

Beim Korrigieren diagnostizieren

Sprachdiagnostische Kompetenz von Lehrkräften der Sekundarstufe II im Forschungsprojekt DIEGO

Florian Seemann^{1,*}

¹ Goethe-Universität Frankfurt am Main

* Kontakt: Institut für Psycholinguistik und Didaktik der deutschen Sprache
(Arbeitseinheit Deutsch als Zweitsprache), Norbert-Wollheim-Platz 1,

60323 Frankfurt am Main

E-Mail: seemann@em.uni-frankfurt.de

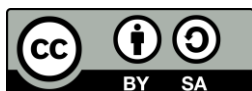
Zusammenfassung: Sprachbildung gilt als durchgängige Querschnittsaufgabe aller Unterrichtsfächer. Um Lernende angemessen und zielgerichtet entsprechend einem individuellen Sprachförderbedarf unterstützen zu können, müssen Lehrkräfte die sprachlichen Kompetenzen der Lernenden akkurat und differenziert diagnostizieren können. Dies ist die Grundbedingung für die Ableitung von Förderbedarf und Förderschwerpunkten.

In dem vorliegenden Beitrag wird eine Studie zur sprachdiagnostischen Kompetenz von Deutschlehrkräften der Sekundarstufe II vorgestellt. Der Fokus liegt hier auf der sprachdiagnostischen Kompetenz im Bereich der Textproduktion.

51 Lehrkräfte wurden im Rahmen dieser Studie des Forschungsprojektes DIEGO gebeten, gemäß ihrer gängigen Korrekturpraxis zwei Schüler*innentexte zu bewerten und aus ihnen Förderschwerpunkte abzuleiten.

Die Ergebnisse zeigen große Unterschiede und Streuungen sowie eine mehrheitliche defizitäre Akkuratheit der Urteile von Lehrkräften. In diesem Beitrag wird die Akkuratheit des Lehrkrafturteils anhand verschiedener Facetten (Fehlerzahl, Korrektur im Text, Förderschwerpunkte) mit einem Expert*innenmaßstab gemessen. Auf Grundlage der Ergebnisse werden Implikationen für eine fächerübergreifende Lehrkräftebildung aufgezeigt.

Schlagerwörter: Lehrerurteil; Sprachdiagnostik; Leistungsbeurteilung; Aufsatzbeurteilung; Diagnose; Kompetenz; Genauigkeit; Urteil; Gymnasiale Oberstufe; Korrektur; Fehler



1 Einleitung

Spätestens seit den Ergebnissen der PISA-Studie am Anfang dieses Jahrtausends ist die Schlüsselfunktion der Kompetenzen in der (Bildungs-)Sprache Deutsch für Bildungserfolg in der Bundesrepublik Deutschland evident (vgl. Baumert & Schümer, 2001). Die Kompetenz, Texte verfassen zu können, ist „in der literalisierten Gesellschaft unverzichtbar“ (Geist & Krafft, 2019, S. 96). Gerade in der Sekundarstufe sind von Lernenden produzierte zusammenhängende Texte essenziell, um in Klassenarbeiten und Klausuren oder bei Abschlussprüfungen wie dem Abitur zu reüssieren.

Damit Lernende hier erfolgreich sein können, wird Sprachbildung als durchgängige Querschnittsaufgabe betrachtet. Lernende sollen in allen Fächern in ihren (bildungs-)sprachlichen Fähigkeiten gefördert werden. Ferner soll Sprachbildung und Sprachförderung in allen Etappen des Bildungswegs stattfinden.

Um Lernende sprachlich zu fördern, benötigen Lehrkräfte sprachdiagnostische Kompetenz. Eine akkurate und differenzierte Sprachdiagnose ist die „Grundlage für individuelle Förderung“ (Michalak, 2012, S. 57) sowie Voraussetzung zur Herstellung der „Passung des Unterrichts“ (Terhart, 2000, S. 15).

Ausgehend von diesem Zusammenhang werden in diesem Beitrag ausgewählte Daten aus dem Forschungsprojekt („Diagnostische Expertise von Deutschlehrkräften der gymnasialen Oberstufe“ (DIEGO) vorgestellt. Lehrkräfte der Sekundarstufe II stehen im Zentrum dieser Studie, da sie erstens im Vergleich zu anderen Abschnitten des Bildungssystems weniger im Forschungsfokus sind. Zweitens unterrichten sie an einer Scharnierstelle zwischen Schule und Ausbildung, Studium und Beruf. Ihre Urteile entscheiden nicht nur über die Qualifizierung und Förderung von Lernenden, sondern auch über deren Bildungskarriere. In DIEGO wurde die sprachdiagnostische Kompetenz von Deutschlehrkräften der gymnasialen Oberstufe mittels eines praxisnahen Studiendesigns untersucht. Es wird die Homogenität sowie die Akkuratheit von Lehrkrafturteilen bei Schülertexten betrachtet. Dies geschieht anhand von Fehlerzahlen, der exemplarischen Analyse der Korrektur im Schülertext sowie der Bestimmung von Förderschwerpunkten. Abschließend werden aus den Ergebnissen Implikationen für Forschung und die fächerübergreifende Lehrkräftebildung abgeleitet. Die vorgestellten Daten beziehen sich exemplarisch auf Deutschlehrkräfte, ein fachübergreifender Bezug wird im Beitrag ebenfalls anvisiert.

2 Theoretischer Hintergrund

Mittlerweile ist es fachdidaktischer Konsens, dass die Förderung von sprachlichen Kompetenzen Aufgabe aller Fächer ist, durchgängig stattfinden soll und Fachunterricht sprachsensibel sein sollte (vgl. z.B. Gogolin et al., 2020; Leisen, 2020; Peuschel & Burkard, 2019). Zentral ist dabei die Rolle von Textproduktion für Bildungserfolg (vgl. Geist & Krafft, 2019, S. 96).

2.1 Textproduktion und Bildungserfolg

Sprachbildung und Sprachförderung werden zumeist mit dem Deutsch- bzw. dem Fremdsprachenunterricht assoziiert, da Sprache und sprachliche Kompetenzen hier genuine Gegenstände des Faches sind:

„Während im Deutschunterricht, im DaZ-Unterricht, im Fremdsprachenunterricht und in Maßnahmen der Sprachförderung der Fokus auf sprachlichen Strukturen und formaler Korrektheit der mündlichen und schriftlichen Texte und der Redebeiträge von Schüler*innen liegt, geht es im Fachunterricht der Sekundarstufen nur sehr selten um eine Arbeit an Grammatik und Orthographie“ (Peuschel & Burkard, 2019, S. 94).

Kompetenzen zum Verfassen von Texten sind jedoch fächerübergreifend notwendig, etwa „um die in vielen Fächern (nicht nur im Fach Deutsch, auch in Geschichte, Literatur, Pädagogik, Philosophie, Religion, Sozialwissenschaften usw.) in der Oberstufe immer wieder zu schreibenden Inhaltsangaben anfertigen zu können“ (Scholten-Akoun, 2018, S. 237). Peuschel & Burkard (2019, S. 94) berichten von Erkenntnissen zum naturwissenschaftlichen Unterricht, „dass die Form der sprachlichen Realisierung einer fachlich richtigen Äußerung Einfluss auf deren fachliche Bewertung haben kann, wobei fachlich richtige Antworten mit sprachlichen Fehlern tendenziell auch als fachlich schlechter bewertet werden“. Zudem zeigen Studien, dass sich „schwache Orthografie und fehlerhafte Grammatik über Gebühr negativ auf Beurteilungen auswirken“ (Baurmann, 2014, S. 265).

Beurteilt werden oftmals Aufsätze als diejenige „Prüfungsform, die von der Grundschule bis zum Abitur eine wesentliche Rolle spielt und in den Vorschriften der Länder als wesentlicher Bestandteil der Zeugnisnote vorgeschrieben ist“ (Wiprächtiger-Gepfert, 2011, S. 175). Aufgrund der Bedeutung von Klausuren und schriftlichen Abiturprüfungen determinieren Defizite in der Textproduktion den Bildungserfolg in der gymnasialen Oberstufe und den damit verbundenen Zugang zur Hochschulbildung.

Dies wird offenbar an der direkten Auswirkung der sprachlichen Leistung auf Noten. In der hessischen Oberstufenverordnung (OAVO) z.B. erfolgt bei sprachlichen Defiziten, die anhand eines Fehlerindex gemessen werden, ein Abzug von bis zu zwei Notenpunkten (vgl. HKMB, 2025, Anlage 9b). Hierfür gewertete Fehler betreffen die beiden Basisdimensionen Sprachrichtigkeit (Kriterien Orthografie und Grammatikalität) und Sprachangemessenheit (Kriterium Wortwahl) aus dem Basiskatalog von Becker-Mrotzek & Böttcher (2014, S. 128f).

Trotz inhaltlich guter Leistungen können durch mangelnde Sprachkompetenz deutliche Nachteile entstehen, die sich wiederum kumulativ negativ auf alle Fächer auswirken können (Baumert & Schümer, 2001; Seipp, 2012). Die Novellierung der hessischen OAVO im Jahr 2025 führt im Fach Deutsch zur obligatorischen Berücksichtigung der sprachlichen Darstellungsleistung von 30 Prozent bei der Bewertung von Klausuren und Abituarbeiten ab dem Landesabitur 2027. Zusätzlich ist weiterhin ein Abzug aufgrund eines zu hohen Fehlerindex möglich.

Mehrere Studien weisen faktischen Sprachförderbedarf beim Verfassen von Texten während und nach der Sekundarstufe II nach (z.B. Cakir, 2009; Frerichs & Kaleta, 2015; Larew, 2014; Überblick bei Geist & Krafft, 2019, S. 100ff.). Bezugnehmend auf die Feststellung Pohls, „Abiturienten [kommen] mit Schreibfähigkeiten an die Universitäten [...], die für die besonderen Anforderungen wissenschaftlichen Schreibens in der Regel nicht ausreichen“ (2011, S. 3), verweisen Niederdorfer et al. (2018) auf wissenschaftliche Erkenntnisse über mangelnde Fähigkeiten beim Verfassen von im Studium benötigten Texten. Scholten-Akoun (2018) erkennt bei Lehramtsstudierenden ebenfalls Sprachförderbedarf: „Die curricular mit dem mittleren Schulabschluss, spätestens aber mit dem Abitur, vorausgesetzten Kompetenzen erwiesen sich als keineswegs vollständig gesichert“. (Scholten-Akoun, 2018, S. 242). Diese Ergebnisse unterstreichen den Bedarf bzw. die Notwendigkeit von durchgängiger Sprachbildung und Sprachförderung, sodass diese nicht „tatsächlich zum Luxusproblem – für alle Schüler, die ihre sprachlichen Probleme nur durch kostspieligen Nachhilfeunterricht beheben können“ (Strecker, 2010, S. 269) werden.

2.2 Diagnostisches Potenzial von Schüler*innentexten

Damit eine differenzierte und auf die individuelle „Textproduktionskompetenz“ (Geist & Krafft, 2019, S. 100) ausgerichtete Förderung auch in der Sekundarstufe II stattfinden kann, bedarf es einer Sprachdiagnostik, die sich aus Schüler*innentexten aufgrund ihres hohen sprachdiagnostischen Potenzials speisen kann, wenn diese von Lernenden selbstständig produziert werden. Diktate hingegen sind wegen lebens- und arbeitsweltlicher

sowie abiturprüfungsrelevanter Bedeutungslosigkeit keine brauchbaren Texte. Es herrscht fachdidaktische Einigkeit darüber, „[d]ass Diktate nur in sehr eingeschränkter Form geeignet sind, um über die Rechtschreibfähigkeiten von SuS (gleich welchen sprachlichen Hintergrunds) Auskunft zu geben“ (Geist & Krafft, 2019, S. 92). Zielführender sind eigene Textprodukte, die ohne Prüfungsdruck in „Lernsituationen“ (Peschel, 2024, S. 40) entstehen, da sie auch über den Wortschatz, die Verwendung passender Konjunktionen bei Nebensätzen, die Textstrukturierung oder die Lesendenführung Aufschluss geben oder auch Spuren des Planens und Überarbeitens beinhalten (vgl. Becker-Mrotzek & Böttcher, 2014, S. 129ff.). Im Rahmen einer Hausaufgabe verfasste Texte sind vor dem Hintergrund aktueller Entwicklungen Anwendungen der textgenerierenden künstlichen Intelligenz nur bedingt geeignet, da man hier eher indirekt die Prompting Skills der Lernenden diagnostiziert und die Gefahr besteht, dass die KI hier zum Ghostwriter wird (vgl. Steinhoff, 2025).

Texte, die in „Prüfungssituationen“ (Peschel, 2024, S. 40) entstehen, sind trotz des Prüfungsdrucks eine wichtige sprachdiagnostische Gelegenheit (vgl. Achilles, 2008, S. 36). Durch einen zeitlichen Rahmen und den Ausschluss digitaler Hilfsmittel kann ein ohnehin vorhandener Moment der gymnasialen Oberstufe zur Leistungsbewertung gleichzeitig zur Sprachdiagnostik genutzt werden, denn man kann eindeutig „Korrekturhandeln als wesentliche Tätigkeit des Lehrers“ (Becker-Mrotzek & Böttcher, 2014, S. 133) betrachten und die Textbeurteilung besitzt „eine zentrale Rolle“ (Peschel, 2024, S. 39) in vielen Fächern neben Deutsch.

In Textklausuren diagnostizieren die Lehrkräfte bei ihrem Korrekturhandeln implizit und explizit. Sie melden dies den Lernenden auf unterschiedliche Weise zurück, auch wenn dies ggf. nicht im Fokus der Lehrkräfte ist, sodass sich folgende sprachdiagnostischen Facetten in dieser „schülerspezifische[n] Situation“ (Karst, 2012) zeigen:

- Fehlerzahl und Fehlerindex (inkl. Punktabzug & Kategorisierung wie „kein Punktabzug“, „geringer Punktabzug“, „hoher Punktabzug“);
- Klassifizierung als Fehler (=Abweichung von sprachlicher Norm) durch Anstreichen, Versehen mit Korrekturzeichen und Einbezug in den Fehlerindex;
- Markierung (z.B. Schlangenlinie) als Abweichung von sprachlicher Norm, aber nicht als für den Fehlerindex relevanter Fehler;
- Korrekturzeichen mit unterschiedlichem Differenzierungsgrad (vgl. LP 1/LP 52 in Kap. 4.3);
- Kommentare am Rand und unter dem Text;
- prüfende Bewertung (vgl. Becker-Mrotzek & Böttcher, 2014, S. 123) der sprachlichen Darstellungsleistung (z.B. (fach-)sprachliche Qualität, Struktur etc.).

Um diese vielen sprachdiagnostischen Momente in Textklausuren der gymnasialen Oberstufe zu nutzen und akkurate Urteile zu treffen, bedarf es einer fundierten sprachdiagnostischen Kompetenz, denn „durch das korrekte Beurteilen von Schreibleistungen sind Lehrkräfte in der Lage, den Unterricht sowie Fördermaßnahmen angemessen an die Leistungen der Schüler*innen anzupassen“ (Strahl, 2024, S. 222).

2.3 Sprachdiagnostische Kompetenz von Lehrkräften

Spätestens seit der PISA-Studie um die Jahrtausendwende gilt diagnostische Kompetenz als „wichtige Facette professioneller Kompetenz“ (Wagner, 2025, S. 100) von Lehrkräften. Diagnostische Kompetenz ist

„ein Bündel von Fähigkeiten, um den Kenntnisstand, die Lernfortschritte und die Leistungsprobleme der einzelnen Schüler sowie die Schwierigkeiten verschiedener Lernaufgaben im Unterricht fortlaufend beurteilen zu können, sodass das didaktische Handeln auf diagnostische Einsichten aufgebaut werden kann“ (Weinert, 2000, S. 14).

In der Forschungspraxis wird diagnostische Kompetenz oftmals mit Urteilsgenauigkeit gleichgesetzt (vgl. Schrader, 2012, S. 43). Präziser ist der Begriff der Urteilsakkuratheit als *judgment accuracy* (vgl. Praetorius & Südkamp, 2017, in Bezug auf Spinath, 2005).

Forschung zu diagnostischer Kompetenz fokussiert primär Grundschulen (vgl. Karing, 2009) und nur äußerst selten die Sekundarstufe II. Vergleichsmaßstab zur Untersuchung der Urteilsgenauigkeit sind die Selbsteinschätzung von Lernenden (vgl. z.B. Hosenfeld et al., 2002; Spinath, 2005) oder die Schüler*innenleistung, die mittels eines Tests oder eines Diagnoseinstrumentes gemessen wurde (z.B. Geist, 2014; Karing, 2009; PISA-Studien).

Durch die Ergebnisse internationaler Vergleichsstudien wurde eine defizitäre diagnostische Kompetenz von Lehrkräften konstatiert (vgl. Helmke, 2012). Prenzel et al. (2007, S. 6) sprechen diesbezüglich von einer diagnostischen Schwäche der Lehrkräfte. Bräuer & Winkler (2012) berichten in ihrem Überblick, dass fachübergreifend eine gering ausgeprägte Urteilsgenauigkeit bei Lehrkräften zu konstatieren sei. Dabei ist jedoch eine große Streuung zwischen den einzelnen Lehrenden zu beobachten (Karing, 2009; Lorenz & Artelt, 2009).

Studien zur Bewertung von Schüler*innenleistungen zeigen Unterschiede bei der Notengebung (z.B. Ingenkamp, 1995). Laut der MARKUS-Studie seien Noten zwischen Klassen nicht vergleichbar (Helmke & Jäger, 2002). Mit Verweis auf die Studien von Birkel (2003) sowie Birkel & Birkel (2002) konstatiert Strahl (2024, S. 227), dass es „häufig [...] an der Übereinstimmung bei der Beurteilung der gleichen Texte durch verschiedene Lehrkräfte“ mangelt.

Es existieren „noch viele Forschungsdesiderata im Bereich der Textbeurteilung und Schreibförderung in den Unterrichtsfächern“ (Petersen et al., 2024, S. 9), obwohl dies eine fächerübergreifend relevante und quantitativ häufig betriebene Tätigkeit von Lehrkräften ist. Beispielsweise sind „Untersuchungen zu Fragen der Beurteilung von Texten als Ausgangspunkt für Leistungsbewertung und Förderung in allen Unterrichtsfächern sind [...] eher rar“ (Petersen et al., 2024, S. 9).

Diese Studie greift dieses Desiderat auf, indem die sprachdiagnostische Kompetenz von Lehrkräften im Rahmen der Beurteilung von Texten untersucht wird.

Aufgrund der Kritik an der Reduzierung diagnostischer Kompetenz auf die Akkuratheit von Urteilen (z.B. Jürgens, 2012; Schrader, 2012) schlägt Helmke (2012, S. 119) mit diagnostischer Expertise ein um das Wissen über Urteilstendenzen und -fehler sowie methodische Fähigkeiten ein erweitertes Konzept vor. Davon ausgehend werden unter Bezugnahme auf Arbeiten von Baumert & Kunter (2006), Helmke (2012), Jäger (2007) und Müller (2014) als Elemente der sprachdiagnostischen Kompetenz Fachwissen und fachdidaktisches Wissen, Verfügbarkeit diagnostischer Methoden, Wissen über Urteilstendenzen/-fehler, Urteilsakkuratheit sowie aus Ergebnissen abgeleitete Förderimplikationen gefasst. In Anlehnung an Geist (2014), die ein umfassenderes Konzept diagnostischer Kompetenz für Sprachförderkräfte in hessischen Vorlaufkursen in den Blick nimmt, wird in diesem Beitrag der Begriff der sprachdiagnostischen Kompetenz genutzt. Urteilsakkuratheit wird Südkamp et al. (2012) folgend als „Übereinstimmung zwischen dem Lehrkrafturteil und der Leistung der Schülerin oder des Schülers bestimmt“ (Strahl, 2024, S. 223) definiert.

Zur Untersuchung der sprachdiagnostischen Kompetenz wurden Lehrkräften Schüler*innentexte vorgelegt. Die Lehrkräfte wurden gebeten, die Texte unter Einbezug gängiger Korrekturzeichen zu korrigieren und zu bewerten. Zudem wurden sie gebeten, Förderschwerpunkte abzuleiten.

Folgende Fragen stehen im Mittelpunkt der Studie:

- 1) Wie homogen sind Urteile der Lehrkräfte hinsichtlich der sprachlichen Kompetenzen von Lernenden in der Sekundarstufe II?
- 2) Wie akkurat sind diese Urteile?
- 3) Leiten Lehrkräfte korrekte Förderschwerpunkte auf Grundlage ihrer Urteile ab?

3 Methodisches Vorgehen

3.1 Stichprobe

Die Datenerhebung fand im Rahmen des Forschungsprojets DIEGO¹ statt. Die Population der Studie besteht aus hessischen Deutschlehrkräften der gymnasialen Oberstufe. Um nicht zu viele die Schulform betreffende Faktoren in die Untersuchung einfließen zu lassen (vgl. Karing, 2009), beschränkt sich die Studie auf Lehrkräfte allgemeinbildender öffentlicher Schulen.

An der Studie nahmen 51 Deutschlehrkräfte teil. Diese arbeiteten zum Zeitpunkt der Erhebung an allgemeinbildenden Gymnasien ($N = 19$), an Gesamtschulen mit Oberstufe ($N = 9$) oder an Oberstufenschulen inkl. Abendgymnasien ($N = 22$), einmal erfolgte keine Angabe. 35 der 51 Teilnehmenden sind weiblich (= 68,6 %). Die Lehrpersonen sind zum Erhebungszeitpunkt zwischen 27 und 63 Jahren alt ($M = 43$; $SD = 10,59$). Ihre Berufserfahrung beträgt 2,5 bis 37 Jahre. ($M = 13,52$; $SD = 10,54$).

3.2 Untersuchungsdesign

3.2.1 Schüler*innentexte zur Beurteilung

Das Untersuchungsdesign basiert auf zwei realen im Rahmen einer Klausur entstandenen Schüler*innentexten, die sich inhaltlich mit Georg Büchners Dramenfragment „Woyzeck“ und Theodor Fontanes Roman „Irrungen, Wirrungen“ beschäftigen. Es kommt daher der Forderung nach größerer Praxisnähe deutschdidaktischer Forschung nach (z.B. Bräuer & Winkler, 2012). Die Texte stammen von Lernenden mit einem diagnostizierten Sprachförderbedarf.²

Aufgrund der regelmäßigen Korrekturtätigkeit von (Deutsch-) Lehrkräften mussten die Proband*innen kein ihnen unbekanntes Forschungs- oder Bewertungsinstrument kennenlernen, sodass dies keinen Einfluss auf das Lehrkräfteurteil hatte (vgl. Strahl, 2024).

Zusammen mit der Instruktion erhielten die Teilnehmenden Informationen über die Lernenden und die Texte (Gymnasium, 13. Klasse, Grundkurs Deutsch, Aufgabenstellung). Die begrenzte Mitteilung von Informationen über die Lernenden diente dazu, die Merkmale der Lernenden auf ihre Leistung (vgl. Südkamp et al., 2012) zu reduzieren und den Einfluss von *beliefs* (vgl. Hachfeld et al., 2012) auf das Lehrkräfteurteil möglichst auszuschließen. Ferner wurde bei der Kennzeichnung der Unterlagen mit der Bezeichnung „SchülerIn“ eine geschlechtsneutrale Begrifflichkeit gewählt, da trotz uneindeutiger Befunde zum Einfluss des Geschlechts (vgl. Wagner, 2025, S. 103) zumindest Hinweise in einer Studie von Lorenz (2011) dahingehend existieren, „dass Mädchen konsequent höhere Werte im Wortschatz [...] zugesprochen werden“ (Wagner, 2025, S. 117). Um Auswirkungen der Handschrift auf das Urteil auszuschließen (vgl. Baurmann, 2014, S. 265), wurden die Texte vom Forschenden handschriftlich abgeschrieben. Der Text von „SchülerIn 1“ umfasst 1822 Wörter, jener von „SchülerIn 2“ besitzt 807 Wörter.

¹ Das Forschungsprojekt DIEGO (PI: Prof. Dr. Petra Schulz, Goethe-Universität Frankfurt am Main) wurde von der Gemeinnützigen Hertie-Stiftung und der Akademie für Bildungsforschung und Lehrerbildung an der Goethe-Universität Frankfurt unterstützt.

² Der diagnostizierte Sprachförderbedarf der Lernenden ist dadurch legitimiert, dass sie im Sprachförderprojekt FJM (später: „Vorsprung“) an der Goethe-Universität Frankfurt am Main teilnahmen und der Sprachförderbedarf seitens der universitären Expert*innen diagnostiziert worden war. Im Rahmen des Projektes förderten speziell ausgebildete Studierende Schülerinnen und Schüler der Sekundarstufe II mit Blick auf den Ausbau bildungssprachlicher Fähigkeiten (vgl. Frerichs & Kaleta, 2015; Strecker, 2010).

3.2.2 Expert*innenurteil als Goldstandard

Um die Urteilsakkuratheit der Lehrkräfte bestimmen zu können, wurde ein Expert*innenurteil (= Goldstandard) eingeholt (vgl. z. B. Lohaus, 2008). Hierzu wurden die Schüler*innentexte fünf Expert*innen vorgelegt. Dabei handelte es sich um Deutschlehrkräfte, die als Auszubildende in der zweiten Phase der Lehrkräftebildung tätig waren.

Werden Expert*innen in Studien herangezogen, wird meist eine ungerade Zahl Personen gewählt. Wenn sich deren Mehrheit dann für ein Urteil entscheidet, dient dies als Maßstab. Im Wissen über ihr Expert*innendasein erhielten sie die gleichen Aufgaben (vgl. Kap. 3.3), d.h. sie wurden gebeten, die Texte zu korrigieren und mögliche Fehler Schwerpunkte zu bestimmen. Sie bearbeiteten die Unterlagen individuell und fungierten somit als unabhängige Rater*innen. Bei Analysen im Text wie etwa bei der Klassifizierung als Fehler durch die Studienteilnehmenden oder der Klassifizierung der Fehler wurde als Maßstab angesetzt, dass mindestens drei von fünf Expert*innen unabhängig voneinander etwas als Fehler angestrichen hatten.

Strahl (2024) weist auf die besondere Problematik bei der Forschung zur Textbeurteilung und zu diagnostischer Kompetenz hinsichtlich des Vergleichsmaßstabs bzgl. des Goldstandards hin, anhand dessen die Akkuratheit der Lehrkrafturteile gemessen wird. Bei allen Problematiken in der Forschung und auch divergierenden Urteile zwischen den Expert*innen konstatiert sie aber auch, dass „dieses Vorgehen allerdings am häufigsten eingesetzt [wird] und [...] eines der besten Maße dar[stellt], die aktuell vorliegen“ (Strahl, 2024, S. 231).

3.3 Durchführung

Im Rahmen der Studie wurden die teilnehmenden Lehrkräfte gebeten, zwei Schüler*innentexte zu korrigieren. Die Aufgabenstellung lautete wie folgt:

„Korrigieren Sie die beiden Schülertexte entsprechend Ihrer gängigen Korrekturpraxis.

- *Das umfasst die Markierung von Fehlern nach den üblichen Kategorien (R, Gr, Z etc.) und Kommentierung am Rand.*
- *Geben Sie eine Note (in Notenpunkten) und berechnen Sie den Fehlerquotienten.*
- *Notieren Sie dazu bitte Notenpunkte (vor eventuellem) Punktabzug, Fehlerzahl, Fehlerquotient sowie Punktabzug unter den Text in die dafür vorgesehenen Felder.“*

Unter den Texten befand sich ein eingefügtes Ergebnisfeld, in das die Lehrkräfte Fehlerzahl, Note und Punktabzug notierten, was eine einheitliche Auswertung und einen Vergleich der Ergebnisse ermöglichte.

Ergänzend erhielten die Teilnehmenden einen Auswertungsbogen, auf dem sie Förderschwerpunkte ankreuzen konnten, die in der damals gültigen OAVO in Hessen als Fehlerkategorien aufgeführt waren (Rechtschreibung, Zeichensetzung, Grammatik, Ausdruck).

Den Lehrkräften wurden die Texte und Aufgabenstellungen postalisch zugesandt. Sie partizipierten freiwillig und ohne Incentive.³

³ Zur Rekrutierung wurden zunächst Schulen kontaktiert, woraufhin sich Lehrkräfte sich zur Studienteilnahme meldeten. Anschließend wurden die Forschungsinstrumente postalisch verschickt. Die Rücklaufquote betrug 44,7 Prozent.

4 Darstellung ausgewählter empirischer Ergebnisse aus DIEGO

4.1 Expert*innenurteile

In einem ersten Schritt der Auswertung wurden die Expert*innenurteile bestimmt. Die Fehlerzahl unter der Klausur ist ein für die Studie zentraler Wert aufgrund seiner Bedeutung für den Fehlerindex und Punktabzug sowie als diagnostische Rückmeldung zur Sprachkompetenz in den Bereichen Sprachrichtigkeit und Sprachangemessenheit für die Lernenden. Dazu wurde auf Basis der von den Expert*innen bestimmten Fehlerzahlen ein Korridor gebildet, der die *range* der Expert*innenfehlerzahl abbildet. Dieser Korridor liegt für den Text von „SchülerIn 1“ zwischen 30 und 55 Fehlern und beim Text von „SchülerIn 2“ zwischen 69 und 93 Fehlern. Die Urteile von Lehrkräften, deren angegebene Fehlerzahl innerhalb dieses Korridors liegt, werden als „Expert*innenurteil getroffen“ ausgewertet. Ihr Urteil ist damit akkurat. Die Urteile von Lehrkräften, die unterhalb dieses Korridors liegen, wurden als „Überschätzung“ analysiert, diejenigen von Teilnehmenden, die oberhalb dieses Korridors lagen, wurden als „Unterschätzung“ gewertet. Deren Urteil ist damit nicht akkurat. Nach den Förderschwerpunkten befragt, gaben die Expert*innen an, dass bei „SchülerIn 2“ alle vier möglichen Förderschwerpunkte (Rechtschreibung, Zeichensetzung, Grammatik, Ausdruck) förderbedürftig sind. Bezüglich „SchülerIn 1“ urteilten alle Expert*innen unabhängig voneinander übereinstimmend, dass Ausdruck und Grammatik Förderschwerpunkte sind, Rechtschreibung und Zeichensetzung hingegen nicht.

Für die Auswertung der Akkuratheit der Förderschwerpunkte wurde eine Vierfeldertafel verwendet (vgl. Geist, 2014, S. 197):

Tabelle 1: Vierfeldertafel zur Auswertung der Urteilsakkuratheit der Förderschwerpunkte (eigene Darstellung nach Geist, 2014, S. 197)

Bereich X (SchülerIn 1/2)		Expert*innen: Förderbedarf im Bereich X	
		Ja	Nein
Lehrkraft: Förderbedarf im Bereich X:	Ja	Korrekt-Positiv	Falsch-Positiv
	Nein	Falsch-Negativ	Korrekt-Negativ

4.2 Fehlerzahlen und Expert*innenkorridor

Die folgende Darstellung gibt einen deskriptiven Überblick über die Werte, die die teilnehmenden Lehrkräfte ($N = 51$) unter der Klausur notierten.

Tabelle 2: Ergebnisse unter der Klausur (eigene Berechnung)

	SchülerIn 1 (N = 51)	SchülerIn 2 (N = 51)
Note vor Punktabzug in Notenpunkten (NP)	Ø 8,66 (N = 48) SD = 2,15 range: 4–13	Ø 4,72 (N = 47) SD = 1,79 range: 1–9
Fehlerzahl	Ø 62,15 SD = 19,12 range: 18 – 119	Ø 94,65 SD = 24,31 range: 54–199
Fehlerquotient	Ø 3,38 SD = 1,07 range: 0,99–6,50	Ø 11,56 SD = 3,01 range: 6,69–24,66
Punktabzug in NP	Ø 1,24 SD = 0,06 range: 0–3	Ø 3,82 SD = 0,39 range: 3 – 4 ⁴
Endnote in NP	Ø 7,43 (N = 48) SD = 2,24 range: 2–11	Ø 1,28 (N = 47) SD = 1,36 range: 0–5

Sowohl bei der Note als auch bei der Fehlerzahl wird „SchülerIn 1“ besser beurteilt wird als „SchülerIn 2“.

Neben der großen *range* der Noten zeigen sich bei der unter der Klausur notierten Fehlerzahl ebenfalls sehr große Unterschiede, was sich auf den Fehlerindex und den Punktabzug auswirkt.

4.2.1 „SchülerIn 1“

Für „SchülerIn 1“ wurde als Expert*innenkorridor aufgrund der Expert*innenurteile eine Fehlerzahl von 30 bis 55 festgelegt. Der Mittelwert der Fehlerzahlen bei dieser*die-
sem Lernenden von 62,19 Fehlern liegt nicht im Expertenkorridor; damit wäre das Expert*innenurteil verfehlt und das Lehrkrafturteil nicht akkurat. Da die Fehlerzahl über dem hier als Goldstandard angesehenen Wert liegt, wird die Leistung von „SchülerIn 1“ hier als schlechter beurteilt und ihre*seine anhand dieses Textes sichtbar werdende Kompetenz wird unterschätzt.

⁴ Zum Zeitpunkt der Studie galt in Hessen noch ein Punktabzug von bis zu 4 NP im Fach Deutsch, der zweistufig von 0 bis 8 gestaltet war, d.h. FI >2 bedeutet 1 NP Abzug usw. (vgl. OAVO, 2010).

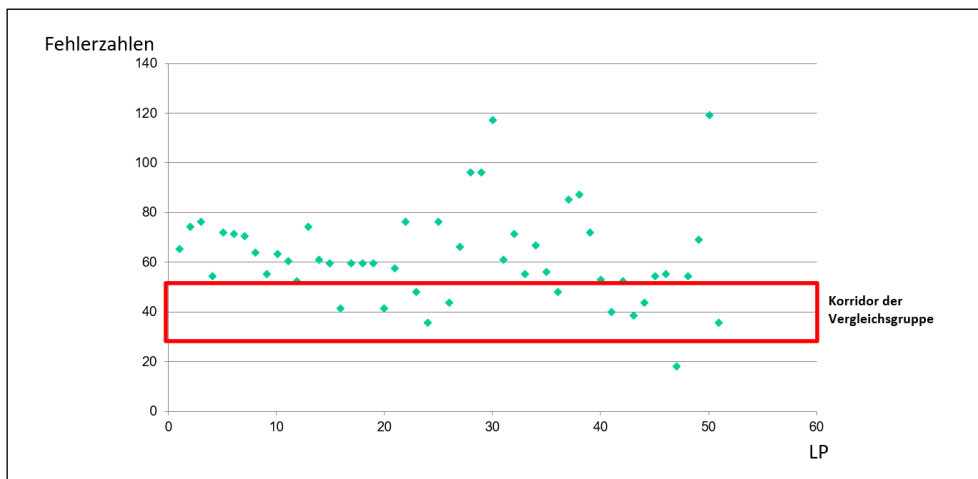


Abbildung 1: Streudiagramm Fehlerzahlen mit Expert*innenkorridor für SchülerIn 1 (LP = Lehrperson; eigene Darstellung)

19 Lehrkräfte (= 37,2 %) stimmen mit dem Urteil eindeutig überein. 31 Teilnehmende (= 60,78 %) befinden sich mit ihrem Urteil über dem Expert*innenkorridor und unterschätzen damit „SchülerIn 1“. Das Urteil einer Lehrkraft (= 2 %) liegt unterhalb des Korridors, was als Überschätzung gewertet wird.

Dabei zeigt sich, dass die große *range* eher durch Extremwerte an beiden Rändern entsteht. Gleichzeitig wird die aufgrund der Standardabweichung ersichtliche große Streuung vor Augen geführt (vgl. Tab. 2). Die Minderheit stimmt mit dem dort eingetragenen Expert*innenkorridor überein; die deutliche Mehrheit beurteilt die Schüler*innenleistung schlechter und es kommt damit zu einer Unterschätzung dieser.

4.2.2 „SchülerIn 2“

Für „SchülerIn 2“ wurde als Expert*innenkorridor aufgrund der Expert*innenurteile eine Fehlerzahl von 69 bis 93 festgelegt. Der Mittelwert der Fehlerzahlen bei diesem Lernenden von 94,65 Fehlern liegt nicht im Expert*innenkorridor. Damit ist das Expert*innenurteil nicht getroffen und das Lehrkrafterteil nicht akkurat. Es zeigt sich die gleiche Tendenz wie bei „SchülerIn 1“ – eine mehrheitliche Unterschätzung der Schüler*innenleistung.

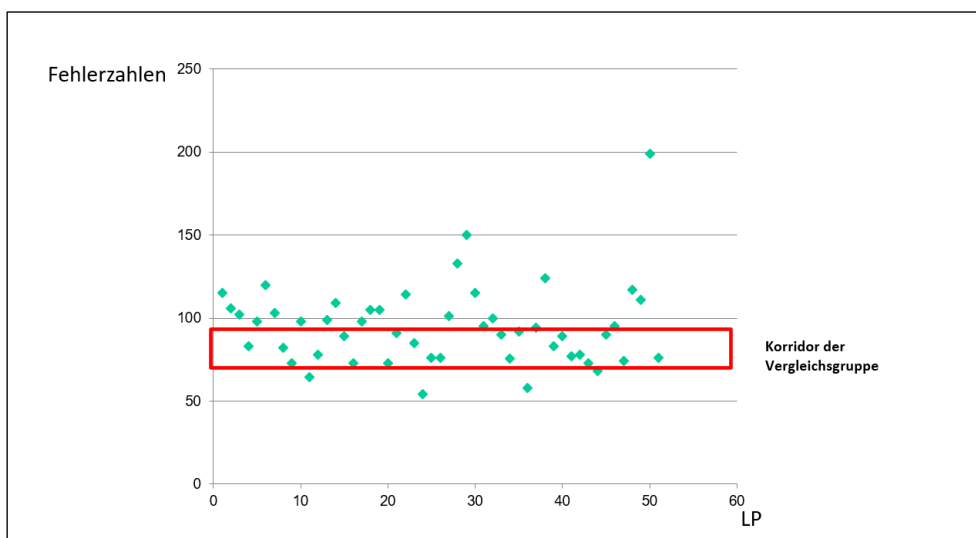


Abbildung 2: Streudiagramm Fehlerzahlen mit Expert*innenkorridor für „SchülerIn 2“ (eigene Darstellung)

Das Urteil von 22 Lehrkräften (= 43,1 %) entspricht dem Expert*innenurteil. Bei 25 Lehrkräften (= 49 %) kommt es zu einer Unterschätzung der Schüler*innenleistung. Das Urteil von vier Studienteilnehmenden (= 4,8 %) wird als Überschätzung der Schüler*innenleistung gewertet, da sie eine geringere Fehlerzahl als die Expert*innen diagnostiziert haben und unterhalb des Expert*innenkorridors liegen. Grundsätzlich ähnelt das Ergebnis jenem Bild, was sich schon bei „SchülerIn 1“ gezeigt hat: wenige Extremwerte, mehrheitlich keine Übereinstimmung mit dem Expert*innenurteil und damit keine akkuraten Urteile der meisten Lehrkräfte sowie eine Tendenz zu Überschätzung, wie auch der Vergleich des Mittelwertes mit dem Expert*innenkorridor hat vermuten lassen.

4.3 Textimmanente Korrektur

Im Rahmen der Studie wurde zudem die Akkuratheit für die Korrektur im Text sowie bezüglich der verwendeten Korrekturzeichen erfasst, was eine tiefere Analyse als in den meisten Studien zur diagnostischen Kompetenz darstellt. Durch ein wortorientiertes Codierungssystem wurden alle Fehlermarkierungen und die verwendeten Korrekturzeichen erfasst und in Bezug zum Goldstandard in dieser Untersuchung gesetzt. Das Codierungssystem bezieht sich auf die in den Fehlerindex zählenden Basisdimensionen Sprachrichtigkeit und Sprachangemessenheit. Das Vorgehen wird anhand zweier Fallbeispiele im Folgenden veranschaulicht.

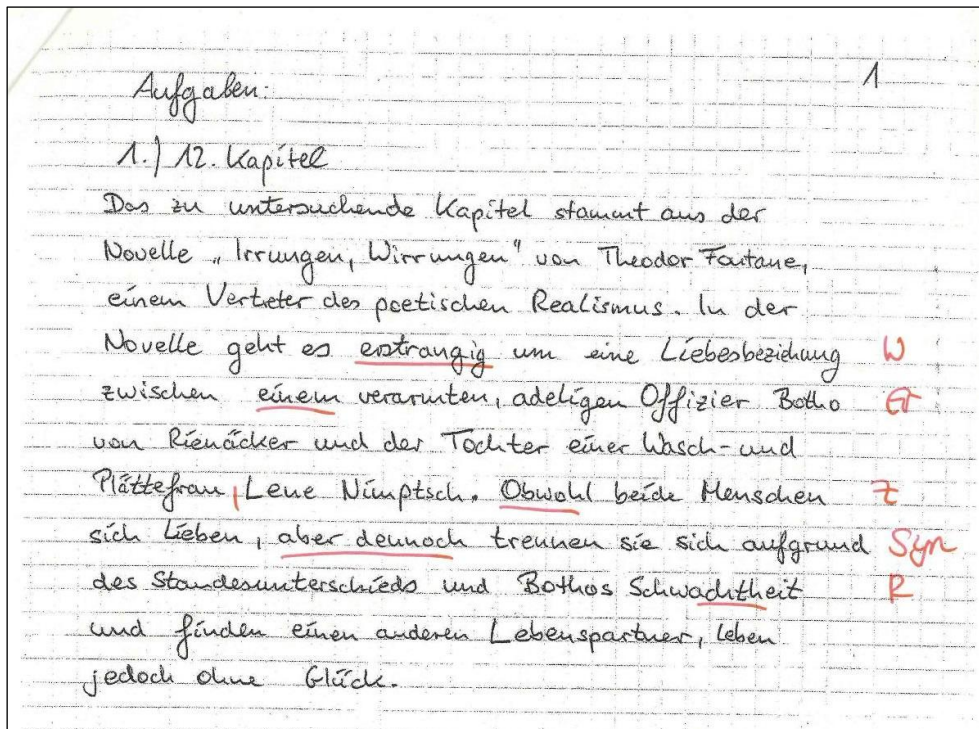


Abbildung 3: Korrektur des Textanfangs von „SchülerIn 1“ durch Lehrperson (LP) 1 (Forschungsprojekt DIEGO)

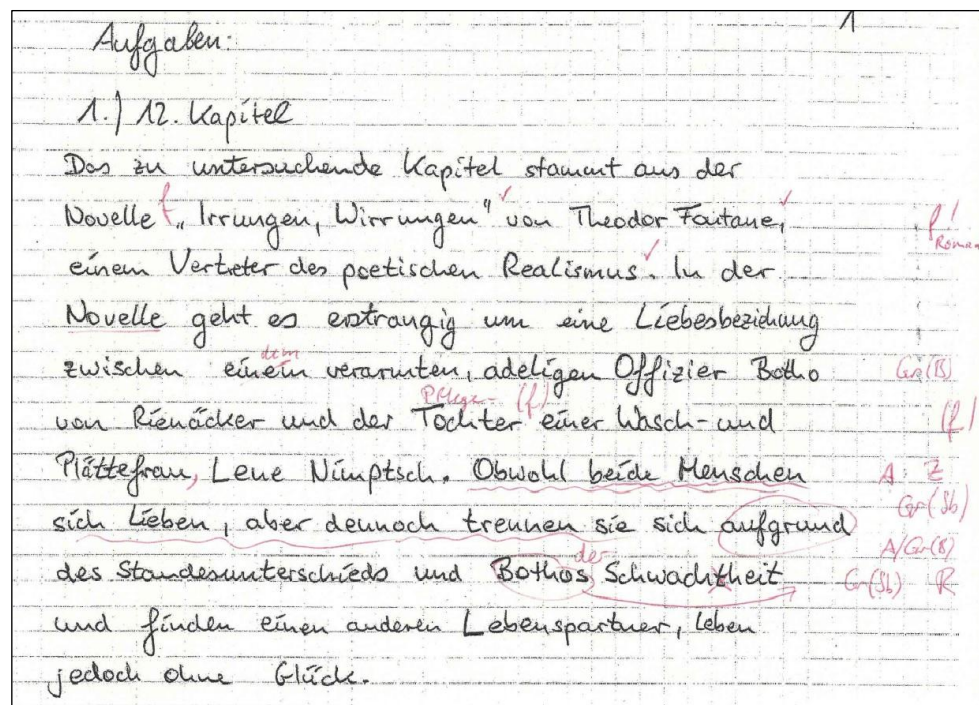


Abbildung 4: Korrektur des Textanfangs von „SchülerIn 1“ durch LP 52 (Forschungsprojekt DIEGO)

Schon auf den ersten Blick fallen Unterschiede im Korrekturverhalten zwischen Lehrperson (LP) 1 und LP 52 auf. Während LP 1 in diesem Abschnitt fünf sprachliche und fehlerindexrelevante Fehler der Basisdimensionen Sprachrichtigkeit und Sprachangemessenheit anstreicht, sind es bei LP 52 sieben.

Im Folgenden werden die von „SchülerIn 1“ genannte Textsorte „Novelle“ (Z. 4) und das Wort „erstrangig“ (Z. 6) exemplarisch betrachtet, die zur Basisdimension Sprachangemessenheit gehören. Für die Expert*innen stellt die Textsorte „Novelle“ keinen sprachlichen Fehler dar. Inhaltlich ist dies falsch, da es sich bei Fontanes Werk um einen Roman handelt. Sowohl LP 1 als auch LP 52 haben dies auch nicht als sprachlichen Fehler markiert, LP 52 hat dies inhaltlich als fehlerhaft gekennzeichnet und mit der richtigen Textsorte als Randkommentar versehen. Zehn Teilnehmende (= 19,6 %) markierten dies ebenso als sprachlichen Fehler.

Für LP 1 ist „erstrangig“ an dieser Stelle fehlerhaft und sie klassifiziert dies mit dem Korrekturzeichen „W“ als lexikalischen Fehler. LP 52 hingegen hat hier keine Fehlermarkierung vorgenommen, obwohl LP 52 mehr Fehler in diesem Abschnitt konstatiert hat. Insgesamt haben 17 Teilnehmende (= ein Drittel der Stichprobe) an dieser Stelle einen Fehler markiert und im Vergleich zum Expert*innenmaßstab kein akkurates Urteil abgegeben. Es zeigt sich bei LP 1 und LP 52 kein eindeutiges Muster bezüglich Quantität und Qualität der Fehler.

Diese beiden Beispiele zeigen, dass die Basisdimension Sprachangemessenheit ein Bereich ist, in dem Unterschiede bestehen. Die vorliegenden Daten zeigen bei einem Blick auf alle Probanden in der Basisdimension Sprachrichtigkeit auch deutliche Unterschiede. Es sind markante Differenzen bei der Beurteilung der Zeichensetzung festzustellen. Die beiden Fallbeispiele zeigen stellvertretend in der gesamten Proband*innen-gruppe zu erkennende Unterschiede im Bereich Grammatik.

Bei einem Blick auf die Korrekturzeichen am Rand von LP 1 und LP 52 springt ins Auge, dass sie teilweise die gleichen Zeichen verwenden. Bei LP 52 unterscheidet sich der Differenzierungsgrad bei Grammatikfehlern. Diese Lehrkraft unterscheidet diese, was eine differenziertere Diagnose und auch Rückmeldung an die Lernenden darstellt, wenn damit weitergearbeitet wird. LP 1 differenziert auch Grammatik (Gr) und Satzbau (Syn). Zudem kommt bei ihr das Korrekturzeichen „W“ vor. Es zeigt sich bei einem Blick in die korrigierten Texte eine Vielzahl und große Varietät mit unterschiedlichem Differenzierungsgrad bei den Korrekturzeichen. Lehrkräfte, die das Wort „Schwachheit“ (Z. 11) als fehlerhaft kennzeichneten, benutzten als Korrekturzeichen: R, A, W, L, R/A, Gr oder Wort.

4.4 Förderschwerpunkte

Da die Expert*innen bei „SchülerIn 1“ alle vier möglichen Förderschwerpunkte (Rechtschreibung, Zeichensetzung, Grammatik, Ausdruck) als förderbedürftig benannten (vgl. Kap. 4.1), wird hier die Akkuratheit der Urteile zu den Förderschwerpunkten bei „SchülerIn 1“ näher betrachtet, da laut dem hier verwendeten Goldstandard nur Ausdruck und Grammatik Förderschwerpunkte sind, Rechtschreibung und Zeichensetzung hingegen nicht.

Hinsichtlich des Förderschwerpunkt Grammatik kommen 46 Lehrkräfte (= 90,2 %) zu einem korrekt-positiven Urteil. Fünf Lehrkräfte (= 9,8 %) benennen im Unterschied zum Expert*innenurteil Grammatik als förderbedürftig (= falsch-negativ) und überschätzen damit die in der vorliegenden Klausur gezeigte Kompetenz.

Beim Förderschwerpunkt Ausdruck zeigt sich ein insgesamt vergleichbares Bild: 44 Lehrkräfte (= 86,3 %) fällen ein korrekt-positives Urteil, sieben Lehrkräfte (= 13,7 %) kommen zu einem falsch-negativen Urteil und überschätzen die Ausdrucksfähigkeit von „SchülerIn 1“.

Bei den Förderschwerpunkten Grammatik und Ausdruck fällt deutlich auf, dass die Teilnehmenden zu großer Mehrheit das Expert*innenurteil treffen und damit korrekt-positive bzw. akkurate Urteile fällen.

Bei den zu Sprachrichtigkeit gehörenden Bereichen Rechtschreibung und Zeichensetzung zeigen sich andere Ergebnisse. 44 Lehrkräfte (= 86,3 %) benennen Rechtschreibung als Förderschwerpunkt. Sie geben damit ein falsch-positives und nicht akkurates Urteil bezüglich des Förderbedarfs ab. Sieben Lehrkräfte (= 13,7 %) kommen zu einem korrekt-negativen Urteil, das damit akkurat ist.

Weitestgehend vergleichbar sind die Resultate beim möglichen Förderschwerpunkt Zeichensetzung. 40 Lehrkräfte (= 78,4 %) bestimmen hier einen Förderschwerpunkt. Damit ist das Urteil falsch-positiv und ist nicht akkurat. Elf Lehrkräfte (= 21,6 %) sind in ihrem Urteil korrekt-negativ und somit akkurat, da sie in Übereinstimmung mit dem Expert*innenurteil Zeichensetzung nicht als Förderschwerpunkt sehen.

Bei den beiden Bereichen, die die Expert*innen als nicht als Förderschwerpunkt angesehen haben, ist eine deutliche Diskrepanz zu erkennen. Die große Mehrheit der Teilnehmenden benennt diese Bereiche als Förderschwerpunkt und gibt ein falsch-positives Urteil ab, was als Unterschätzung der Schülerleistung gewertet werden kann. Zudem kann geschlussfolgert werden, dass damit auch keine Lehrkraft die Kompetenz der Lernenden in Rechtschreibung und Zeichensetzung als Teile von Sprachrichtigkeit überschätzt.

5 Diskussion der Ergebnisse

Die vorliegende Studie untersuchte die sprachdiagnostische Kompetenz von Deutschlehrkräften der Sekundarstufe II bei der Korrektur von Schüler*innentexten. Im Mittelpunkt stand erstens die Frage nach der Homogenität der Urteile der Lehrkräfte. Zweitens wurde danach gefragt, wie akkurat die Urteile im Vergleich zu einem Expert*innenurteil sind und drittens sollte herausgefunden werden, ob die aus der Korrektur abgeleiteten Förderschwerpunkte korrekt sind.

Die Ergebnisse zeigen, dass in den durch die 51 Teilnehmenden in denselben Schüler*innentexten bestimmten Fehlerzahlen eine große Streuung und damit Heterogenität im Bereich der Sprachrichtigkeit und Sprachangemessenheit vorliegt. Folglich zeigt sich in den vorliegenden Daten eine Heterogenität von auf Sprache bezogenen Lehrkrafturteilen der Sekundarstufe II im Fach Deutsch. Die Fehlerzahl als numerische Diagnose der Sprachrichtigkeit und Sprachangemessenheit ist eine Rückmeldung an die Lernenden zu deren Sprachkompetenz und Auswirkungen auf das Abitur sind abhängig von der Lehrkraft. Dies deckt sich mit Erkenntnissen der Forschung zur Notengebung und zur diagnostischen Kompetenz, die sich auf die hier untersuchten Werte übertragen lassen. Die exemplarische Analyse der textimmanenten Korrektur veranschaulichte Urteils-heterogenität bei Fehlerbestimmung, -klassifizierung und Korrekturzeichen.

Zweitens zeigen die Ergebnisse mehrheitlich keine Akkuratheit der Urteile. 37,2 Prozent der Lehrkräfte kommen bei „SchülerIn 1“ und 43,1 Prozent bei „SchülerIn 2“ zu einem akkuraten Urteil, das mittels eines Expert*innenkorridors als Goldstandard ermittelt wurde.

Gleichwohl zeigte sich hierbei eine Tendenz zur strengeren Beurteilung bzw. Unterschätzung der Lernenden. Die Unterschätzung der sprachlichen Kompetenz in den Dimensionen Sprachrichtigkeit und Sprachangemessenheit führt zu einer sich auf die individuelle Sprachförderung auswirkende Sprachdiagnose und zu einem höheren Fehlerindex, der direkte Auswirkung auf Bildungserfolg hat und eine unmittelbare Rückmeldung an Lernende darstellt. Dies wurde sowohl bei der Fehlerzahl als auch bei der textimmanenten Korrektur sichtbar und bestätigt Forschungsergebnisse anderer Studien mit anderer Population von Lehrkräften oder anderem Fach als Deutsch. Es wird hier

ebenfalls die Notwendigkeit einer differenzierteren Ausbildung von sprachdiagnostischer Kompetenz konstatiert.

Die Frage, ob Lehrkräfte akkurat Förderschwerpunkte aus ihrer Diagnose ableiten, kann teilweise negiert werden. Beim Vergleich mit dem Expert*innenurteil bei „SchülerIn 1“ wurde beleuchtet, dass Lehrkräfte bei Grammatik und Ausdruck als Förderschwerpunkte mehrheitlich richtig-positive Urteile, jedoch mehrheitlich falsch-positive Urteile bei Rechtschreibung und Zeichensetzung fällen.

Wenn Lehrkräfte bei „SchülerIn 1“ alle Förderbereiche als förderbedürftig rückmelden, dann führt dies zu einer Überforderung der Lernenden, die in der Folge nicht zielgerichtet an der eigenen Sprachkompetenz arbeiten können. Ebenso bedient Sprachförderung, etwa durch die Auswahl und Bereitstellung von differenzierendem Sprachfördermaterial, Übungen für die Lerngruppe und oder die „Passung des Unterrichts“ (Terhart, 2000, S. 15), nicht die notwendigen Sprachförderbedürfnisse, sodass anzunehmen ist, dass in diesem Fall bei „SchülerIn 1“ ineffektiv die Schreibkompetenz im Kriterium Orthografie der Basisdimension Sprachrichtigkeit (vgl. Becker-Mrotzek & Böttcher, 2014) gefördert wird. Hier kann auch eine Übereinstimmung mit Forschungsergebnissen zu anderen Zielgruppen festgestellt werden.

Zusammenfassend zeigen die Ergebnisse übereinstimmend mit der Forschungslage eine zu fördernde sprachdiagnostische Kompetenz bei vielen Lehrkräften. In Anlehnung an andere empirische Ergebnisse zur diagnostischen Kompetenz und zur Beurteilung von Schüler*innentexten zeigt sich auch anhand der Einblicke in das Forschungsprojekt DIEGO, „dass die Bewertung von Texten ein fehleranfälliger Prozess ist“ (Strahl, 2024, S. 234).

Einschränkend muss bedacht werden, dass Fragen in diesem Beitrag lediglich anhand der zwei Basisdimensionen Sprachrichtigkeit und Sprachangemessenheit untersucht wurden und andere Elemente der Schreibkompetenz und Textqualität außer Acht blieben. Allerdings sind diese beiden Basisdimensionen aufgrund ihrer Bedeutung für den Fehlerindex hinsichtlich Bildungserfolg und Diagnostik äußerst relevant in der schulischen Praxis und die Ergebnisse deuten darauf hin, dass Heterogenität und mangelnde Akkuratheit auch bei anderen Basisdimensionen zu vermuten sind.

Durch den Fokus auf Deutschlehrkräfte der gymnasialen Oberstufe wird die Forschung zur (sprach-)diagnostischen Kompetenz erweitert. Die untersuchten Facetten der Urteilsakkuratheit ermöglichen zudem differenziertere Aussagen über die sprachdiagnostische Kompetenz von Lehrkräften.

Selbstredend zeigen auch die Expert*innen eine gewisse Heterogenität ihrer Urteile, was „ein Problem der Forschung zur Textbewertung“ (Strahl, 2024, S. 234) darstellt. Mit dem Expert*innenkorridor wurde hier mittels eines praxisnahen Instruments der Versuch unternommen, die Akkuratheit von Lehrkrafturteilen möglichst nach an der Korrekturrealität zu erheben, was gegenwärtig das angemessenste Vorgehen für Forschung zur diagnostischen Kompetenz bei der Textbeurteilung ist (vgl. Strahl, 2024).

Gleichsam aufschlussreich wäre eine Replikationsstudie (vgl. Birkel & Birkel, 2002) mit den Texten aus DIEGO, um Entwicklungen mit Blick auf die Veränderung der Curricula der ersten beiden Lehrkräftebildungsphasen sowie Fortbildungsangeboten zu untersuchen. Wegen der fachübergreifenden Relevanz von Textproduktion wäre es erstrebenswert, die sprachdiagnostische Kompetenz von Lehrkräften anderer Fächer mit dem vorliegenden Erhebungsinstrument oder/und mit fachlich bezogenen Sample-Texten zu untersuchen, denn wie Forkarth & Manzel (2024) bezogen auf Politiklehrkräfte darstellen, fehlen hier noch Studien.

Der seit Jahren festgestellte Forschungsbedarf zu Einflussfaktoren auf die diagnostische Kompetenz (Artelt & Gräsel, 2009; Bräuer & Winkler, 2012; Strahl, 2024; Südkamp, 2010) besteht weiterhin und eine wie hier angelegte Studie könnte noch mehr Einflussfaktoren erheben und zu erwartende Unterschiede oder fehlende Akkuratheit bei Urteilen ergründen. Hier könnte etwa die Berufserfahrung (vgl. Dünnebier et al., 2009)

betrachtet werden. Eine Studie mit Lehramtsstudierenden könnte mit Ergebnissen der Forschungsprojekte ASSET („Assessing Students‘ English Texts“, vgl. Machts et al., o.J.) und ASSET-G („Assessing Students‘ Texts – German“, vgl. Möller et al., o.J.) in Relation gesetzt werden, wonach „erfahrene Lehrkräfte strenger und somit ungenauer urteilen als Lehramtsstudierende“ (Strahl, 2024, S. 233).

Weiterhin wäre es relevant zu untersuchen, welche Fördermaßnahmen Lehrkräfte aus ihrer Diagnose ableiten, denn erst dann kann eine Diagnose auch sprachlernwirksam werden.

Um weiteren Erklärungsfaktoren auf den Grund zu gehen, wären Laut-Denk-Protokolle dienlich, um über die Korrektur- und Fehlerklassifizierungsentscheidungen qualitative Daten zu gewinnen.

6 Implikationen für eine fächerübergreifende Lehrkräftebildung

Die Resultate dieser Studie untermauern mit Blick auf die großen Streuungen und die in der Minderheit der Stichprobe zu findende Akkuratheit die Notwendigkeit, dass die Aufgabe der durchgängigen Sprachbildung im Rahmen der Lehrkräftebildung als Querschnittsaufgabe gesehen werden muss. Die Entwicklung der lehrkräfteseitigen Fähigkeit zu akkuraten Urteilen und die Sensibilisierung für Urteilsfehler als Teil sprachdiagnostischer Kompetenz muss in allen Phasen der Lehrkräftebildung verankert sein, und zwar vor allem in den Unterrichtsfächern, „in denen schriftsprachliche Leistungen bewertet werden“ (Strahl, 2024, S. 235). Dies betrifft insbesondere Deutsch und gesellschaftswissenschaftliche Fächer (vgl. z. B. Forkarth & Manzel, 2024).

Wie berichtet, zeigen sich in der Urteilsakkuratheit und auch in der Ableitung von Förderschwerpunkten Unterschiede zwischen den untersuchten Deutschlehrkräften. Es ist davon auszugehen, dass sich diese Differenzen ebenfalls bei Lehrkräften anderer Fakultas offenbaren. Wenn schon bei Deutschlehrkräften Urteile in den hier untersuchten Basisbereichen differieren, kann davon ausgegangen werden, dass dies bei Lehrkräften ohne Fakultas im Fach Deutsch ebenfalls der Fall ist (vgl. Forkarth & Manzel, 2024).

Unter der Annahme, dass „das Fachwissen hingegen [...] zu genaueren Beurteilungen“ (Strahl, 2024, S. 233) führt, ist linguistisches Wissen und Wissen rund um Sprachdiagnostik ein notwendiger Lerninhalt für (angehende) Lehrkräfte aller Fächer.

Aufgrund der Indizien für die Lernbarkeit der Textbeurteilung (vgl. Peschel, 2024, S. 55) erscheint eine Integration von Inhalten der Textbeurteilung nicht nur aus Gründen der Bildungsgerechtigkeit, sondern vor allem zur Förderung der Schreibkompetenzen von Lernenden erstrebenswert.

Vorstellbar wären hier Lehrveranstaltungen, in denen Studierende ähnlich wie die Teilnehmenden dieser Studie (möglichst authentische) Schüler*innentexte korrigieren, eine Fehlerzahl bestimmen, Fehler kennzeichnen und klassifizieren sowie Förderschwerpunkte bestimmen. Anschließend vergleichen die Studierenden ihr Urteil mit einem Expert*innenstandard und reflektieren abschließend theoretisch fundiert und diskursiv. Aufgrund der gezeigten essenziellen Rolle der Kompetenz beim Verfassen von Texten, dem Potenzial von Texten zur Sprachdiagnostik und den gezeigten Ergebnissen für Deutschlehrkräfte bedarf es der obligatorischen Verankerung solcher Lehrkonzepte in allen Studiengängen zu Unterrichtsfächern, in denen Textproduktion gefragt ist, um Sprachbildung als Querschnittsaufgabe zu erfüllen. Für eine möglichst professionsnahe Ausbildung sprachdiagnostischer Kompetenz kann dafür plädiert werden, dass „die Arbeit an authentischem Material, also an echten Schülertexten passiert“ (Peschel, 2024, S. 56). Strukturell sinnvoll wäre Sprachdiagnostik als fächerübergreifender Lerninhalt für Fächer wie Deutsch oder Gesellschaftswissenschaften, um voneinander und miteinander zu lernen.

Peschel (2024) folgend wäre es sogar noch gewinnbringender, wenn die theoretische Auseinandersetzung mit praktischen Elementen wie Schulpraktika verknüpft wäre, wobei die Studierenden noch mehr Einblicke in die Beobachtung des Schreibprozesses im Fachunterricht erhalten könnten.

Da diagnostische Kompetenz nur dann sinnvoll und ertragreich für Lernende ist, wenn Ergebnisse der Diagnostik mit entsprechenden Fördermaßnahmen kombiniert werden können (Schrader & Helmke, 2001, S. 53), sollte darüber hinaus die Ausbildung diagnostischer Kompetenz nicht lediglich auf die Akkuratheit der Urteile fokussiert sein trainiert werden.

Für die Schulpraxis wäre ebenfalls angeraten, Textkorrekturen auf die sprachlichen Kompetenzen fachübergreifend zu vergleichen und zu reflektieren, gerade auch in Hinblick auf die Korrekturzeichen. Verschiedene Fehlerzeichen sind für die Unterrichtspraxis relevant. Wenn Schülerinnen und Schüler derselben Schule unterschiedlichen Korrekturzeichen begegnen, können sie die damit differenzierende Rückmeldung zur Weiterentwicklung möglicherweise nicht zielgerecht verstehen und es verwirrt sie. Entscheidend scheint eine Einbettung in den Unterricht oder eine Transparenz der Korrekturzeichen. Teilweise gibt es bereits Vereinbarungen an Schulen; in jüngster Vergangenheit hat das Hessische Kultusministerium einheitliche Korrekturzeichen veröffentlicht.

Eine Förderung sprachdiagnostischer Kompetenz fächerübergreifend zu betrachten, wäre nicht nur für alle Phasen der Lehrkräftebildung ein Gewinn, sondern vor allem für Lernende, deren Sprachförderung auf einer differenzierten und akkuraten Diagnose fußt.

Literatur und Internetquellen

- Achilles, H. (2008). Diagnostisches Potential von Klassenarbeiten und schriftlichen Übungen. In S. Kliemann (Hrsg.), *Diagnostizieren und Fördern in der Sekundarstufe I. Schülerkompetenzen erkennen, unterstützen und ausbauen* (S. 36–45). Cornelsen.
- 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>
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Leske und Budrich. https://doi.org/10.1007/978-3-322-83412-6_10
- Baurmann, J. (2014). Schulisches Schreiben. In M. Kämper-van den Boogaart (Hrsg.), *Deutsch-Didaktik. Praxishandbuch für die Sekundarstufe I und II* (5., überarb. Neuauflage) (S. 251–267). Cornelsen.
- Becker-Mrotzek, M. & Böttcher, I. (2014). *Schreibkompetenz entwickeln und beurteilen* (5., Auflage). Cornelsen. <https://doi.org/10.2307/j.ctvddzg18.5>
- Birkel, P. (2003). Aufsatzbeurteilung – ein altes Problem neu untersucht. *Didaktik Deutsch*, 8 (15), 46–63.
- Birkel, C. & Birkel, P. (2002). Wie einig sind sich Lehrkräfte bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht*, 12, 129–224.
- Bräuer, C. & Winkler, I. (2012). Aktuelle Forschung zu Deutschlehrkräften. Ein Überblick. *Didaktik Deutsch*, 33, 74–91.
- Cakir, Y. (2009). *In der Zweitsprache Deutsch Seminararbeiten schreiben. Analyse zum makro- und mikrostrukturellen Textwissen türkischer Muttersprachlerinnen und Muttersprachler. Eine explorative Studie*. Universitätsverlag Rhein-Ruhr.
- Dünnebieber, K., Gräsel, C. & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. Eine experimentelle Studie zu Ankereffekten. *Zeitschrift für Pädagogische Psychologie*, 23 (3/4), 187–195. <https://doi.org/10.1024/1010-0652.23.34.187>
- Forkarth, C. & Manzel, S. (2024). Diagnoseinstrumente zur Beurteilung und Förderung von Schreibkompetenzen im sozialwissenschaftlichen Fachunterricht. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Texte schreiben in allen Unterrichtsfächern. Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 131–148). Waxmann.
- Frerichs, D. & Kaleta, M. (2015). (Sprachliche) Förderung in der Sekundarstufe II und im Übergang zur Hochschule. In J. Boesten, W. Ostermann, & T. Helmig (Hrsg.), *Sprache bildet! Auf dem Weg zu einer durchgängigen Sprachbildung* (S. 210–224). Verlag an der Ruhr.
- Geist, B. (2014). *Sprachdiagnostische Kompetenz von Sprachförderkräften*. De Gruyter. <https://doi.org/10.1515/9783110308556>
- Geist, B. & Krafft, A. (2019). *Deutsch als Zweitsprache* (2., aktual. Aufl.). Narr Francke Attempto. <https://doi.org/10.24053/9783823393399>
- Gogolin, I., Lengyel, D., Bainski, C., Lange, I., Michel, U., Rutten, S. & Scheinhardt-Stettner, H. (2020). *Durchgängige Sprachbildung. Qualitätsmerkmale für den Unterricht* (2., überarb. u. erw. Aufl., FörMig Material, Bd. 10). Waxmann.
- Hachfeld, A., Schroeder, S., Anders, Y., Hahn, A. & Kunter, M. (2012). Multikulturelle Überzeugungen. Herkunft oder Überzeugung? Welche Rolle spielen der Migrationshintergrund und multikulturelle Überzeugungen für das Unterrichten von Kindern mit Migrationshintergrund? *Zeitschrift für Pädagogische Psychologie*, 26 (2), 101–120. <https://doi.org/10.1024/1010-0652/a000064>

- HKM (Hessisches Kultusministerium). (2010). *Oberstufen- und Abiturverordnung (OAVO) vom 20. Juli 2009 in der Fassung vom 1. Juni 2010*. https://www.rv.hessenrecht.hessen.de/bshe/document/hevr-OSt_AbiVHEpIVZ
- HKMB (Hessisches Ministerium für Kultus, Bildung und Chancen). (2025). *Oberstufen- und Abiturverordnung (OAVO) vom 20. Juli 2009 in der Fassung vom 4. Juni 2025*. https://www.rv.hessenrecht.hessen.de/bshe/document/hevr-OSt_AbiVHEV14IVZ
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (4., aktual. Aufl.). Klett/Kallmeyer.
- Helmke, A. & Jäger, R.S. (2002). (Hrsg.). *Die Studie MARKUS – Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Verlag Empirische Pädagogik.
- Hosenfeld, I., Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (Zeitschrift für Pädagogik, 45. Beiheft, S. 65–82). Beltz.
- Ingenkamp, K. (1995). *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (9. Aufl.). Beltz.
- Jäger, R.S. (2007). *Beobachten, beurteilen und fördern. Lehrbuch für die Aus-, Fort- und Weiterbildung*. Verlag Empirische Pädagogik.
- Jürgens, E. (2012). Diagnosefunktion: Fehlanzeige. Ist eine Neuorientierung der Beurteilungspraxis nötig? In C. Fischer (Hrsg.), *Diagnose und Förderung statt Notengebung? Problemfelder schulischer Leistungsbeurteilung* (S. 23–44). Waxmann.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23, 197–209. <https://doi.org/10.1024/1010-0652.23.34.197>
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern*. Waxmann.
- Larew, O. (2014). Zum Stellenwert der Sprachkompetenz in Programmen zum wissenschaftlichen Schreiben. Erfahrungen aus dem Programm für Lehramtsstudierende mit Deutsch als Zweitsprache „Zwischen den Sprachen“. *Zielsprache Deutsch*, 1, 3–19.
- Leisen, J. (2020). Sprachbildung im sprachsensiblen Politikunterricht. Zur Bedeutung von Bildungssprache und Fachsprache im Politikunterricht. *Sprachbildung im Politikunterricht*. (Sonderheft Wochenschau Verlag 20s/2020, S. 10–20). <https://doi.org/10.46499/1430.1285>
- Lohaus, D. (2008). *Leistungsbeurteilung*. Dt. Wiss.-Verl.
- Lorenz, C. (2011). *Diagnostische Kompetenz von Grundschullehrkräften. Strukturelle Aspekte und Bedingungen*. University of Bamberg Press. <https://doi.org/10.20378/irb-333>
- Lorenz, C. & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23, 211–222. <https://doi.org/10.1024/1010-0652.23.34.211>
- Machts, N., Möller, J. & Keller, S. (o.J.). *ASSET. Assessing Students' English Texts*. https://www.ipl.uni-kiel.de/de/PaedPsych_LuL/forschung/asset-assessing-students-english-texts
- Michalak, M. (2012). Von der Sprachstandsdiagnose zur sprachlichen Förderung. In M. Michalak & M. Kuchenreuther (Hrsg.), *Grundlagen der Sprachdidaktik Deutsch als Zweitsprache* (S. 57–84). Schneider Verlag Hohengehren.
- Möller, J., Kilian, J., Strahl, F. & Balik, E. (o.J.). *ASSET-G. Assessing Students' Texts – German*. <https://www.asset-g.uni-kiel.de/de>

- Müller, A. (2014). Profession und Sprache: Die Sicht der (Zweit-)Spracherwerbsforschung. In T. Betz & P. Cloos (Hrsg.), *Kindheitspädagogische Beiträge. Kindheit und Profession – Konturen und Befunde eines Forschungsfeldes* (S. 66–83). Beltz Juventa.
- Niederdorfer, L., Ebner, C. & Schmölzer-Eibinger, S. (2018). Entwicklung wissenschaftlicher Textkompetenz von SchülerInnen in der Zweitsprache. In W. Griebhaber, S. Schmölzer-Eibinger, H. Roll & K. Schramm (Hrsg.), *Schreiben in der Zweitsprache Deutsch. Ein Handbuch* (S. 121–134). De Gruyter. <https://doi.org/10.1515/9783110354577-009>
- Peschel, C. (2024). Kann man das Beurteilen von (Schüler-)Texten lernen? Einblicke in diagnostisches Vorgehen von Lehrkräften. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Texte schreiben in allen Unterrichtsfächern. Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 39–59). Waxmann.
- Petersen, I., Reble, R. & Kilian, J. (2024). Einleitung. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Texte schreiben in allen Unterrichtsfächern. Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 7–17). Waxmann.
- Peuschel, K. & Burkard, A. (2019). DaZ, Sprachbildung und der geistes- und gesellschaftswissenschaftliche Unterricht. In K. Peuschel & A. Burkard (Hrsg.), *Sprachliche Bildung und DaZ in den geistes- und gesellschaftswissenschaftlichen Fächern* (S. 53–101). Narr Francke Attempo. <https://doi.org/10.24053/9783823391678>
- Pohl, T. (2011). Wissenschaftlich Schreiben. Begriff, Erwerb und Fördermaximen. *Der Deutschunterricht*, 63 (5), 2–11.
- 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* (S. 13–18). Waxmann.
- Prenzel, M., Gogolin, I. & Krüger, H.-H. (2007). Editorial. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (Zeitschrift für Erziehungswissenschaft, Sonderheft 8, S. 5–8). VS-Verlag. https://doi.org/10.1007/978-3-531-90865-6_1
- Scholten-Akoun, D. (2018). Sprachstandstests im hochschulischen Kontext. In W. Griebhaber, S. Schmölzer-Eibinger, H. Roll & K. Schramm (Hrsg.), *Schreiben in der Zweitsprache Deutsch. Ein Handbuch* (S. 229–246). De Gruyter. <https://doi.org/10.1515/9783110354577-016>
- Schrader, F.-W. (2012). Was wissen wir über Diagnostizieren und Fördern durch Lehrer? *Pädagogik*, 6, 42–45.
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilungen durch Lehrer. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Beltz.
- Seipp, B. (2012). Deutsch als Zweitsprache in der Lehrerbildung – Spagat zwischen Hoffnung und Realität. In M. Michalak & M. Kuchenreuther (Hrsg.), *Grundlagen der Sprachdidaktik Deutsch als Zweitsprache* (S. 1–26). Schneider Verlag Hohengehren.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1/2), 85–95. <https://doi.org/10.1024/1010-0652.19.12.85>
- Steinhoff, T. (2025). Künstliche Intelligenz als Ghostwriter, Writing Tutor und Writing Partner. Zur Modellierung und Förderung von Schreibkompetenzen im Zeichen der Automatisierung und Hybridisierung der Kommunikation als Beispiel des Schreibens mit ChatGPT in der 8. Klasse. In C. Albrecht, J. Brüggemann, T. Kretschmann & C. Meier (Hrsg.), *Personale und funktionale Bildung im Deutschunterricht. Theoretische, empirische und praxisbezogene Perspektiven* (S. 85–99). J. B. Metzler. https://doi.org/10.1007/978-3-662-69640-8_7

- Strahl, F. (2024). Die diagnostische Kompetenz von Lehrkräften bei der Beurteilung von Texten. Theoretische Grundlagen, Einflussfaktoren, aktuelle Forschungsergebnisse und Implikationen für die Praxis. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Texte schreiben in allen Unterrichtsfächern. Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 221–240). Waxmann.
- Strecker, G. (2010). Die Förderung von Deutsch als Zweitsprache in der Sekundarstufe II – ein Luxusproblem? Erste Einsichten aus dem Sprachförderprojekt FJM. In M. Rost-Roth (Hrsg.), *DaZ-Spracherwerb und Sprachförderung Deutsch als Zweitsprache. Beiträge aus dem 5. Workshop Kinder mit Migrationshintergrund* (S. 255–272). Fillibach Verlag.
- Südkamp, A. (2010). *Diagnostische Kompetenz: Zur Genauigkeit der Beurteilung von Schülerleistungen durch Lehrkräfte* [Unveröffentlichte Dissertation]. Christian-Albrechts-Universität zu Kiel.
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Supplement Material of 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>
- Terhart, E. (2000). *Lehr-Lern-Methoden* (3., erw. Auflage). Juventa.
- Wagner, L. (2025). Einflussfaktoren auf die Diagnosekompetenz (angehender) Lehrkräfte – ein systematisches Literaturreview. *Unterrichtswissenschaft*, 53, 99–128. <https://doi.org/10.1007/s42010-024-00215-3>
- Weinert, F.E. (2000). Lehren und Lernen für die Zukunft. Ansprüche an das Lernen in der Schule. *Pädagogische Nachrichten Rheinland-Pfalz*, 2, 1–16.
- Wiprächtiger-Geppert, M. (2011). Sprachliche Leistungen beurteilen. In M. Budde, S. Riegler & M. Wiprächtiger-Geppert (Hrsg.), *Sprachdidaktik* (S. 163–178). Akademie Verlag. <https://doi.org/10.1524/9783050052823.163>

Beitragsinformationen

Zitationshinweis:

Seemann, F. (2026). Beim Korrigieren diagnostizieren. Sprachdiagnostische Kompetenz von Lehrkräften der Sekundarstufe II im Forschungsprojekt DIEGO. *HLZ – Herausforderung Lehrer*innenbildung*, 9 (2), 153–174. <https://doi.org/10.11576/hlz-8454>

Eingereicht: 03.11.2025 / Angenommen: 15.04.2026 / Online verfügbar: 22.05.2026

ISSN: 2625–0675



Dieses Werk ist freigegeben unter der Creative-Commons-Lizenz CC BY-SA 4.0 (Weitergabe unter gleichen Bedingungen). Diese Lizenz gilt nur für das Originalmaterial. Alle gekennzeichneten Fremdinhalte (z.B. Abbildungen, Fotos, Tabellen, Zitate etc.) sind von der CC-Lizenz ausgenommen. Für deren Wiederverwendung ist es ggf. erforderlich, weitere Nutzungsgenehmigungen beim jeweiligen Rechteinhaber einzuholen. <https://creativecommons.org/licenses/by-sa/4.0/legalcode.de>

English Information

Title: Diagnosing While Correcting. Language Diagnostic Competence of Secondary School Teachers in the DIEGO Research Project

Abstract: Language development is considered a cross-curricular task that cuts across all subjects. In order to provide learners with appropriate and targeted support tailored to their individual language development needs, teachers must be able to accurately and precisely assess learners' language competencies. This is the fundamental prerequisite for identifying support needs and determining areas of focus.

This article presents a study on the language diagnostic competence of German teachers at the upper secondary level. The focus here is on language diagnostic competence in the area of text production.

As part of the DIEGO study, 51 teachers were asked to evaluate two student texts in accordance with their usual correction practices and to identify areas of focus for support.

The results reveal significant differences and variations, as well as a generally low level of judgment accuracy in teachers' assessments. In this article, the accuracy of teachers' assessments is measured against an expert standard using various criteria (number of errors, corrections within the text, areas for improvement). Based on the results, implications for interdisciplinary teacher education are highlighted.

Keywords: teachers' judgment; language assessment; competence assessment; text assessment; diagnosis; competence; accuracy; judgment; upper secondary school; corrective feedback; errors