

Aspekte diagnostischer Kompetenz von Grundschullehrkräften im Rechtschreibunterricht – Eine empirische Analyse der Urteilsakkuratheit hinsichtlich der Rechtschreibfehler im (nicht-)integrierten Schreiben

Patrick Schreyer  · Anke Reichardt  · Johanna Ingenerf 

Angenommen: 27. Februar 2025 / Online publiziert: 5. März 2025
© The Author(s) 2025, korrigierte Publikation 2025

Zusammenfassung Die diagnostische Kompetenz von Lehrkräften ist ein entscheidender Faktor für die Förderung der Grundbildung. Im Beitrag wird die Urteilsakkuratheit als Teilbereich der diagnostischen Kompetenz untersucht und im Kontext der Rechtschreibleistungen von Drittklässler*innen mittels quantitativer Fehleranalysen analysiert. Diese wird in zwei unterschiedlichen Schreibdimensionen erfasst: im integrierten Schreiben (Texte) und im nicht-integrierten Schreiben (Tests). Die Ergebnisse zeigen, dass Lehrkräfte die Rechtschreibleistungen ihrer Schüler*innen, gemessen an der Fehleranzahl auf 100 Wörter, weitgehend akkurat einschätzen können. Gleichzeitig orientieren sich die Einschätzungen stärker an den Ergebnissen des nicht-integrierten Schreibens, einer primär schulisch relevanten Schreibdimension, und weniger an einer didaktisch perspektivierten und auf Grundbildung bezogenen Fähigkeit.

Schlüsselwörter Rechtschreibung · Diagnostische Kompetenz · Urteilsakkuratheit · Grundschule · Schreibkompetenz

✉ Patrick Schreyer
Institut für Erziehungswissenschaft, Universität Kassel, Kassel, Deutschland
E-Mail: schreyer@uni-kassel.de

Anke Reichardt · Johanna Ingenerf
Institut für Schulpädagogik und Grundschuldidaktik, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Deutschland

Anke Reichardt
E-Mail: anke.reichardt@paedagogik.uni-halle.de

Johanna Ingenerf
E-Mail: johanna.ingenerf@paedagogik.uni-halle.de

Aspects of diagnostic competence of primary school teachers in spelling lessons—An empirical analysis of judgement accuracy with regard to spelling errors in (non-)integrated writing

Abstract The diagnostic competence of teachers is a crucial factor in promoting basic education. This paper examines judgment accuracy as a subdomain of diagnostic competence, focusing on the spelling performance of third-grade students through quantitative mistake analyses. The study captures this competence in two different writing dimensions: integrated writing (texts) and non-integrated writing (tests). The results show that teachers can largely assess their students' spelling performance accurately, measured by the number of mistakes per 100 words. At the same time, the assessments are more strongly aligned with the results of non-integrated writing, a primarily school-relevant writing dimension, rather than a didactically oriented and basic education-related skill.

Keywords Spelling and orthography · Diagnostic competence · Judgement accuracy · Primary school · Writing competence

1 Diagnostische Kompetenzen von Lehrkräften

Diagnostische Kompetenz bezeichnet die Fähigkeit von Lehrkräften, Lernvoraussetzungen, Lernprozesse und Lernleistungen von Schüler*innen möglichst präzise einschätzen und analysieren zu können (Schrader 2013, S. 154). Als Kernelement der Unterrichtsgestaltung wird sie als (mit-)entscheidend für die Unterrichtsqualität betrachtet (Artelt und Gräsel 2009; Charalambous und Praetorius 2020). Sie gilt als integraler Bestandteil der professionellen Handlungskompetenz von Lehrkräften (Baumert und Kunter 2006), wird als ein mehrdimensionales Konstrukt aufgefasst (von Aufschnaiter et al. 2015) und zeigt fachspezifische Charakteristika (Lorenz und Artelt 2009). Die vorliegende Untersuchung im Kontext der INTERFACH-Videostudie konzentriert sich auf die Beurteilung eines Teilaspekts diagnostischer Kompetenz von Lehrkräften, der die Einschätzung der Lernvoraussetzungen ihrer Schüler*innen in dritten Grundschulklassen betrifft. Konkret liegt der Fokus auf der Urteilsakkuratheit der Lehrkräfte in Bezug auf die Rechtschreibfähigkeiten ihrer Schüler*innen, gemessen am Anteil korrekt bzw. falsch geschriebener Grapheme.

Für die vorliegende Untersuchung wird zur Einordnung der diagnostischen Kompetenz das Modell professioneller Handlungskompetenz nach Baumert und Kunter (2006) herangezogen. Diese wird als Teil des Professionswissens verstanden und verschiedenen Bereichen zugeordnet. Demnach spielt diagnostische Kompetenz für das allgemeine pädagogische Wissen eine Rolle (ebd., S. 489), ist aber ebenso für das fachdidaktische Wissen relevant (ebd., S. 495). Die Abgrenzung von fachdidaktischem und allgemeinpädagogischem Wissen geht im Bereich des Diagnostizierens jedoch mit Schwierigkeiten einher (Bremerich-Vos 2019). Dies zeigt sich etwa daran, dass in einigen Bereichen eine „Integration mehrerer Kompetenzfacetten des fachdidaktischen und pädagogischen Wissens“ (Brunner et al. 2011, S. 217) erforderlich ist – wie etwa im Bereich der Leistungsbeurteilung. Das Modell professioneller

Handlungskompetenz verweist somit darauf, dass bei dem Konstrukt diagnostischer Kompetenz von einer mehrdimensionalen Kompetenz mit fachspezifischen Ausprägungen auszugehen ist (von Aufschnaiter et al. 2015; Praetorius et al. 2012).

1.1 Urteilsakkuratheit

Da es keine Hinweise auf eine übergreifende diagnostische Kompetenz gibt, ist ein gezielter Blick auf einzelne fachspezifische Bereiche erforderlich (Lorenz und Artelt 2009). Diese Untersuchung konzentriert sich auf eine domänenspezifische Ausprägung der diagnostischen Kompetenz von Grundschullehrkräften im Bereich der Rechtschreibung. Zur Operationalisierung wird der etablierte Komponentenansatz herangezogen, der die Urteilsakkuratheit¹ untersucht (Schrader und Helmke 1987; Schrader und Praetorius 2018). Lehrkräfteurteile gelten dann als akkurat, wenn sie in der Lage sind, bestimmte Zielkriterien (z. B. die Anzahl der in einem Test zu lösenden Aufgaben oder die Leistung in einem Test) entsprechend ihrer tatsächlichen Ausprägung bei einzelnen Schüler*innen oder der gesamten Klasse korrekt einzuschätzen. Durch den Fokus auf die Akkuratheit soll insbesondere die Qualität der Urteile von Lehrkräften im Sinne der Validität erfasst werden.

Für die Untersuchung der Urteilsakkuratheit haben sich unterschiedliche Maße etabliert (Karst 2012; Schrader und Helmke 1987; Südkamp et al. 2008). Besonders im deutschsprachigen Raum zählen die drei Komponenten der Diagnoseakkuratheit nach Schrader und Helmke (1987) zu den wichtigsten. Diese umfassen die Niveauelemente, die Differenzierungskomponente und die Vergleichs- oder Rangkomponente. Die Niveauelemente bezieht sich auf die Differenz zwischen den Mittelwerten aller Urteile einer Lehrkraft und dem Mittelwert der korrespondierenden Kriteriumswerte innerhalb der Klasse. Die Differenzierungskomponente betrifft die Einschätzung der Streuung der Schüler*innenleistungen innerhalb einer Klasse. Die Vergleichskomponente bestimmt die zutreffende Einschätzung der relativen Leistungsposition einzelner Schüler*innen innerhalb der Klasse und ist „operationalisiert als Produkt-Moment-Korrelation zwischen den vom Lehrer vorhergesagten und den in seiner Klasse tatsächlich erzielten Leistungen“ (ebd., S. 31).

Die Gleichsetzung von diagnostischer Kompetenz mit Urteilsakkuratheit wird jedoch zunehmend kritisch diskutiert, da hinsichtlich ihrer Eignung als empirisches Konstrukt ein Forschungsdesiderat besteht (Artelt und Gräsel 2009; Praetorius et al. 2012). Darüber hinaus stellt die Urteilsakkuratheit eine spezifische Perspektive auf diagnostische Kompetenz dar, da sie lediglich den Teilbereich abbildet, der sich auf die Genauigkeit der Vorhersage eines bestimmten Schüler*innen-Merkmals bezieht. Durch die Entscheidung für ein Messinstrument, mit dem das entsprechende Schüler*innen-Merkmal erhoben wird, kann nur der Teil der diagnostischen Kompetenz abgebildet werden, der die Akkuratheit der Vorhersage *dieses* Testergebnisses be-

¹ Der Begriff der *Urteilsakkuratheit* wird nach Praetorius und Südkamp (2017) gegenüber dem Begriff der *-genauigkeit* bevorzugt, da Genauigkeit eher mit Reliabilität statt, wie hier beabsichtigt, mit Validität assoziiert wird. Zudem entspricht der Begriff der Akkuratheit eher dem englischen Begriff *judgment accuracy*.

trifft. Situationsnahe Faktoren, die Lehrkräfte im Schulalltag in ihre Einschätzungen einbeziehen, bleiben dadurch ausgeblendet (Karst 2017).

1.2 Befunde zur Urteilsakkuratheit

Die Urteilsakkuratheit von Lehrkräften hinsichtlich der Schüler*innenleistungen wurde in mehreren Metaanalysen untersucht. Hoge und Coladarci (1989) berichten, dass Lehrkräfte im Durchschnitt die Leistungen ihrer Schüler*innen moderat einschätzen können, mit einer Median-Korrelation von $r = 0,66$. Schaut man sich die Urteilsübereinstimmungen der drei Studien im Bereich *Language Arts* (LA) an, dann zeigen diese Werte mit $r = 0,70$ – $0,76$ noch höhere Werte als der Median aller Studien. Eine ähnlich moderate Korrelation konnten Südkamp et al. (2012) in ihrer Metaanalyse zu Sprach- und Mathematikleistungen feststellen mit einem Zusammenhang von $r = 0,63$. In Bezug auf LA sind die Befunde jedoch deutlich uneindeutiger; so finden sich in diesem Bereich zum Teil keine ($r = -0,03$) bis hin zu sehr hohen ($r = 0,84$) Korrelationen zwischen den lehrer*innenseitigen Beurteilungen der Sprach- und den tatsächlich gemessenen Schüler*innenleistungen.

Ein genauerer Blick auf die Studien, die in den Metaanalysen aus dem Bereich LA einbezogen wurden, zeigt, dass kaum bereichsspezifische Untersuchungen zur Rechtschreibung existieren. Der Großteil dieser Studien fokussiert auf die Beurteilungsgenauigkeit hinsichtlich der Lesefähigkeiten (z. B. Begeny et al. 2008; Feinberg und Shapiro 2003), des Textverständnisses (z. B. Karing 2009; Lorenz und Artelt 2009) oder des Wortschatzes der Schüler*innen (z. B. Bang et al. 2009; Lorenz und Artelt 2009). Nur wenige Studien berücksichtigen Aspekte der Rechtschreibung, wobei diese *spelling* (Fletcher et al. 2001) oder *writing* (Bailey und Drummond 2006; Gallant 2009) als Teilbereiche übergeordneter Fähigkeiten im Bereich der Sprachfertigkeiten (*literacy skills*) in den LA konzeptualisieren. Alle genannten Studien verwendeten Leistungstests zur Messung der Schüler*innenleistungen und sind somit überwiegend auf der Ebene des nicht-integrierten Schreibens (siehe Abschn. 3.1) anzusiedeln. Eine Studie, die sich verstärkt auf die Ebene des integrierten Schreibens konzentriert, indem sie die Schreibfähigkeiten von Schüler*innen anhand spontan verfasster Aufsätze untersucht, ist die von Burns und Symington (2003). Die Aufsätze wurden mithilfe des Test of Written Language-3 (TOWL) und dem darin enthaltenen *Spontaneous Writing Quotient* (SWQ) eingeschätzt, der neben der kontextbezogenen Sprache und dem Aufbau des Textes auch dessen Rechtschreibung bewertet. In der Studie zeigen sich moderate Korrelationen zwischen den Schüler*innenleistungen und den Einschätzungen der Lehrkräfte von $r = 0,47$.

Obwohl unklar bleibt, ob die Fächerdomäne einen Einfluss auf die Urteilsakkuratheit hat, fanden Hopkins et al. (1985) unterschiedlich hohe Korrelationen in verschiedenen Fächern: Während die Korrelationen in naturwissenschaftlichen und gesellschaftlichen Fächern niedrig waren, zeigten sich in Mathematik und den sprachbezogenen Fächern sowie im Bereich des Lesens eher hohe Korrelationen. Eckert et al. (2006) stellten fest, dass die Übereinstimmung der Einschätzungen bei den Lesefähigkeiten höher war als bei der Mathematikleistung. Demaray und Elliot (1998) berichteten von einer sehr hohen Übereinstimmung in der Vorhersage der Lese- ($r = 0,82$) und der Rechtschreibleistung ($r = 0,86$), während der Zusam-

menhang in Mathematik ebenfalls moderat, aber etwas geringer als in den beiden anderen Fächerdomänen war ($r = 0,66$). Im Gegensatz dazu fanden Südkamp et al. (2012) keine Unterschiede in der Höhe von Urteilen in verschiedenen Fächern und auch keine Unterschiede bei bereichsspezifischen gegenüber generischen Urteilen.

2 Rechtschreibleistungen unter Berücksichtigung des Handlungszusammenhangs

Die vorliegende Untersuchung konzentriert sich auf die Beurteilung der Rechtschreibfähigkeiten der Schüler*innen durch die Lehrkräfte zwischen den mittels eines Einschätzungsbogens erfassten Beurteilungen der Rechtschreibfähigkeiten der Schüler*innen durch die Lehrkräfte sowie den tatsächlichen Rechtschreibleistungen der Schüler*innen. Diese werden durch zwei unterschiedliche Erhebungsverfahren ermittelt, zum einen beim Verfassen eines eigenen Textes zu Bildimpulsen, dem sog. integrierten Schreiben, und zum anderen durch einen standardisierten Test (Hamburger Schreibprobe; May et al. 2013), dem sog. nicht-integrierten Schreiben. Es ist davon auszugehen, dass die jeweils gezeigten Rechtschreibfähigkeiten sich je nach Schreiddimension voneinander unterscheiden (die Anlage unserer Studie ist vergleichbar mit Fay 2010). Im Rahmen der vorliegenden Untersuchung wird daher nicht Rechtschreibkompetenz als Gesamtkonstrukt betrachtet, sondern Rechtschreibleistungen in unterschiedlichen Schreibkontexten, die mittels quantitativer Fehleranalysen bewertet werden. Gemäß der Bildungsstandards (KMK 2022, S. 13) sollen Schüler*innen *Texte* richtig schreiben können. Diese Annahme beruht auf der Prämisse, dass Rechtschreibung „keinen Eigenwert, sondern *dienende Funktion*“ (Abraham et al. 2009, S. 63) hat, um die schriftsprachliche Verständigung zu erleichtern. Kompetente Schreibende besitzen eine Sensibilität für mögliche Fehler und können den Fokus beim Textschreiben bewusst auf die Orthographie lenken (Balhorn und Büchner 2005).

Im Bereich der Rechtschreibung hat sich demgegenüber die Leistungsfeststellung durch nicht-integriertes Schreiben, wie beispielsweise das schulisch nach wie vor übliche Verschriften nach *Diktat*, etabliert und ist weit verbreitet, denn obgleich die Integration der Orthographie in den Bereich des *Textschreibens* für die unterrichtliche Praxis adäquater wäre, so „problematisch ist sie für die testdiagnostische Umsetzung“ (Böhme und Bremerich-Vos 2012, S. 28). Entsprechend dieser Praxis ist wenig über die Kompetenz der Lehrkräfte in Bezug auf die Einschätzung der Leistung beim integrierten Schreiben bekannt (Reichardt 2015). Doch gerade beim Textschreiben, zu dem u. a. das orthographisch richtige Schreiben gehört, handelt es sich um eine Fähigkeit, die Schüler*innen im Unterricht erwerben sollen, da sie zu den zentralen Aspekten schulischer Grundbildung zählt (Köller et al. 2022). Denn das Textschreiben ist für die Schüler*innen in allen Fächern von grundlegender Relevanz für das ungehinderte Lernen, denn es zeigt sich, dass orthographisch bedingte Unterbrechungen beim Textschreiben einen mittleren Einfluss auf die Textqualität und einen geringen, aber signifikanten Einfluss auf Formulierungsflüssigkeit und den Schreibfluss haben (Lemke 2021). Lehrkräfte sind in diesem Sinne einerseits für deren unterrichtliches Handeln und damit für den Lernfortschritt der Schüler*innen

entscheidend, andererseits aber auch für die Vergabe von Noten und damit für Bildungschancen (Brunner et al. 2011).

Orthographisches Schreiben wird in den Bildungsstandards dem Kompetenzbereich Schreiben zugeordnet: Schüler*innen „können Texte [...] in den Kernbereichen der Orthografie richtig schreiben“ (KMK 2022, S. 13). Auch die Konkretisierung der Anforderungen fokussiert die Anwendung durch die Kinder, und zwar v. a. auf Textebene: häufig vorkommende Inhalts- und Funktionswörter richtig schreiben, Beachten der Zeichensetzung, Fehleraufmerksamkeit, Rechtschreibstrategien und Rechtschreibhilfen kennen und anwenden, Überprüfen und Korrigieren von Texten auf orthografische Richtigkeit (ebd.). Die Verortung der Orthographie in den Kontext des Textschreibens unterscheidet sich von einem früheren Verständnis, bei dem Rechtschreibfähigkeit mit dem Korrektheitsgrad beim Diktatschreiben gleichgesetzt wird (Becker 2019) und entsprechend auch von den Konzeptionen der meisten Diagnoseverfahren, welche Rechtschreibfähigkeiten zumeist mittels diktiertter Wörter überprüfen. Entsprechend dieser Disparität müssen die orthographischen Fähigkeiten hinsichtlich des Handlungszusammenhangs differenziert werden, denn während beim freien Schreiben die orthographischen Spezifika des Wortmaterials weitestgehend der tatsächlichen Verteilung in der deutschen Sprache entsprechen, enthält das Wortmaterial eines Diktats oder Tests gebündelt mehrere Fehlerfallen (Fay 2022).

2.1 Orthographie beim integrierten und nicht-integrierten Schreiben

Ludwig (1995) unterscheidet mit dem integrierten und nicht-integrierten Schreiben jeweils Formen der Schreibpraxis. Unter integriertem Schreiben ist die Einbettung des Schreibens als eine Teilhandlung bei der Textproduktion zu verstehen, während nicht-integriertes Schreiben vom Prozess der Textproduktion abgetrennt wird (ebd.). Instrumente, mit denen das Rechtschreibkönnen erfasst wird, lassen sich ebenfalls entlang beider Schreibdimensionen unterscheiden: im integrierten Schreiben, indem frei geschriebene Texte die Grundlage der Analysen bilden; im nicht-integrierten Schreiben, indem Schüler*innen diktierte Wörter und Sätze verschriften. Anzunehmen ist, dass beide Handlungszusammenhänge ermöglichen, Rechtschreibfähigkeiten abzubilden, jedoch als unterschiedliche Facetten eines Konstrukts.

Qualitative Fehleranalysen deuten darauf hin, dass die Schreibdimension, also ob Wörter und Texte frei oder nach Diktat geschrieben werden, einen erheblichen Einfluss auf die Leistungen hat. Untersuchungen von Risel (2011) zeigen, dass die Interpunktion im Diktat durchschnittlich zu 11 % fehlerhaft ist, im Aufsatz hingegen zu ca. 29 %. Diesbezüglich ist ein umgekehrtes Verhältnis bei der Großschreibung zu beobachten: Im Diktat machen Fehler dieser Kategorie ca. 22 % der Fehler aus, im Aufsatz dagegen ca. 11 %. Insbesondere die Ergebnisse der Untersuchung von Fay (2010), durchgeführt in Klasse 4 ($N = 30$ Schüler*innen), bestätigen, dass die Leistungen im integrierten Schreiben und nicht-integrierten Schreiben nicht miteinander vergleichbar sind: Es zeigt sich, dass nur in zwei von zehn erhobenen Fehlerkategorien die Leistungen im Test besser ausfallen als im freien Text. Die Autorin führt dies darauf zurück, dass im Rechtschreibtest eine Verdichtung orthographischer Phänomene vorliegt – auf wenige Wörter verteilt, treten aus testkonzeptionellen Gründen gebündelt Fehlerfallen auf. Demzufolge seien die orthographischen Anforderungen

im Test weitaus höher als im frei verfassten Text. „Derartige Befunde lassen Zweifel aufkommen an der Praxis, Rechtschreibkompetenz zu bestimmen über Leistungen in Rechtschreibtests. Zumal dies weder den curricularen Kompetenzbeschreibungen entspricht noch alltagsweltlichen und beruflichen Kompetenzanforderungen“ (Becker 2019, S. 281).

Für die empirische Modellierung und Diagnostik von Rechtschreibfähigkeiten lässt sich demnach festhalten, dass auf der Grundlage der durchgeführten Erhebung lediglich Aussagen über die Fähigkeiten in der jeweiligen Schreibdimension zulässig sind, in der sie erhoben wurde.

2.2 Instrumente zur Erfassung der Rechtschreibleistungen in unterschiedlichen Schreibdimensionen

Rechtschreibanalyseinstrumente werden eingesetzt zur Individualdiagnostik, Vergleichsmessung, Kompetenzmodellierung oder Unterrichtsentwicklung sowie zur Überprüfung produktiver oder reflexiver Rechtschreibfähigkeiten.

Auf Individualebene kann die Diagnostik des Leistungsstandes von Schüler*innen mittels normierter Tests z. B. Selektionsentscheidungen (etwa zur Legastheniediagnostik) begründen. In Klassenstufe 3 findet neben der HSP 3 (May 2013) auch der Deutsche Rechtschreibtest für das dritte und vierte Schuljahr (DERET 3–4+; Schneider 2018) häufig Anwendung. Ein Vorzug geschlossener, normierter Verfahren liegt insbesondere in der empirisch abgesicherten Evaluierung; individuelle Leistungen werden auf der Grundlage einer Stichprobe bewertet, die als repräsentativ für die Gesamtheit aller Lernenden eines Jahrgangs betrachtet werden kann (vgl. Überblick in Mischo und Wahl 2015).

Frei geschriebene Texte werden aus testökonomischen Gründen selten zur Messung der Rechtschreibfähigkeiten herangezogen. Der zeitliche Aufwand qualitativer Fehleranalysen ist in der unterrichtlichen und empirischen Praxis groß. Entsprechende Instrumente wurden vorrangig für den Einsatz individualdiagnostischer Fragen konzipiert, etwa im Kontext von Förderunterricht oder Legasthenie-Therapie. Zu den bekanntesten orthographischen Analyseinstrumenten und Fehlerschlüsseln zur Auswertung der Rechtschreibleistung in freien Texten zählen die Dortmunder Rechtschreibfehler-Analyse (DoRA; Loeffler und Meyer-Schepers 1992), die Aachener Förderdiagnostische Rechtschreibanalyse (AFRA; Herné und Naumann 2005) und die teilstandardisierte Oldenburger Fehleranalyse (OLFA; Thomé und Thomé 2004, 2023).

Im vorliegenden Projekt wurden für die Auswertung der Rechtschreibleistungen der Schüler*innen der dritten Klassenstufe sowohl die Hamburger Schreibprobe (nicht-integriertes Schreiben) als auch die OLFA (integriertes Schreiben) eingesetzt. Die detaillierten Analysen und individuellen Fähigkeitsprofile der Schüler*innen werden für eine weiterführende Teilstudie benötigt, fließen jedoch nicht in die hier berichtete Untersuchung ein, sondern für die HSP ausschließlich die Graphemtreffer und für die OLFA die Anzahl der Fehler auf 100 Wörter. Dabei handelt es sich jeweils um aussagekräftige Werte für die über alle Fehler *gemittelte* Leistung, unabhängig von ihrer entwicklungsorientierten Kategorisierung. Von Interesse ist für die hier berichtete Untersuchung weniger die qualitative Einordnung der Fehler als

vielmehr die Unterscheidung der Leistungen je nach erhobener Schreibdimension (Test vs. Text).

2.3 Forschungsfragen

Im Rahmen des Projekts sind zwei Forschungsfragen von Interesse:

FF 1 Wie gut können Grundschullehrkräfte im Fach Deutsch mithilfe eines globalen, fähigkeitsunspezifischen Einschätzungsbogens die Rechtschreibfähigkeiten ihrer Schüler*innen einschätzen? Basierend auf den Ergebnissen früherer Studien wird erwartet, dass im Hinblick auf die Niveau- und Differenzierungskomponente jeweils eine Überschätzung der Schüler*innenfähigkeiten bzw. deren Streuung bei den Einschätzungen durch die Lehrkräfte zu beobachten ist. Im Hinblick auf die Rangkomponente wird ein moderater bis hoher Zusammenhang mit den Ergebnissen im integrierten und nicht-integrierten Schreiben erwartet.

FF 2 Inwiefern orientieren sich die Einschätzungen der Lehrkräfte eher an den gemittelten Ergebnissen beim integrierten Schreiben oder am nicht-integrierten Schreiben?

3 Methodisches Vorgehen

3.1 Stichprobe

Diese Untersuchung wurde im Rahmen der INTERFACH-Videostudie durchgeführt, die Teil des DFG-geförderte Graduiertenkolleg „Fachlichkeit und Interaktionspraxis im Grundschulunterricht“ (INTERFACH) an der Martin-Luther-Universität Halle-Wittenberg und der Universität Kassel ist. Die Erhebungen fanden im Zeitraum zwischen Oktober 2022 und April 2024 statt. Das Studiendesign umfasst u. a. Fragebogenerhebungen, Leistungstests und Videoaufzeichnungen. Insgesamt wurden 19 Deutschlehrkräfte aus dritten Grundschulklassen aus Hessen und Sachsen-Anhalt einbezogen, die alle weiblich sind und über eine durchschnittliche Berufserfahrung von 12,2 ($SD = 8,9$) Jahren verfügen. Die von den Lehrkräften eingeschätzte Schüler*innenstichprobe umfasst 297 Schüler*innen, was 85,0 % der Gesamtstichprobe der Studie entspricht, mit einem durchschnittlichen Alter von 8,8 ($SD = 0,64$) Jahren. Die durchschnittliche Schüler*innenanzahl pro Klasse beträgt 15,63 ($SD = 3,90$; min = 6, max = 24). Neben der verbindlichen Durchführung der HSP hatten die Lehrkräfte die Möglichkeit, zusätzlich auch freie Texte zu Bildimpulsen schreiben zu lassen. An dieser Erhebung nahmen 11 Lehrkräften und 158 Schüler*innen teil (45,5 % der INTERFACH-Videostudie Stichprobe).

3.2 Instrumente

3.2.1 Einschätzungsbogen

Um die diagnostischen Fähigkeiten der Lehrkräfte zu erfassen, wurde ihnen im Rahmen der Vorerhebungen zur Videostudie ein globaler, fähigkeitsunspezifischer Einschätzungsbogen mit folgendem Prompt vorgelegt:

„Bitte schätzen Sie die aktuellen Fähigkeiten aller Schüler*innen Ihrer Klasse im Bereich der Rechtschreibung ein. Schätzen Sie hierzu die Schüler*innen jeweils einzeln im Vergleich zu den Fähigkeiten der anderen Schüler*innen in dieser Klasse ein.“

Auf Basis dieses Prompts füllten die Lehrkräfte den Einschätzungsbogen für jeden einzelnen Schüler bzw. jede einzelne Schülerin auf einer fünfstufigen Skala, von 1 = weit unter dem Klassendurchschnitt, 2 = unter dem Klassendurchschnitt, 3 = in der Nähe des Klassendurchschnitts, 4 = über dem Klassendurchschnitt bis 5 = weit über dem Klassendurchschnitt, aus. Das Instrument nimmt entsprechend der Kategorisierung nach McElvany et al. (2009) im Urteilsbereich die Rechtschreibleistungen und auf der Urteilebene die individuellen Schüler*innen einer Klasse in den Blick. Da die einzuschätzenden Fähigkeiten im Prompt nicht näher spezifiziert werden, kann von einer Globaleinschätzung durch die Lehrkräfte ausgegangen werden.

3.2.2 Hamburger Schreibprobe (HSP)

Die Erfassung des nicht-integrierten Schreibens erfolgte über den von May et al. (2013) entwickelten standardisierten Rechtschreibtest HSP. Die Auswertung erfolgte online über das verlagseigene Diagnose- und Auswertungstool (<https://diagnostik.klett.de/>). Im Rahmen dieser Untersuchung werden nur die Ergebnisse der *Graphemtreffer* (T-Werte) aus dem Testbereich der Gesamtwerte für richtige Schreibungen berücksichtigt.

3.2.3 Oldenburger Fehleranalyse (OLFA)

Für die Analyse des integrierten Schreibens wurden die Texte der Schüler*innen mit der von Thomé und Thomé (2023) entwickelten OLFA, jeweils von zwei unabhängigen Beurteiler*innen, ausgewertet. Es zeigten sich moderate bis sehr gute Übereinstimmungswerte bei den Beurteilungen ($ICC = 0,70-0,99$). Bei den in die Analyse einbezogenen Daten handelt es sich um gemittelte Werte dieser beiden Beurteilungen. Berücksichtigt werden für diese Untersuchung nur die Werte zu den *Fehlern auf 100 Wörtern*. Fehler werden dabei im Rahmen der OLFA-Analyse graphemweise analysiert.

3.3 Durchführung

Während der Vorerhebungen bearbeiteten die Schüler*innen zuerst die HSP und anschließend einen Fragebogen, der u. a. demografische Daten erfasste. Der Rechtschreibtest und der Fragebogen wurden jeweils von geschulten Erhebungspersonen

angeleitet. Bei der HSP waren diese angehalten, sich strikt an die Durchführungsvorgaben zu halten. Für eine zusätzliche Standardisierung der Testungen wurden weitere Aspekte (wie etwa die Organisation der konkreten Testdurchführung) in das Erhebungsskript aufgenommen. Parallel zu diesen Erhebungen füllten die Lehrkräfte einen Fragebogen und den für diese Untersuchung zentralen Einschätzungsbogen aus, bei dem die Lehrkräfte angehalten waren, die Fähigkeiten der Schüler*innen im Bereich der Rechtschreibung zu beurteilen.

Neben der verbindlichen Durchführung der HSP hatten die Lehrkräfte die Möglichkeit, zusätzlich auch Geschichten zu Bildimpulsen schreiben zu lassen (Dehn et al. 2011) als Grundlage für die OLFA-Analysen. Für die Gestaltung der Schreibumgebung wurden zwei Motive gewählt, die als Bildimpulse den Schreibprozess steuern sollten. Die Lehrkräfte erhielten ein umfassendes Materialpaket: zwei Din A3-Plakate, Schreibblätter mit Identifikationsnummern für die Schüler*innen und eine Schritt-für-Schritt-Anleitung zur Durchführung der Schreibaufgabe. Die Durchführung sollte insgesamt 45 min umfassen.

3.4 Analyse

Um die Komponenten der Diagnoseakkuratheit nach Helmke und Schrader (1987) zu berechnen, wurden jeweils die Graphemtreffer der HSP und die Fehler auf 100 Wörter aus der OLFA zusammen mit den fünfstufigen Lehrkräfteeinschätzungen der Schüler*innenfähigkeiten betrachtet. Um die Werte der Graphemtreffer und die Werte der Fehler auf 100 Wörter mit denen der Lehrkräfteeinschätzungen vergleichen zu können, wurde eine Min-Max-Skalierung auf einer Skala von 1–5 für beide Testwerte vorgenommen. Der Vorteil dieser Art der Skalierung liegt darin, dass die Verteilung der Daten nicht verändert wird, was besonders relevant ist, um die Streuung der Daten zu untersuchen.

4 Ergebnisse

4.1 Deskriptive Statistiken

Tab. 1 zeigt die deskriptiven Statistiken der Schüler*innenfähigkeiten, gemessen anhand der Graphemtreffer und den Fehlern auf 100 Wörter, sowie der Lehrkräf-

Tab. 1 Deskriptive Statistiken der Schüler*innenfähigkeiten (Graphemtreffer und Fehler auf 100 Wörter) und der Fähigkeitseinschätzungen der Lehrkräfte (Einschätzungsbogen)

	<i>N</i>	<i>M</i>	<i>SD</i>	min	Max
<i>Schüler*innenfähigkeiten</i>					
Graphemtreffer	295	3,03	1,19	1	5
Fehler auf 100 Wörter ^a	157	3,34	1,26	1	5
<i>Lehrkräfteeinschätzungen</i>					
HSP-Stichprobe	19	2,94	1,16	1	5
OLFA-Stichprobe	11	2,86	1,13	1	5

^aWerte sind invertiert

Tab. 2 Kennwerte der Verteilungen der Akkuratheitskomponenten der Lehrkräfteeinschätzungen jeweils für Graphemtreffer und Fehler auf 100 Wörter

	<i>M</i>	<i>SD</i>	Prozentränge				
			0	25	50	75	100
<i>Niveauelemente</i>							
Graphemtreffer	-0,09	0,58	-1,31	-0,37	-0,24	0,44	0,83
Fehler auf 100 Wörter ^a	-0,49	0,38	-1,11	-0,73	-0,50	-0,28	0,29
<i>Differenzierungskomponente</i>							
Graphemtreffer	0,94	0,38	0,25	0,67	0,90	1,21	1,60
Fehler auf 100 Wörter ^a	0,83	0,41	0,12	0,68	0,71	0,97	1,73
<i>Rangkomponente</i>							
Graphemtreffer	0,67	0,13	0,51	0,61	0,67	0,80	0,92
Fehler auf 100 Wörter ^a	0,60	0,18	0,29	0,46	0,60	0,71	0,84

Daten beziehen sich für Graphemtreffer auf $N = 295$ Schüler*innen und $N = 19$ Lehrkräfte und bei den Werten der Fehler auf 100 Wörter auf $N = 157$ Schüler*innen und $N = 11$ Lehrkräfte

Arithmetisches Mittel bei Niveau- und Differenzierungskomponente

Median bei Rangkomponente

^aWerte sind invertiert

teinschätzungen. Die Schüler*innentexte hatten im Durchschnitt eine Länge von 149,4 Wörtern ($SD = 83,9$), und die mittlere Fehleranzahl betrug 39,8 Fehler pro 100 Wörter ($SD = 32,3$).

4.2 Akkuratheit der Lehrkräfteeinschätzungen

In diesem Abschnitt werden die Ergebnisse der Verteilungen der Akkuratheitskomponenten der Lehrkräfteeinschätzungen für die Graphemtreffer und die Fehler auf 100 Wörter dargestellt. Tab. 2 gibt eine Übersicht über die mittleren Werte (M), Standardabweichungen (SD) und Prozentränge der Abweichungsmaß-, Niveau-, Differenzierungs- und Rangkomponenten.

Die Berechnung der *Niveauelemente* erfolgt durch die Ermittlung der Differenz zwischen den Einschätzungen der Lehrkräfte und den tatsächlichen Testwerten. Positive Abweichungen von Null deuten auf eine Überschätzung hin, während negative Abweichungen auf eine Unterschätzung hinweisen. Je näher der ermittelte Wert an Null liegt, desto präziser ist die Einschätzung. Für die Graphemtreffer ergibt sich eine mittlere Niveauelemente von $M = -0,09$ ($SD = 0,58$) und für die Fehler auf 100 Wörter liegt der Wert bei $M = -0,49$ ($SD = 0,38$). Diese Ergebnisse deuten in beiden Fällen auf eine leichte Unterschätzung der Rechtschreibfähigkeiten der Schüler*innen durch die Lehrkräfte hin. Ein gepaarter t-Test ergibt für die Graphemtreffer jedoch keine signifikante Differenz zwischen den Mittelwerten der tatsächlichen Leistungen der Schüler*innen und den Einschätzungen der Lehrkräfte ($t(18) = -0,66$; $p = 0,52$). Im Gegensatz dazu zeigt der Mittelwertsunterschied der Fehler auf 100 Wörter eine signifikante Differenz ($t(10) = -4,25$; $p = 0,002$). Diese Ergebnisse legen nahe, dass es bei den Graphemtreffer keine systematischen Unter- oder Überschätzungen der Fähigkeiten durch die Lehrkräfte gibt. Demgegenüber werden diese bei dem Wert Fehler auf 100 Wörter systematisch unterschätzt.

Die *Differenzierungskomponente* wird durch das Verhältnis der Streuungen der Lehrkräfteeinschätzungen und der Streuung der tatsächlichen Leistungswerte ermittelt. Ein Wert von 1 zeigt eine perfekte Akkuratheit an. Werte über 1 deuten auf eine Überschätzung und niedrigere Werte auf eine Unterschätzung der Streuung der Schüler*innenleistungen hin. Bei der Analyse der Graphemtreffer ergab der t-Test keine signifikante Abweichung der mittleren Streuung $M = 0,94 (SD = 0,38)$ von 1 ($t(18) = -0,73; p = 0,47$). Obwohl die mittlere Streuung für die Werte der Fehler auf 100 Wörter mit $M = 0,83 (SD = 0,41)$ etwas niedriger ist, zeigte auch hier die Analyse keine signifikante Abweichung der mittleren Streuung von 1 ($t(10) = -1,39; p = 0,20$). Zusammenfassend deuten die Ergebnisse darauf hin, dass die Lehrkräfte im Durchschnitt die Streuungen der Schüler*innenleistungen sowohl bei den HSP- als auch bei den OLFA-Werten korrekt einschätzen.

Die Berechnung der *Rangkomponente* ergibt sich aus den Rangkorrelationen zwischen den Einschätzungen der Lehrkräfte und den tatsächlichen Leistungen der Schüler*innen. Diese Korrelation kann Werte von -1 bis 1 annehmen, wobei ein negativer Wert einen negativen Zusammenhang und ein positiver Wert einen positiven Zusammenhang beschreibt. Gemäß Kuckartz et al. (2013) wird die Stärke des Zusammenhangs klassifiziert als kein Zusammenhang ($0,00 \leq r < 0,10$), geringer Zusammenhang ($0,10 \leq r < 0,30$), mittlerer Zusammenhang ($0,30 \leq r < 0,50$), hoher Zusammenhang ($0,50 \leq r < 0,70$) und sehr hoher Zusammenhang ($0,70 \leq r < 1,00$). Die Ergebnisse zeigen für die Werte der Graphemtreffer mit $M_r = 0,67$ einen hohen Zusammenhang zwischen der Einschätzung der Lehrkräfte und den Schüler*innenleistungen. Für die Werte der Fehler auf 100 Wörter zeigt sich mit $M_r = 0,60$ ein leicht geringerer Zusammenhang, der aber ebenfalls als hoch eingeschätzt werden kann. Insgesamt können die Lehrkräfte die Rangordnungen der Schüler*innenleistungen in beiden Schreibdimensionen auf einem hohen Niveau einschätzen.

4.3 Vergleiche zwischen den Maßen von HSP und OLFA

Zur Untersuchung der Unterschiede in der Akkuratheit der Lehrkräfteeinschätzungen für Graphemtreffer und Fehler auf 100 Wörter führten wir eine 2×2 -ANOVA mit wiederholten Messungen durch. Diese Analyse untersucht die Haupteffekte der Urteilsart (Graphemtreffer vs. Fehler auf 100 Wörter) und der Akkuratheitskomponenten (Niveauelemente und Differenzierungskomponente) sowie die Interaktion zwischen diesen Faktoren. Um die Rangkomponenten zu vergleichen, führten wir einen z-Test für die Gleichheit zweier abhängiger Korrelationen durch, wie von Steiger (1980) vorgeschlagen. Abb. 1 visualisiert die deskriptiven Statistiken der Akkuratheitskomponenten und verdeutlicht die Unterschiede und Gemeinsamkeiten zwischen den beiden Bewertungsarten.

Die statistische Analyse ergibt einen signifikanten Haupteffekt der Urteilsart, $F(1, 10) = 79,73; p < 0,001$, was darauf hinweist, dass die Genauigkeit der Urteile zwischen Graphemtreffer und Fehler auf 100 Wörter unterschiedlich ist. Der Interaktionseffekt zwischen Urteilsart und Akkuratheitskomponenten ist nicht signifikant, $F(1, 10) = 1,02; p = 0,34$, d.h. die Art des Urteils hat einen signifikanten Einfluss auf die Genauigkeit der Lehrkräfteeinschätzungen. Der Haupteffekt der Ge-

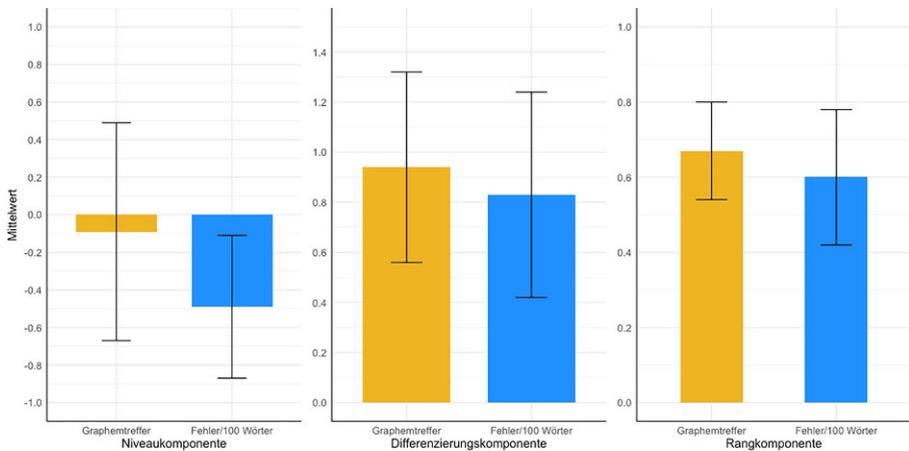


Abb. 1 Deskriptive Statistiken (M, SD) der Akkuratheitskomponenten unterschieden nach der Art der Beurteilung

nauigkeitskomponente war ebenfalls nicht signifikant, $F(1, 10) = 0,59$; $p = 0,46$, was darauf hinweist, dass es keine signifikanten Unterschiede zwischen den Mittelwerten der verschiedenen Komponenten (Niveau- und Differenzierungskomponente) gibt

Die Ergebnisse des z-Test zeigen, dass die Korrelationen zwischen den Rangkomponenten von Graphemtreffer und Fehler auf 100 Wörter nicht signifikant unterschiedlich sind ($z = 0,93$; $p = 0,36$). Dies deutet darauf hin, dass die Rangkomponenten von Graphemtreffer und Fehler auf 100 Wörter ähnlich stark mit den Schüler*innenleistungen korrelieren. Die berechnete Effektstärke ($d = 0,76$) zeigt jedoch, dass der Unterschied in der Rangkomponente im Verhältnis zur Streuung hoch ist. Die fehlende statistische Signifikanz könnte auf die kleine Stichprobengröße ($N = 11$) und die hohe Variabilität innerhalb der Daten zurückzuführen sein.

5 Diskussion

Die Studie untersucht, wie genau globale Lehrkräfteeinschätzungen der Rechtschreibfähigkeiten von Schüler*innen in dritten Klassen sind und auf welche Aspekte diese domänenspezifische Beurteilung abzielt. Die Ergebnisse zeigen, dass es Lehrkräften gut gelingt, die Rechtschreibfähigkeiten ihrer Schüler*innen in Bezug auf deren Fehlerhäufigkeit, gemessen an zwei unterschiedlichen Leistungsmaßen, einzuschätzen. Dies kann als grundlegende Voraussetzung dafür gewertet werden, dass Lehrkräfte dahingehend professionalisiert sind, rechtschreibschwache Kinder in ihrer Klasse zu identifizieren. Gerade in der Grundschule ist dieser Teilaspekt der diagnostischen Kompetenz, auf den die Urteilsakkuratheit hinweist, von besonderer Bedeutung, da Lehrkräfte mit einer hohen Heterogenität in den Lernvoraussetzungen konfrontiert sind (Martschinke et al. 2003).

Zugleich zeigt sich, dass sich die Einschätzungen der Lehrkräfte in höherem Maß an den Leistungen im nicht-integrierten Schreiben orientieren. Die Leistungen im freien Schreiben werden demgegenüber unterschätzt. Das ist dahingehend bemerkenswert, da anzunehmen wäre, dass sie mit den Rechtschreibleistungen ihrer Schüler*innen im integrierten Schreiben tagtäglich und im Rahmen von informellen Lernstandserhebungen wiederkehrend befasst sind. Stattdessen orientieren sie sich möglicherweise stärker an punktuellen standardisierten Erhebungen des nicht-integrierten Schreibens in Form von Leitungstests (Diktat) oder normierten Verfahren (Test). Vor dem Hintergrund, dass standardisierte Verfahren oft mit Fehlerfallen arbeiten, indem u. a. überdurchschnittlich komplexe Wörter mit mehreren orthographischen Besonderheiten aus dem Peripheriebereich der dt. Schrift verschriftet werden müssen, während freie Texte die tatsächliche Häufigkeit orthographischer Abweichungen widerspiegeln, zeigt sich, dass Lehrkräften diese sprachstatistische Gegebenheit möglicherweise nicht bewusst ist. Hier gilt es, Lehrkräfte dafür zu sensibilisieren, dass ihre Schüler*innen bereits über grundlegende Fähigkeiten verfügen, wenn sie eigene Texte mit den ihnen geläufigen Wörtern in weiten Teilen korrekt verschriften. Dabei handelt es sich um eine Fähigkeit, die zur sprachlichen Grundbildung zählt und ihre Relevanz in vielfältigen gesellschaftlichen Kontexten entfaltet. Die Anwendung von Rechtschreibregeln im eigenen Schreibprozess unterstützt Schüler*innen in der Entwicklung einer umfassenden schriftsprachlichen Fähigkeit. Darüber hinaus beschränkt sich die Rechtschreibung nicht auf den Deutschunterricht, sondern wirkt fächerübergreifend.

Die vorliegende Studie deutet darauf hin, dass sich Lehrkräfte in ihren Einschätzungen eher an Erhebungen mit diktierten Wörtern orientieren, die sich weder sprachstatistisch, also im Hinblick auf von Kindern tatsächlich verwendeten Wörtern, noch rechtschreibdidaktisch begründen lassen, sondern von schulischen Testsituationen in Form von Diktaten o. a. überformt sind. Bei dieser Form des nicht-integrierten Schreibens handelt es sich jedoch um eine ausschließlich schulisch auftretende Schreibdimension (Ludwig 1995) mit eigener normativer Kraft, die kaum alltagspraktische Relevanz besitzt und daher kritisch zu hinterfragen ist.

Bei der Interpretation der Ergebnisse dieser Studie sollten mehrere Limitationen berücksichtigt werden. Eine wichtige Limitation besteht darin, dass, anders als von Hoge und Coladarsi (1989) bemerkt, die Akkuratheit der Lehrkräfteeinschätzungen besser beurteilt werden kann, wenn die Einschätzungen der Lehrkräfte das gleiche Maß bzw. die gleiche Einheit besitzen wie die zu bestimmende Fähigkeit. Die so von den Lehrkräften vorgenommene Globaleinschätzung, ist deswegen problematisch, da sie nicht gezielt auf eine spezifische Rechtschreibfähigkeit und somit auch auf keine konkrete der hier untersuchten Schreibstrategien gerichtet war. Folglich prägt das subjektive Verständnis der Lehrkraft das einzuschätzende Schüler*innenmerkmal, was die Objektivität und Vergleichbarkeit der Ergebnisse beeinträchtigen kann.

Zusätzlich ist die Stichprobengröße, insbesondere in Bezug auf die Werte aus den frei geschriebenen Texten, mit insgesamt 11 Klassen als eher niedrig zu beurteilen. Dies zeigt sich besonders in den vergleichenden Analysen zu HSP und OLFA, bei denen die meisten Unterschiede in den Mittelwerten keine Signifikanz aufweisen, trotz teilweise recht hoher Effektstärken. Eine größere Stichprobe könnte

dazu beitragen, die statistische Power zu erhöhen und möglicherweise signifikante Unterschiede aufzuzeigen.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Abraham, U., Beisbart, O., Koß, G., & Marenbach, D. (2009). *Praxis des Deutschunterrichts: Arbeitsfelder, Tätigkeiten, Methoden* (6. Aufl.). Donauwörth: Auer.
- Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(34), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>.
- von Aufschnaiter, C., Cappell, J., Dübbelde, G., Ennemoser, M., Mayer, J., Stiensmeier-Pelster, J., Sträßer, R., & Wolgast, A. (2015). Diagnostische Kompetenz: Theoretische Überlegungen zu einem zentralen Konstrukt der Lehrerbildung. *Zeitschrift für Pädagogik*, 61(5), 738–758. <https://doi.org/10.25656/01:12723>.
- Bailey, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment*, 11(3–4), 149–178.
- Balhorn, H., & Büchner, I. (2005). *Denkwege in die Rechtschreibung: A-o-m 3: Lehrerkommentar*. Hamburg: Verlag für pädagogische Medien.
- Bang, H. J., Suárez-Orozco, C., Pakes, J., & O'Connor, E. (2009). The importance of homework in determining immigrant students' grades in schools in the USA context. *Educational Research*, 51(1), 1–25. <https://doi.org/10.1080/00131880802704624>.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>.
- Becker, T. (2019). Orthographie. In S. Jeuk & J. Settinieri (Hrsg.), *Sprachdiagnostik Deutsch als Zweitsprache: Ein Handbuch* (S. 277–298). Berlin, Boston: De Gruyter Mouton.
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. (2008). Teachers' perceptions of students' reading abilities: an examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43–55. <https://doi.org/10.1037/1045-3830.23.1.43>.
- Böhme, K., & Bremerich-Vos, A. (2012). Beschreibung der im Fach Deutsch untersuchten Kompetenzen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik* (S. 19–33). Münster: Waxmann.
- Bremerich-Vos, A. (2019). Zum Professionswissen von (zukünftigen) Deutschlehrkräften. Empirische Befunde und offene Fragen. *Didaktik Deutsch*, 24(46), 47–63.
- Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 215–234). Münster: Waxmann.
- Burns, M. K., & Symington, T. (2003). A comparison of the spontaneous writing quotient of the test of written language (3rd ed.) and teacher ratings of writing progress. *Assessment for Effective Intervention*, 28(2), 29–34. <https://doi.org/10.1177/073724770302800203>.

- Charalambous, C. Y., & Praetorius, A.-K. (2020). Creating a forum for researching teaching and its quality more synergistically. *Studies in Educational Evaluation*, 67, 100894. <https://doi.org/10.1016/j.stueduc.2020.100894>.
- Dehn, M., Merklinger, D., & Schüler, L. (2011). *Texte und Kontexte: Schreiben als kulturelle Tätigkeit in der Grundschule*. Seelze: Kallmeyer.
- Demaray, M. K., & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: a comparison of actual and predicted performances. *School Psychology Quarterly*, 13(1), 8–24. <https://doi.org/10.1037/h0088969>.
- Eckert, T. L., Dunn, E. K., Coddling, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: an examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools*, 43(3), 247–265. <https://doi.org/10.1002/pits.20147>.
- Fay, J. (2010). *Die Entwicklung der Rechtschreibkompetenz beim Textschreiben: Eine empirische Untersuchung in Klasse 1 bis 4*. Frankfurt a. M.: Peter Lang.
- Fay, J. (2022). Möglichkeiten und Grenzen des Freien Schreibens für den Rechtschreiberwerb. In C. Röber & H. Olfert (Hrsg.), *Schriftsprach- und Orthographieerwerb: Erstlesen, Erstschreiben* (2. Aufl. S. 433–448). Baltmannsweiler: Schneider Verlag Hohengehren. überarb. und erw. Aufl.
- Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, 18(1), 52–65. <https://doi.org/10.1521/scpq.18.1.52.20876>.
- Fletcher, J., Tannock, R., & Bishop, D. V. M. (2001). Utility of brief teacher rating scales to identify children with educational problems: experience with an Australian sample. *Australian Journal of Psychology*, 53(2), 63–71. <https://doi.org/10.1080/00049530108255125>.
- Gallant, D. J. (2009). Predictive validity evidence for an assessment program based on the work sampling system in mathematics and language and literacy. *Early Childhood Research Quarterly*, 24(2), 133–141. <https://doi.org/10.1016/j.ecresq.2009.03.003>.
- Herné, K.-L., & Naumann, C. L. (2005). *Aachener Förderdiagnostische Rechtschreibfehler-Analyse. AFRA*. Aachen: Alfa Zentaurus. Systematische Einführung in die Praxis der Fehleranalyse mit Auswertungshilfen zu insgesamt 33 standardisierten Testverfahren als Kopiervorlagen
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, 59(3), 297–313. <https://doi.org/10.3102/00346543059003297>.
- Hopkins, K. D., George, C. A., & Williams, D. D. (1985). The concurrent validity of standardized achievement tests by content area using teachers' ratings as criteria. *Journal of Educational Measurement*, 22(3), 177–182.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsreich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 197–209. <https://doi.org/10.1024/1010-0652.23.34.197>.
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern*. Münster: Waxmann.
- Karst, K. (2017). Akkurate Urteile – Die Ansätze von Schrader (1989) und McElvany et al. (2009). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: theoretische und methodische Weiterentwicklung* (S. 21–24). New York, Münster: Waxmann.
- KMK (2022). Bildungsstandards für das Fach Deutsch Primarbereich. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-Primarbereich-Deutsch.pdf. Zugriffen: 10. Jan. 2025. Beschluss der Kultusministerkonferenz vom 15.10.2004, i.d.F. vom 23.06.2022.
- Köller, O., Thiel, F., Van Ackeren, I., Anders, Y., Becker-Mrotzek, M., Cress, U., Diehl, C., Kleickmann, T., Lütje-Klose, B., Prediger, S., Seeber, S., Ziegler, B., Kuper, H., Stanat, P., Maaz, K., & Lewalter, D. (2022). *Basale Kompetenzen vermitteln – Bildungschancen sichern*. Bonn: SWK. Perspektiven für die Grundschule. Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK)
- Kuckartz, U., Rädiker, S., Ebert, T., & Schehl, J. (2013). *Statistik: Eine verständliche Einführung*. Wiesbaden: VS.
- Lemke, V. (2021). *Rechtschreiben beim Textschreiben. Zum Einfluss der Orthographie auf die Schreibflüssigkeit und die Qualität von Schülertexten*. Münster: Waxmann.
- Loeffler, I., & Meyer-Schepers, U. (1992). *DoRA: Dortmunder Rechtschreibfehler-Analyse zur Ermittlung des Schriftsprachstatus rechtschreibschwacher Schüler*. Dortmund: ILT.
- Lorenz, C., & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 211–222.

- Ludwig, O. (1995). Integriertes und nicht-integriertes Schreiben. Zu einer Theorie des Schreibens: Eine Skizze. In J. Baumann & R. Weingarten (Hrsg.), *Schreiben: Prozesse, Prozeduren und Produkte* (S. 273–287). Wiesbaden: VS. https://doi.org/10.1007/978-3-322-97050-3_13.
- Martschinke, S., Kammermeyer, G., & Pietsch, M. (2003). Jedes Kind ist anders. Jede Klasse ist anders. Ergebnisse aus dem KILIA-Projekt zur Heterogenität im Anfangsunterricht. *Zeitschrift für Erziehungswissenschaft*, 6, 257–257.
- May, P., Malitzky, V., & Vieluf, U. (2013). *HSP 1-10: Hamburger Schreibprobe: Manual/Handbuch: Diagnose orthografischer Kompetenz: zur Erfassung der grundlegenden Rechtschreibstrategien*. Stuttgart: Verlag für pädagogische Medien.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Hartz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 223–235.
- Mischo, C., & Wahl, S. (2015). Standardisierte Testverfahren zur Messung von Lese- und Rechtschreibfertigkeiten. In C. Röber & H. Olfert (Hrsg.), *Schriftsprach- und Orthographieerwerb: Erstlesen, Ersts Schreiben* (S. 327–346). Baltmannsweiler: Schneider Verlag Hohengehren.
- Praetorius, A.-K., & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: theoretische und methodische Weiterentwicklungen* (S. 13–18). Münster: Waxmann.
- Praetorius, A.-K., Lipowsky, F., & Karst, K. (2012). Diagnostische Kompetenz von Lehrkräften: Aktueller Forschungsstand, unterrichtspraktische Umsetzbarkeit und Bedeutung für den Unterricht. In R. Lazarides & A. Ittel (Hrsg.), *Differenzierung im mathematisch-naturwissenschaftlichen Unterricht* (S. 115–146). Bad Heilbrunn: Klinkhardt.
- Reichardt, A. (2015). *Rechtschreibung im Textraum: Modellierungen der Schreibkompetenz in der Grundschule*. Duisburg: Gilles & Francke.
- Risel, H. (2011). *Arbeitsbuch Rechtschreibdidaktik*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Schneider, W. (2018). Diagnostik von Schriftsprachkompetenz bei Kindern und Jugendlichen. *Lernen und Lernstörung*, 7(4), 209–214.
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerbildung*, 31(2), 154–165.
- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern. Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27–52.
- Schrader, F.-W., & Praetorius, A.-K. (2018). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost, J. R. Sparfeldt & S. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 92–98). Weinheim: Beltz.
- Siekmann, K., & Thomé, G. (2018). *Der orthographische Fehler. Grundzüge der orthographischen Fehlerforschung und aktuelle Entwicklungen* (2. Aufl.). Oldenburg: isb, Institut für sprachliche Bildung – Verlag, aktualisierte Auflage
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245–251. <https://doi.org/10.1037/0033-2909.87.2.245>.
- Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der Simulierte Klassenraum. Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22(3–4), 261–276. <https://doi.org/10.1024/1010-0652.22.34.261>.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: a meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>.
- Thomé, G., & Thomé, D. (2004). *Oldenburger Fehleranalyse. OLFA. (Version 2.0). Instrument und Handbuch zur Ermittlung der orthographischen Kompetenz aus freien Texten ab Klasse 3 und zur Qualitätssicherung von Fördermaßnahmen*. Oldenburg: Igel-Verl. Wiss.
- Thomé, G., & Thomé, D. (2023). *OLFA 3–9+: Oldenburger Fehleranalyse für die Klassen 3–9 und darüber. Instrument und Handbuch zur Ermittlung der orthographischen Kompetenz und Leistung aus freien Texten für die Entwicklung effektiver Fördermaßnahmen* (7. Aufl.). Oldenburg: isb, überarbeitete Auflage

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.