtschreibstrategie gebildet wird.

Autoren Tacke et al. ändern jedoch dieses Konzept willkürlich. Insofern ist der Befund, dass das von ihnen veränderte Auswertungskonzept nicht trägt, eher als Kritik an der von ihnen vorgenommenen Konzeptänderung zu werten.

Im Handbuch der HSP wird die Profilauswertung und die Differenzierung von Profiltypen ausführlich begründet und empirisch belegt. Abbildung 2 zeigt eines der beiden ausführlich im Handbuch dargestellten Fallbeispiele, bei dem gerade die Analyse des unausgewogenen Strategieprofils auf die Schwierigkeiten des Lernprozesses verweist, die sich aus der Gesamtleistung (PR 24) nicht ohne Weiteres erschließen lassen. Nach der von Tacke et al. vorgenommenen Veränderung würde dieses Strategieprofil jedoch nicht relevant sein, da keiner der Strategiewerte den PR 50 überschreitet.

Leider nehmen die Autoren auch bei diesem Aspekt das Handbuch der HSP nicht zur Kenntnis, sondern konstruieren ein eigenes Konzept, das weder theoretisch noch empirisch mit dem Original vergleichbar ist.

- Auf die Kritik der Autoren an der Bestimmung der alphabetischen Strategie soll hier nur kurz eingegangen werden. Die vorgebrachte Kritik ist nicht in den empirisch gewonnenen Daten fundiert, sondern hat den Status theoretischer Überlegungen. Die Kritiker zitieren nicht die entsprechenden Ausführungen im Handbuch, sondern stattdessen aus einer nicht zur Veröffentlichung bestimmten E-Mail-Korrespondenz.

Vielleicht verwechseln die Autoren deshalb phonologische und phonemische Zugriffsweisen, wenn sie behaupten, das Konzept der alphabetischen Strategie würde die der Dialektsprache entsprechenden, jedoch von der orthographisch relevanten Standardlautung abweichende Schreibungen ungenügend berücksichtigen. Dass diese Kritik nicht überzeugt, zeigen die Autoren selbst, wenn sie (ironisch?) als Alternative vorschlagen, die Werte für die alphabetische Strategie „für alle in Frage kommenden Mundarten separat zu normieren“. (Eine Steigerung wäre die Idee, gleich die Standardorthografie zugunsten vieler regionaler Mundartorthografien zu ersetzen.) Hier sei auf das im Herbst erscheinende, ergänzte Handbuch der HSP verwiesen. Darin wird an Beispielen verschiedener Dialekte gezeigt, dass es eben nicht so ist, wie die Autoren Tacke et al. spekulieren: Tatsächlich weisen Kinder aus Regionen mit starken Dialektinflüssen (z.B. Schwaben, Sachsen) nicht geringere, sondern tendenziell eher höhere Leistungen in der alphabetischen Strategie auf. Der Grund liegt wahrscheinlich darin, dass starke Dialektsprecher gezielt lernen müssen, sich beim Rechtschreiben an der hochsprachlichen Standardlautung zu orientieren. Ihre Selbstinstruktion lautet: „Schreibe, wie die Hochsprache klingt.“ Beherrschen sie dies - und daran führt beim Erwerb der orthographischen Kompetenz kein Weg vorbei -, setzen sie diese Strategie bewusster ein als Schreiber, deren Umgangssprache der Hochsprache ähnlicher ist. „Dialektfreie“ Sprecher mögen sich dann ein wenig länger der irreführenden Strategie anvertrauen: „Schreibe, wie du sprichst.“

Vergleich der praktischen Brauchbarkeit von DRT und HSP

Da von Tacke et al. die Brauchbarkeit der HSP generell in Frage gestellt wird, sollen hierzu einige aktuelle Ergebnisse einer eigenen Untersuchung vorgestellt werden.⁴ Im Rahmen der Forschungsarbeiten für die Aktualisierung und Erweiterung des HSP-Konzepts wurden bei Schülern einer Teilstichprobe neben der HSP (May, 1998) auch der DRT (Müller, 1990, 1997; Grund, Haug & Naumann, 1994) erhoben.⁵ Zudem wurden die Lehrkräfte gebeten, die Recht-

⁴ Die folgende Darstellung beschränkt sich auf die Grundschuljahrgänge. Eine vollständige Übersicht der Ergebnisse, die auch auf die Sekundarstufe beziehen, erscheint in der nächsten Auflage des Handbuchs der HSP.

⁵ Als Vergleichstest wurde der DRT herangezogen, da er ebenfalls bundesweit verbreitet ist, in verschiedenen Jahrgangsversionen vorliegt und wie die HSP nicht nur einen Gesamtwert für die Rechtschreibleistung ausweist, sondern auch eine qualitative Fehleranalyse ermöglicht.

schreibleistungen der Kinder auf einer Notenskala zu bewerten. Ausgehend von den Berichten aus der Praxis, in denen von Unterschieden bei der Anwendung verschiedener Rechtschreibtests berichtet wird, wurde die Frage untersucht, inwieweit die Tests hinsichtlich der diagnostischen Aussagekraft vergleichbar sind und wie die praktische Brauchbarkeit der Tests von den Lehrkräften, die sie einsetzen, beurteilt wird.

Die Rekrutierung der teilnehmenden Klassen geschah auf dem Wege eines schriftlichen Aufrufs, der sich an die Teilnehmer einer Tagung der Deutschen Gesellschaft für Lesen und Schreiben im Herbst 1999 richtete. Zusätzliche Testklassen wurden über die PH Schwäbisch-Gmünd (Dr. Plieninger) und die Universität Siegen (Prof. Brügelmann) vermittelt. Es handelt sich demnach um eine anfallende Stichprobe freiwillig teilnehmender Klassen, auf deren Auswahl von Seiten der Autoren kein Einfluss genommen wurde.

Die Teilstichprobe besteht aus Klassen in den Bundesländern Baden-Württemberg, Bayern, Hamburg, Hessen und Nordrhein-Westfalen. Tabelle 2 zeigt die Anzahl der Klassen und Kinder in den einzelnen Klassenstufen. Die Tests wurden durch die Lehrkräfte der teilnehmenden Klassen durchgeführt und zentral durch eingewiesene Hilfskräfte ausgewertet. Die Lehrkräfte wurden gebeten, das durchschnittliche Leistungsniveau ihrer Klasse auf einer dreistufigen Skala einzuschätzen. Dabei ordnete die große Mehrzahl der Lehrkräfte ihre Klassen in die Mittelkategorie einer dreistufigen Leistungsskala zu, die übrigen 15 Prozent schätzten ihre Klassen als unterdurchschnittlich ein. Demnach bilden die teilnehmenden Klassen nach Einschätzung der Lehrkräfte eine insgesamt knapp durchschnittlich leistungsstarke Stichprobe. In den teilnehmenden Grundschulen konnten mit einer Ausnahme jeweils alle Klassenstufen einbezogen werden. Wenn man davon ausgehen kann, dass die Leistungen der Klassen stark durch das schulische Umfeld beeinflusst werden, können die Leistungen der einzelnen Klassenstufen miteinander in etwa als vergleichbar gelten. Tabelle 2 zeigt die Ergebnisse der bisher ausgewerteten Klassen.⁶

Der Vergleich zwischen HSP und DRT hinsichtlich der mittleren T-Werte, die den Handbüchern entnommen bzw. aus den dort angegebenen Prozenträngen errechnet wurden, ergibt eine deutlich geringere Schwankung bei den HSP-Werten: Die mittleren T-Werte differieren maximal um 3.0 Punkte; dagegen unterscheiden sich die mittleren T-Werte der DRT-Jahrgänge um bis zu 6.4 Punkten. Bei den Streuwerten ergeben sich keine eindeutigen Unterschiede. Dies belegt, dass die HSP-Ergebnisse der einzelnen Jahrgänge besser miteinander vergleichbar sind, was bei einem Test zur langfristigen Lernentwicklungsdiagnose notwendig ist. Dagegen belegen die deutlich höheren Mittelwertschwankungen beim DRT, dass die Eichstichproben, aus denen die Vergleichswerte gewonnen wurden, sich deutlich unterscheiden. Für eine langfristige Entwicklungsdiagnose ist dies ein erhebliches Problem.

⁶ Da einige Klassen beide Tests vor dem in den Handbüchern vorgegebenen Zeitpunkten durchführten, fallen die Ergebnisse generell etwas schwächer aus als in den jeweiligen Eichstichproben.

Tabelle 2: Vergleich zwischen HSP und DRT in den Klassenstufen 1 bis 4: Durchschnittswerte und Zusammenhänge mit der Schulnote

Bezugsgröße	Klasse 1		Klasse 2		Klasse 3		Klasse 4	
	Rohwerte	T-Werte	Rohwerte	T-Werte	Rohwerte	T-Werte	Rohwerte	T-Werte
HSP: Mittelwert	48.2	48.9	129.9	48.8	171.1	47.9	254.3	45.9
HSP: Standardabweichung	7.9	10.0	9.3	8.5	13.9	8.6	15.9	7.7
DRT: Mittelwert	11.9	44.0	17.8	49.3	26.6	42.9	26.2	45.9
DRT: Standardabweichung	9.3	10.2	6.3	7.7	10.1	10.0	8.7	8.9
Korrelation HSP/DRT (*)	-.79	.84	-.77	.80	-.78	.85	.79	.77
Korrelation HSP/Rechtschreibnote (*)	-.89	-.87	-.80	-.79	-.82	-.85	-.73	-.71
Korrelation DRT/Rechtschreibnote (*)	-.85	-.79	.75	-.74	.74	-.75	-.72	-.72
Anzahl teilnehmender Klassen	8		9		12		11	
Anzahl Kinder	176		194		256		236	

(*) Die verschiedenen Vorzeichen ergeben sich durch die unterschiedliche Ausrichtung der Werte (Notenskala: Ziffern 1-6, HSP: Zahl der Graphentreffer, DRT 1 und 2: Anzahl der Wortfehler, DRT 3 und 4: Anzahl der richtigen Wörter)

Die bessere Übereinstimmung der Jahrgangswerte beweist noch nicht, dass die absolute Höhe der Vergleichswerte besser abgestimmt ist. Da wir nicht über eine bundesweite Repräsentativerhebung mit beiden Testverfahren verfügen, ziehen wir als Gütekriterium die Übereinstimmung mit der Lehrerbeurteilung heran. Die beiden Tests korrelieren sowohl in den Rohwerten als auch in den T-Werten hochsignifikant miteinander, was dafür spricht, dass beide im Kernbereich der erfassten Fähigkeit übereinstimmen. Dennoch liegen Werte für die Korrelation mit der Rechtschreibnote bei der HSP generell etwas höher als beim DRT. Entscheidend für die von Tacke et al. in Frage gestellte Güte der Vergleichsnormen ist die Tatsache, dass dieser für die HSP auszuweisende engere Zusammenhang mit den Schulnoten nicht nur für die Rohwerte gilt, sondern auch in drei der vier Jahrgänge für die T-Werte. Dies beweist, dass nicht nur die Erfassung der Rechtschreibleistung mit der HSP (Graphentreffer-Methode) enger mit dem Expertenurteil übereinstimmt als die DRT-Ergebnisse, sondern dass auch die Normierung der HSP den Einschätzungen der Lehrkräfte näher kommt als die Normen des DRT.

Um die Brauchbarkeit der Tests für die Schulpraxis zu ermitteln, baten wir die teilnehmenden Lehrkräfte, die beiden von ihnen durchgeführten Rechtschreibtests mit Hilfe der üblichen Notenskala hinsichtlich der Aspekte praktische Handhabung der Tests, Auswahl der Wörter und Sätze, diagnostische Genauigkeit, Möglichkeit zur Ableitung von Fördermaßnahmen und Hinweise zur Unterrichtsplanung zu beurteilen und ihre Bewertungen zu begründen. Beide Tests erhielten von den Lehrkräften eher positive Bewertungen (mittlere Bewertung: 2.6). Dennoch liegen die mittleren Beurteilungen der Lehrkräfte bei allen fünf Fragen für die HSP günstiger (Durchschnittsbewertung: 2.3) als für den DRT (2.9). Vor allem hinsichtlich der praktischen Handhabbarkeit und der Auswahl der Wörter und Sätze beurteilen die Lehrkräfte die HSP im Durchschnitt deutlich positiver. Dies gilt insbesondere für die unteren beiden Klassenstufen.

Als Begründung dafür wird u.a. angeführt, dass die HSP für die Kinder in der Art der Darbietung, der Wortauswahl und Aufmachung motivierender sei und als „nicht so schwierig“ empfunden wird.

Zusammenfassend kann aufgrund dieser Ergebnisse festgestellt werden, dass die HSP nicht nur nach ihrem theoretischen Konzept für eine entwicklungsbezogene Förderdiagnose geeignet ist, sondern auch der empirischen Überprüfung im Vergleich mit alternativen Konzepten standhält. Allerdings sollte diese in dafür geeigneter Form erfolgen.

Literaturhinweise:

- Grund, M., Haug, G. & Naumann, C.L. (1994). *Diagnostischer Rechtschreibtest für 4. Klassen. Beiheft mit Anleitung und Normentabellen*. Weinheim: Beltz.
- May, P. (1998). *Diagnose orthographischer Kompetenz. Zur Erfassung der grundlegenden Rechtschreibstrategien mit der Hamburger Schreib-Probe (4. Aufl.)*. Hamburg: Verlag für pädagogische Medien.
- May, P. (2000). *Lernförderlichkeit im schriftsprachlichen Unterricht. Effekte des Klassen- und Förderunterrichts in der Grundschule auf den Lernerfolg. Ergebnisse der Evaluation des Projekts „Lesen und Schreiben für alle“ (PLUS)*. Hamburg: Behörde für Schule, Jugend und Berufsbildung.
- Müller, R. (1990, 1997). *Diagnostischer Rechtschreibtest für 1., 2. und 3. Klassen. Beihefte mit Anleitung und Normentabellen*. Weinheim: Beltz.