BzL BEITRÄGE ZUR LEHRERINNEN- UND LEHRERBILDUNG
Zeitschrift zu Theorie und Praxis der Aus- und Weiterbildung von Lehrerinnen und Lehrern

44. Jahrgang – Heft 2/2026

E-Offprint

Hinweis zum Copyright

Die «Beiträge zur Lehrerinnen- und Lehrerbildung» (BzL) sind eine Open-Access-Zeitschrift ohne Embargo-Frist für die einzelnen Artikel.

Die Autorinnen und Autoren sind frei, die in der Zeitschrift «BzL» publizierte Version («version of record», d.h. den hier vorliegenden E-Offprint) unter der Lizenz Creative Commons Attribution 4.0 International (CC BY 4.0) über weitere Kanäle (z.B. Repositorien, Plattformen, Websites) öffentlich zugänglich zu machen.



Beiträge zur Lehrerinnen- und Lehrerbildung
In Trägerschaft der Schweizerischen Gesellschaft für Lehrerinnen- und Lehrerbildung (SGL)
Erscheint dreimal jährlich

ISSN 2296-9632

Zitiervorschlag: Mezger, R. & Lindauer, N. (2026). Formative Beurteilung diskursiver Synthesen von Studierenden mit holistischem Benchmark-Rating. *Beiträge zur Lehrerinnen- und Lehrerbildung, 44* (2), 176–193. https://doi.org/10.36950/bzl.44.2.2026.10416

www.bzl-online.ch

Editorial

Afra Sturm, Christian Brühwiler, Dorothee Brovelli, Doreen Flick-Holtsch, Bruno Leutwyler, Sandra Moroni und Markus Weil 157

Schwerpunkt

Beurteilen in der Lehrer:innenbildung

Franz Eberle Studierfähigkeit für die Studiengänge der Lehrpersonenbildung für die Volksschule 160

Res Mezger und Nadja Lindauer Formative Beurteilung diskursiver Synthesen von Studierenden mit holistischem Benchmark-Rating 176

Matthias Baumgartner und Eric Jeisy Formative Beurteilung in berufsfeld- und evidenzorientierten Lehr-Lern-Arrangements der Lehrpersonenbildung – ein Typologie- und Topologiemodell 194

Doris Moser-Frötscher, Mirjam Egli Cuenat, Wilfrid Kuster, Thomas C. Messerli, Thomas Roderer und Ingo Thonhauser Berufsspezifische Sprachkompetenz von Fremdsprachenlehrpersonen: Konzeption und Operationalisierung der PROF-L-Prüfung 212

Eric Jeisy und Charlotte Nüesch Digitale Leistungsüberprüfungen passend gestalten: Erfahrungen und Empfehlungen aus elf Pilotprojekten in der Lehrer:innenbildung 232

Forum

Maren Krempin, Simon Moses Schleimer, Anna Gronostaj und Corinna Gottmann Evidenzbasierte Schulentwicklung und Ausbildung von Change Agents: Der Doppelhelixansatz 248

Cornelia Gräsel, Stefan Keller und Anna-Katharina Praetorius Unterrichtsplanung meets Unterrichtsqualitätsforschung: Konzeption und Anwendungsmöglichkeiten des heuristischen Planungstools planTEACH 272

Thomas Rey, Eva Weingartner und Valentin Unger Adaptive Planungskompetenz im Umgang mit Heterogenität: Studentische Perspektiven auf Lerngelegenheiten im Spannungsfeld von Evidenzorientierung und Praxisbedeutsamkeit 296

Susanne Prediger, Christoph Selter, Lars Holzäpfel, Daniela Götze und Bettina Rösken-Winter Fortbildung ausrichten an Kernpraktiken: Kohärenz durch fünf zentrale unterrichtliche Anforderungssituationen im deutschlandweiten Fortbildungsprogramm QuaMath 315

Esther Brunner und Barbara Drollinger-Vetter Weiterbildungsangebote für Mathematikdidaktikerinnen und Mathematikdidaktiker: Ein Beitrag zur Professionalisierung einer Berufsgruppe 337

Marco Galle und Anja Winkler Lernzentrierte Kooperation in Modulbeschreibungen der deutschsprachigen Lehrpersonenbildung in der Schweiz 354

Rubriken

Buchbesprechungen

Plaumauer, C. & Stumpner, S. (Hrsg.). (2024). Über den Tafelrand hinaus denken! Hochschuldidaktische Impulse und Anregungen zur Fallarbeit in der Lehrer:innenbildung. Münster: Waxmann (Franziska Zellweger) 374

Caduff, C. & Weiss, S. (Hrsg.). (2025). Zwischen Theorie und Praxis. Die Qualifizierung von Dozierenden in der Berufsbildung. Bern: hep (Belinda Aeschlimann) 376

Porsch, R., Reintjes, C. & Bellenberg, G. (Hrsg.). (2025). Heterogen qualifizierte Lehrpersonen an Schulen in Deutschland, Österreich und der Schweiz. Münster: Waxmann (Sandra Matschnigg-Peer) 378

Held, O. (2024). ChatGPT im Geschichtsunterricht. Frankfurt am Main: Wochenschau (Kevin van Loon) 380

Beeli-Zimmermann, S. & Wannack, E. (Hrsg.). (2024). Schulen im Spannungsfeld analoger und digitaler Kommunikation. Bern: hep (Jurgita Lenkauskaitė) 382

Kapidzic, S., Gauly, B., Greiff, S., Martin, S., Massing, N., Zabal, A. & Rammstedt, B. (2025). Adaptives Problemlösen Erwachsener im internationalen Vergleich: Ergebnisse von PIAAC 2023. Münster: Waxmann (Franz Eberle) 384

Neuerscheinungen 386

Zeitschriftenspiegel 388

Editorial

Der Themenbereich «Beurteilen» hat für eine wirksame Lehrer:innenbildung hohe Relevanz. Entsprechend wird die Thematik seit vielen Jahren in unterschiedlichen Facetten bearbeitet, was sich auch in den BzL widerspiegelt. So befassten sich frühere Themenhefte beispielsweise mit der Förderung der diagnostischen Kompetenz von Lehrpersonen (z.B. mit der pädagogischen Diagnostik in Heft 2/2013), während sich andere Hefte mit Leistungsnachweisen in der Lehrer:innenbildung (1/2007) oder Fragen zu wissenschaftlichen Abschlussarbeiten (Doppelnummer 1/2021 und 2/2021) auseinandersetzten. Die grundlegende Aufgabe von Beurteilung im Bildungskontext liegt darin, lern- oder entwicklungsrelevante Merkmale akkurat einzuschätzen. Ein diagnostisches Urteil stellt jedoch auch in der Lehrer:innenbildung keinen Selbstzweck dar, sondern schafft eine fundierte Grundlage für die Feststellung des Lern- oder Leistungsstands der Studierenden, um daraus Massnahmen für die Förderung und die Unterstützung der professionellen Entwicklung von künftigen Lehrpersonen (formative Beurteilung) ableiten zu können oder um Selektionsentscheide (summative Beurteilung) zu fällen, zu begründen und zu legitimieren.

Das vorliegende Themenheft nimmt Fragen rund um das Beurteilen professioneller Kompetenzen in der Ausbildung von künftigen Lehrpersonen auf. Dabei werden aktuelle Themenbereiche und kontroverse Diskussionsfelder aus verschiedenen Perspektiven beleuchtet: Was wird unter Studierfähigkeit in der Lehrer:innenbildung verstanden, wie wird sie erfasst und welche Fördermassnahmen werden daraus abgeleitet? Wie lassen sich professionelle Kompetenzen von (künftigen) Lehrpersonen in unterschiedlichen fachspezifischen Kontexten beurteilen? Mit welchen (innovativen) Methoden werden relevante berufsspezifische Kompetenzen beurteilt und wie werden sie gefördert? Wie lassen sich formative Beurteilungssettings ressourcenschonend umsetzen?

Der einleitende Beitrag von **Franz Eberle** gibt einen Überblick über das System der Zulassung zu Pädagogischen Hochschulen und differenziert dabei zwischen Studierfähigkeit und der angestrebten Berufsbefähigung – zwei Aspekte, die manchmal vermischt werden. Der Beitrag führt genauer aus, welche Kompetenzbündel bei einer allgemeinen und bei einer spezifischen Studierfähigkeit zentral sind und wie sich die verschiedenen Zulassungen hinsichtlich Studienerfolg und Studienabbruch verhalten. Abschliessend zeigt der Autor auf, dass es weiterhin gute Gründe gibt, weshalb Berufsmaturand:innen eine Aufnahme- oder Ergänzungsprüfung benötigen, um zu einem Studium an einer Pädagogischen Hochschule zugelassen zu werden.

Der Beitrag von **Res Mezger und Nadja Lindauer** widmet sich dem Thema der formativen Beurteilung diskursiver Synthesen von künftigen Lehrpersonen. Sie gehen der Frage nach, wie wissenschaftliche Lese- und Schreibkompetenzen von Studienanfänger:innen systematisch erfasst und während der Ausbildung gefördert werden können. Im Beitrag wird ein formatives Assessment auf der Basis eines

holistischen Benchmark-Ratings vorgestellt und es werden darauf abgestimmte Fördermassnahmen näher beleuchtet.

Der Beitrag von **Matthias Baumgartner und Eric Jeisy** adressiert die Funktionen und Möglichkeiten, aber auch die Grenzen formativer Beurteilung in hochschuldidaktischen Lehr-Lern-Arrangements der Lehrpersonenbildung. Im Rahmen einer sportdidaktischen Interventionsstudie wird die Bedeutung von iterativen Coaching-Zyklen und formativer Beurteilung für den Performanzzuwachs nachgewiesen. Schliesslich werden Möglichkeiten für eine aufwandreduzierende Gestaltung derartiger Lehr-Lern-Arrangements (z.B. mittels automatisierter Videoproduktion) erörtert.

In ihrem Beitrag befassen sich **Doris Moser-Frötscher, Mirjam Egli Cuenat, Wilfrid Kuster, Thomas C. Messerli, Thomas Roderer und Ingo Thonhauser** mit der Förderung und der Beurteilung berufsspezifischer Sprachkompetenzen in der Ausbildung von Fremdsprachenlehrpersonen. Am Beispiel der Professionellen Sprachprüfung für Lehrpersonen (PROF-L) wird die Operationalisierung von Sprachkompetenzen in berufsrelevanten Handlungssituationen vorgestellt und das zugrunde liegende Kompetenzverständnis verortet. Zudem werden Einblicke in die institutionelle Rolle und die Nutzung von PROF-L an der HEP Vaud gewährt.

In einem Praxisbeitrag berichten **Eric Jeisy und Charlotte Nüesch** über die Entwicklung und die Erprobung von elf digitalen Leistungsüberprüfungen, die im Rahmen des PgB-Programms P-8 «Stärkung von Digital Skills in der Lehre» von swissuniversities entstanden sind. Die Evaluationsergebnisse weisen darauf hin, dass nicht das digitale Format an sich für die Qualität digitaler Leistungsüberprüfungen entscheidend ist, sondern die Abstimmung von angestrebter Kompetenz, Assessmentfunktion, Aufgabe, Medium sowie technischen und organisatorischen Bedingungen. Aus den Erfahrungen werden ein Orientierungsrahmen und konkrete Hinweise für die Gestaltung digitaler Leistungsüberprüfungen abgeleitet.

Im Forumsteil befassen sich sechs Beiträge mit unterschiedlichen Aspekten der Aus- und Weiterbildung von Lehrpersonen: Im Mittelpunkt des Beitrags von **Maren Krempin, Simon Moses Schleimer, Anna Gronostaj und Corinna Gottmann** steht ein Ansatz für eine evidenzorientierte Schulentwicklung, der sich der Analogie der Doppelhelix bedient. Dieser Ansatz zielt darauf ab, pädagogisches Schulpersonal unter Berücksichtigung aktueller Erkenntnisse der Schulqualitäts-, Schuleffektivitäts- und Schulentwicklungsforschung zu Change Agents auszubilden, die parallel zum Besuch der Fortbildung in ihren Schulen systematisch dazu beitragen, evidenzorientierte Schulentwicklung wirkungsvoll umzusetzen.

Cornelia Gräsel, Stefan Keller und Anna-Katharina Praetorius stellen in ihrem Beitrag das evidenzorientierte Unterrichtsplanungstool planTEACH vor, das zum Ziel hat, die Professionalisierung von (künftigen) Lehrpersonen bei der Unterrichtsplanung

zu unterstützen. planTEACH nimmt Befunde aus der Unterrichtsqualitätsforschung systematisch auf und bietet für jeden Planungsschritt (z.B. Lernzielbestimmung, Diagnostik der Lernvoraussetzungen, Ablauf, Aufgabenwahl) empirisch informierte, praxistaugliche Hinweise an. Zum Beitrag gehört der Online-Anhang «planTEACH: Schritt für Schritt zum guten Unterricht», der einen konkreten Leitfaden mit Fragen und Hinweisen enthält.

Thomas Rey, Eva Weingartner und Valentin Unger stellen die Konzeption und die Evaluation eines Hochschulseminars zu allgemeiner adaptiver Unterrichtsplanungskompetenz vor. Sie stützen sich dabei unter anderem auf empirische Vorarbeiten, bei denen Kriterien zur Gestaltung und Reflexion adaptiver Unterrichtsentwicklung erarbeitet und validiert wurden. Im Seminar wandten Studierende der Sekundarstufe I und II ihr im Seminar erworbenes Wissen auf reale Unterrichtsplanungen an. Die Evaluation zeigt, dass sie ihren Lerngewinn als hoch und praxisbedeutsam einschätzten.

Susanne Prediger, Christoph Selter, Lars Holzäpfel, Daniela Götze und Bettina Rösken-Winter gehen in ihrem Beitrag auf die Frage ein, was eine Ausrichtung an typischen unterrichtlichen Anforderungssituationen sowie (Kern-)Praktiken für die Gestaltung von effektiven Fortbildungen bedeutet. Sie nehmen zunächst sowohl aus generischer als auch aus fachspezifischer Perspektive eine theoretische Klärung vor und identifizieren dann fünf mathematikdidaktische Anforderungssituationen, die in das Fortbildungsprogramm QuaMath integriert wurden.

Esther Brunner und Barbara Drollinger-Vetter stellen ein Entwicklungsprojekt vor, das sich an Mathematikdidaktiker:innen richtet. Sie unterscheiden zunächst zwischen der für die Lehrpersonenbildung typischen doppelten Professionalisierung und einer kollektiven sowie individuellen Professionalisierung. Mit Blick auf Letztere führten sie eine Befragung bei amtierenden Mathematikdidaktiker:innen durch, um den Weiterbildungsbedarf genauer eruieren zu können. Abschliessend stellen sie verschiedene Weiterbildungsangebote vor.

Marco Galle und Anja Winkler führten zur Frage, inwiefern sich in den öffentlich zugänglichen Modulbeschreibungen von zwölf Pädagogischen Hochschulen der Deutschschweiz das Konzept der lernzentrierten Kooperation zeigt, eine quantitative Dokumentenanalyse durch. Diese Dokumentenanalyse ergänzten sie um eine Netzwerkkarte, um die Verhältnisse von Kooperation und Lernzentrierung besser abbilden zu können. Damit können sie zeigen, dass Kooperation in den Modulbeschreibungen primär als kommunikative Tätigkeiten beschrieben wird und dass die Beschreibungen selten gleichzeitig auch mit Bezug zu Lernzentrierung erfolgen. Daraus leiten die Autor:innen einen Handlungsbedarf für Pädagogische Hochschulen ab.

Afra Sturm, Christian Brühwiler, Dorothee Brovelli, Doreen Flick-Holtsch, Bruno Leutwyler, Sandra Moroni und Markus Weil

Formative Beurteilung diskursiver Synthesen von Studierenden mit holistischem Benchmark-Rating

Res Mezger und Nadja Lindauer

Zusammenfassung Diskursive Synthesen sind ein wesentlicher Bestandteil von wissenschaftlichen (Abschluss-)Arbeiten, wie sie die Studierenden an Pädagogischen Hochschulen im Zuge ihrer Qualifizierung für die akademische Praxis verfassen müssen. Die Erarbeitung diskursiver Synthesen erfordert ausgebaute wissenschaftliche Lese- und Schreibkompetenzen, die bei den Studierenden nur bedingt vorausgesetzt werden können. Um sie während der Ausbildung gezielt zu fördern, bedarf es einer systematischen Erfassung des Lernstands bei den Studienanfänger:innen. Im Beitrag wird ein formatives Assessment, das auf dem holistischen Benchmark-Rating basiert, vorgestellt und ausblickend auf daran anknüpfende Fördermassnahmen eingegangen.

Schlagwörter diskursive Synthese – wissenschaftliches Lesen und Schreiben – formative Beurteilung – holistisches Benchmark-Rating

Formative assessment of students' discursive syntheses with holistic benchmark rating

Abstract Discursive syntheses are a central component of academic (thesis) papers, as required of students at universities of teacher education. Generating discursive syntheses demands advanced academic reading and writing skills, however, that cannot be taken for granted at the beginning of studies. To foster these skills effectively throughout the study programme, a systematic assessment of first-year students' learning progress is essential. This article introduces a formative assessment based on a holistic benchmark rating and offers an outlook on corresponding support strategies.

Keywords discursive synthesis – academic reading and writing – formative assessment – holistic benchmark rating

DOI 10.36950/bzl.44.2.2026.10416 | **ISSN** 2296-9632

Dieses Werk steht unter der Lizenz Creative Commons Namensnennung 4.0 International.

1 Ausgangslage

Die Lehrer:innenbildung in der Schweiz erfolgt über ein Studium an einer Pädagogischen Hochschule und beinhaltet die Qualifikation sowohl für die berufliche als auch für die akademische Praxis. Insbesondere bei der Qualifizierung für die akademische Praxis steht das Verfassen einer wissenschaftlichen Qualifikationsarbeit im Zentrum. Wie Moroni, Reusser und Weil (2021) in ihrem Überblick über die wissenschaftlichen Abschlussarbeiten an Schweizer Pädagogischen Hochschulen aufzeigen, müssen die Studierenden in diesem Kontext selbstständig eine berufsrelevante Fragestellung systematisch und nach wissenschaftlichen Kriterien bearbeiten, wobei die spezifischen Anforderungen in Abhängigkeit des Studiengangs variieren (z.B. Bachelorarbeit versus Masterarbeit). An allen Schweizer Pädagogischen Hochschulen werden die Studierenden mit Kursen zu wissenschafts- und forschungsmethodischen Kompetenzen auf das Verfassen der Abschlussarbeit vorbereitet – je nach Hochschule oder Studiengang mit Unterschieden im Umfang (Moroni et al., 2021). Inwieweit in diesen Kursen auch allgemeine akademische Lese- und Schreibkompetenzen, wie sie in diesem Beitrag fokussiert werden und für das Verfassen von Abschlussarbeiten unabdingbar sind, explizit vermittelt werden, bleibt unklar (vgl. Kurzporträts der einzelnen Pädagogischen Hochschulen in BzL 1/2021). Festzuhalten ist, dass die Studierenden zu Beginn des Studiums noch nicht über diese Kompetenzen verfügen. Die (Fach-)Maturitätsarbeit ist dafür «nur» eine gute Vorbereitung (Sturm & Mezger, 2013).

In diesem Beitrag wird mit der Schreibkompetenzabklärung der Pädagogischen Hochschule FHNW ein formativ ausgerichtetes Assessment vorgestellt, mit dem Studienanfänger:innen eine erste Rückmeldung dazu erhalten, wo sie im Erwerb zentraler akademischer Lese- und vor allem Schreibkompetenzen stehen. Wir fokussieren dabei die Frage, wie sich ein holistisches Benchmark-Rating als produktbasiertes Selbst- und Fremdbeurteilungsverfahren mit formativer Ausrichtung umsetzen lässt, und berichten zentrale Ergebnisse aus aktuellen Durchführungen. Abschliessend wird der Blick auf Massnahmen zur Förderung und Unterstützung des Erwerbsprozesses gerichtet, wie sie in den Studiengängen der Pädagogischen Hochschule FHNW umgesetzt sind.

2 Diskursive Synthesen

Wesentliches Merkmal von wissenschaftlichen (Abschluss-)Arbeiten sind diskursive Synthesen. Sie kennzeichnet, dass ausgehend von einer eingegrenzten Fragestellung Informationen aus mehreren mehrheitlich schriftlichen Quellen miteinander in Beziehung gesetzt werden, um zu einer fundierten Antwort zu gelangen (Nelson & King, 2023). Typischerweise finden sich diskursive Synthesen im Theorieteil bzw. bei der Darstellung des Forschungsstands. Mit Blick auf das Produkt lässt sich festhalten, dass diskursive Synthesen keine blosse Aneinanderreihung von Zusammenfassungen

einzelner Quellen ohne Bezugnahme zueinander und keine blosse Reproduktion von Bestehendem darstellen. Vielmehr handelt es sich um Texte mit einer eigenen Makrostruktur, die Inhalte aus anderen Texten aufnehmen und mit Fokus auf die Erreichung der eigenen Diskursziele präsentieren (Nelson & King, 2023; Rickert & Philipp, 2022; Sturm, 2021). Auf Prozessebene resultieren diskursive Synthesen nicht aus einer linearen Herangehensweise, sondern erfordern ein rekursives Vorgehen. Dabei wird das Quellenmaterial in interaktiven Lese- und Schreibprozessen durch Operationen des Zerlegens, Auswählens und Verbindens transformiert (Nelson & King, 2023; Schoor et al., 2021; Vandermeulen, Van Steendam, De Maeyer, Lesterhuis & Rijlaarsdam, 2024). Das Verfassen einer diskursiven Synthese kombiniert folglich zahlreiche verschiedene Prozesse und ist entsprechend komplex und anspruchsvoll.

Studierende werden den hohen Anforderungen, die mit der Erstellung von diskursiven Synthesen einhergehen, nur bedingt gerecht. So zeigen vorliegende Untersuchungen übereinstimmend, dass Studierenden die Integration von Inhalten aus verschiedenen Quellen in einen Text mit eigener Makrostruktur schwerfällt (z.B. Boscolo, Arfé & Quarisa, 2007; Granado-Peinado et al., 2023; Luna, Villalón, Martínez-Álvarez & Mateos, 2022; Luna, Villalón, Mateos & Martín, 2020; Mateos et al., 2018). Es gilt daher, die Studierenden während der Ausbildung im Erwerb der erforderlichen Kompetenzen zu unterstützen. Voraussetzung für die Entwicklung und die Auswahl von gezielt auf die jeweiligen Bedürfnisse abgestimmten Förderangeboten bildet eine vorgängige Beurteilung des Lernstands, wie sie im nächsten Abschnitt beleuchtet wird.

3 Beurteilung von Kompetenzen für das Verfassen diskursiver Synthesen

3.1 Formative Beurteilung

Beurteilungen können unterschiedliche Funktionen haben. Mit Blick auf das Lernen haben formative Beurteilungen einen besonderen Stellenwert. Ihre Funktion besteht darin, den Lernprozess auf die Erreichung eines bestimmten Ziels hin zu steuern – so etwa auf den Erwerb von akademischen Lese- und Schreibkompetenzen für das Verfassen von diskursiven Synthesen. Eine zentrale Rolle spielen dabei die Erfassung des aktuellen Lernstands und die Formulierung eines davon ausgehenden Feedbacks, das den Lernenden das angestrebte Ziel, die bisherige Entwicklung und mögliche Schritte zur Verringerung der Diskrepanz zwischen der gezeigten und der anvisierten Leistung verdeutlicht (Hattie & Timperley, 2007). Neuere Konzeptionen fassen formatives Feedback darüber hinaus als einen interaktiven Prozess, bei dem die Lernenden nicht nur passiv eine Rückmeldung von den Lehrenden empfangen, sondern eine aktivere Rolle übernehmen, indem sie sich mit dem Feedback auseinandersetzen und die richtigen Schlüsse daraus ziehen (Scherer, Graham & Busse, 2024). Diesem Anspruch kann eine Kombination einer Fremd- mit einer Selbstbeurteilung, im Rahmen derer die Lernenden selbst angeleitet eine Einschätzung ihrer Leistung vornehmen,

Rechnung tragen. Fremd- und Selbstbeurteilung können dabei auf vier von Hattie und Timperley (2007) differenzierten Ebenen ansetzen: Produkt, Prozess, Selbstregulation und Selbst. Im vorliegenden Beitrag steht das Produkt in Form eines verfassten Textes im Fokus.

3.2 Beurteilung von Texten

Bei der Beurteilung von Texten lässt sich zwischen analytisch und holistisch ausgerichteten Verfahren unterscheiden. Im deutschsprachigen Raum stehen sowohl in der Schreibdidaktik als auch in der Schreibforschung vor allem Verfahren mit einer hauptsächlich analytischen Ausrichtung im Fokus, so auch bei der Beurteilung diskursiver Synthesen (z.B. Decker & Siebert-Ott, 2019; Rickert & Philipp, 2022; Sturm & Mezger, 2016). Analytisch ausgerichtete Verfahren sind dadurch gekennzeichnet, dass einzelne Textmerkmale über entsprechende Kriterien erfasst und die Texte kriterienweise durchgearbeitet werden (Lindauer, 2024; Weigle, 2002). Verbreitete Kriterien zur Beurteilung von Merkmalen diskursiver Synthesen stellen etwa das Vorhandensein zentraler Inhalte aus den vorliegenden Quelltexten, deren korrekte Wiedergabe oder deren textübergreifende Organisation und Verknüpfung dar (Granado-Peinado et al., 2023; Luna et al., 2022; Mateos et al., 2018; Rickert & Philipp, 2022; Sturm & Mezger, 2016). Im Unterschied zu analytischen Verfahren erfolgt bei holistischen keine separate Beurteilung einzelner Textmerkmale, sondern eine globale Beurteilung des Textes. Grundlage dafür bilden ebenfalls vordefinierte Kriterien, die in einem Globalurteil vereint werden (Weigle, 2002). Neben klassischen holistischen Verfahren, bei denen die Urteilsbildung mit primärer Bezugnahme auf Kriterien erfolgt, existieren weitere Varianten, die eine Urteilsbildung anhand des Vergleichs mit anderen Texten vorsehen (Bouwer, Koster & van den Bergh, 2023; Bouwer, Lesterhuis, Smedt, Keer & De Maeyer, 2024). Dazu zählt das sogenannte holistische Benchmark-Rating, bei dem die zu beurteilenden Texte mit Benchmark-Texten zu vergleichen sind. Diese als erste Referenz zu berücksichtigenden Benchmark-Texte werden kriteriengeleitet ausgewählt und repräsentieren unterschiedliche Leistungsniveaus, die auch in Form von Deskriptoren verdeutlicht werden (Feenstra, 2014; Lindauer, 2024). Holistische Benchmark-Ratings haben sich in verschiedenen Untersuchungen als reliabel und valide erwiesen (Bouwer et al., 2023; Bouwer et al., 2024) und sind entsprechend in der internationalen Schreibforschung, auch spezifisch zur Beurteilung von Synthesen (Raedts, Van Steendam, Grez, Hendrickx & Masui, 2017; Vandermeulen et al., 2020; Van Steendam et al., 2022), fest etabliert. Dabei werden sie häufig für den Wirksamkeitsnachweis durchgeführter Schreibfördermassnahmen genutzt und weisen damit summativen Charakter auf. Zum Einsatz holistischer Benchmark-Ratings für formative Beurteilungen liegen erst vereinzelt Vorschläge vor (Sturm, Lindauer & Sommer, 2023). Für den Hochschulkontext bzw. das wissenschaftliche Schreiben sind uns keine entsprechenden Arbeiten bekannt. Die Implementation des holistischen Benchmark-Ratings als neues Beurteilungsverfahren bei der Schreibkompetenzabklärung stellt folglich ein innovatives Vorgehen dar und ermöglicht den PH-Studierenden einen Einblick in ein alternatives Verfahren.

4 Schreibkompetenzabklärung an der Pädagogischen Hochschule FHNW

4.1 Hintergrundinformationen, Ziele und Fragestellungen

Die Schreibkompetenzabklärung wird an der Pädagogischen Hochschule FHNW seit zwanzig Jahren mit Studienanfänger:innen durchgeführt (ausführlich dazu Sturm, 2009; Sturm & Mezger, 2016). Seit 2012 ist sie in allen Studiengängen der Pädagogischen Hochschule FHNW obligatorisch.[1] Im Durchschnitt nehmen jedes Jahr über 600 Erstsemestrige teil.

Mit der Schreibkompetenzabklärung werden primär zwei Ziele verfolgt: Zum einen sollen die Studierenden ein erstes Feedback zu zentralen akademischen Lese- und Schreibkompetenzen als Teil der Studierfähigkeit erhalten und sich auf dieser Basis gegebenenfalls für ein freiwilliges Förderangebot im kommenden Semester entscheiden. Zum anderen besteht ein wichtiges didaktisches Ziel darin, die Studierenden zu Beginn ihres Erwerbsprozesses des wissenschaftlichen Lesens und Schreibens systematisch mit der Textsorte «Diskursive Synthese» und mit einem elaborierten Verfahren für die Selbstbeurteilung eines Textes zu konfrontieren.

Die Schreibkompetenzabklärung besteht aus drei Teilen: Lese-/Schreibaufgabe, Selbstbeurteilung und Fremdbeurteilung (vgl. Abschnitt 4.2). Im Laufe der Jahre wurde die Schreibkompetenzabklärung sukzessive weiterentwickelt und zuletzt 2024 das holistische Benchmark-Rating als Selbst- und Fremdbeurteilungsverfahren mit formativer Ausrichtung eingeführt. Dazu liegen nun Daten aus zwei Durchführungen vor, die im vorliegenden Beitrag mit drei Fragestellungen untersucht werden:

1) Wie fallen die Beurteilungen der von den PH-Studierenden im ersten Semester verfassten Texte aus?
2) Wie hängen die Selbstbeurteilungen der Studierenden und die Fremdbeurteilungen der PH-Fachpersonen zusammen?
3) Wie schätzen die Studierenden die Selbst- und die Fremdbeurteilung im Vergleich hinsichtlich ihrer Nützlichkeit ein?

Im Folgenden wird der Aufbau der Schreibkompetenzabklärung mit Fokus auf das Beurteilungsverfahren vorgestellt, bevor in Abschnitt 4.5 die Ergebnisse zu den Fragestellungen präsentiert werden.

[1] Eine Ausnahme bildet der Studiengang Sonderpädagogik. Hier ist die Schreibkompetenzabklärung als freiwilliges Angebot etabliert.

4.2 Aufbau der Schreibkompetenzabklärung

4.2.1 Teil 1: Lese-/Schreibaufgabe

Den *ersten Teil* der Schreibkompetenzabklärung bildet die zu bearbeitende *Lese-/Schreibaufgabe.* Sie besteht darin, eine diskursive Synthese auf der Basis von drei Originaltexten anzufertigen. Zum Einsatz kommt eine im PH-Studium situierte Aufgabe (vgl. Abbildung 1), die in leicht adaptierter Form aus dem Projekt «LUDIS» übernommen wurde (Lindauer & Mezger, in Vorbereitung).

> In einer erziehungswissenschaftlichen Seminararbeit bearbeitest du das Thema «Herausforderungen für Lehrpersonen im Schulalltag». Nachdem du im ersten Kapitel der Arbeit einen Überblick über verschiedene mögliche Herausforderungen gegeben hast, behandelst du in den folgenden Kapiteln einige spezifische Herausforderungen genauer, und zwar in Form von argumentativen Synthesen, die sich auf wissenschaftliche Argumente abstützen. Du beginnst mit der spezifischen Herausforderung «Unterrichtsstörungen» und vertrittst in diesem Kapitel den Standpunkt, dass es wichtig ist, Lehrpersonen zum einen über dieses Thema aufzuklären und zum anderen im Umgang damit zu schulen. Diesen Standpunkt unterstützt du mit Argumenten aus den drei vorgegebenen Quelltexten.

Abbildung 1: Lese-/Schreibaufgabe (Variante «Unterrichtsstörungen»).

Zurzeit existieren zwei Varianten der Aufgabe, die thematisch beide Herausforderungen von Lehrpersonen im Schulalltag adressieren, einerseits «Unterrichtsstörungen» (vgl. Abbildung 1), andererseits «Mobbing». Leitend bei der Wahl dieser beiden Unterthemen war, dass sie für den zukünftigen Lehrberuf auf allen Schulstufen relevant sind und eine wissenschaftliche Auseinandersetzung erlauben. Die Aufgabe ist so konzipiert, dass die Studierenden 60 bis 90 Minuten für die Bearbeitung benötigen. Dabei werden die Studierenden darauf hingewiesen, dass bei der knappen Zeit keine perfekten Texte erwartet werden, die in der Regel auch eine sorgfältige Überarbeitung voraussetzen. Neben der Zeit ist die Textlänge beschränkt: Die Synthese darf maximal 700 Wörter umfassen. Diese beiden Vorgaben ergeben sich aus dem Umstand, dass die Schreibkompetenzabklärung für die Studierenden nicht mehr als zwei Stunden der Selbstlernzeit (vgl. Abschnitt 4.3) beanspruchen darf und für die Durchführung der Selbstbeurteilung sowie die Verarbeitung der Fremdbeurteilung noch Zeit reserviert werden muss.

Bei den drei Originaltexten, die in der Lese-/Schreibaufgabe eingesetzt werden, handelt es sich um gekürzte Ausschnitte aus wissenschaftlichen Beiträgen, die das jeweilige Thema aus unterschiedlichen Perspektiven beleuchten und insgesamt 14 (relevante) Argumente enthalten, die eher komplementären und weniger konträren Charakter aufweisen. Die komplementäre Ausrichtung ist praxisnah, geht es doch bei pädagogisch-didaktischen Fragen meist um eine Differenzierung und weniger um eine Entscheidung für eine spezifische Position. So besteht mit Blick auf die eingesetzten Aufgaben ein weitreichender Konsens, dass es sich etwa bei Unterrichtsstörungen um eine He-

rausforderung für Lehrpersonen handelt. Inwiefern beispielsweise das Verhalten der Schüler:innen bzw. dasjenige der Lehrpersonen dazu beiträgt, kann auf der Basis von Quellen differenziert herausgearbeitet werden. Die Textmenge beträgt für die Variante «Unterrichtsstörungen» (ohne Literaturverzeichnis) 4233 Zeichen bzw. 629 Wörter, für die Variante «Mobbing» 4364 Zeichen bzw. 671 Wörter. Die Lesbarkeitsindizes (LIX), die einen Orientierungspunkt bezüglich der Komplexität auf der Textoberfläche liefern, weisen mit 54.84 bzw. 51.1 jeweils einen mittleren Schwierigkeitsgrad aus (Lenhard & Lenhard, 2014–2022).

4.2.2 Teil 2: Selbstbeurteilung

Zwei bis drei Wochen nach der Lese-/Schreibaufgabe findet *Teil 2* der Schreibkompetenzabklärung statt, in dem die Studierenden eine *Selbstbeurteilung* ihrer Lösungen mittels holistischen Benchmark-Ratings vornehmen. Dazu werden den Studierenden zuerst aufgabenspezifische Kriterien vorgestellt, die eine wichtige Grundlage des Benchmark-Ratings darstellen. Im Fokus stehen dabei die folgenden vier Kriterien, die für die Aufgabe als besonders relevant eingestuft werden und in der Schreibforschung zu diskursiven Synthesen verbreitet zum Einsatz kommen (vgl. Abschnitt 3.2): Nennung des eigenen Standpunktes, Aufnahme der zentralen Argumente aus den Quelltexten, Korrektheit der Wiedergabe der Argumente und Zusammenführung der Argumente in einen eigenständigen Text. Weitere Kriterien betreffen den Aufbau, die Kohärenz und die Nachvollziehbarkeit des Textes sowie die Quellenangaben. Kein Kriterium bildet hingegen etwa die sprachformale Korrektheit.

In einem zweiten Schritt erhalten die Studierenden die für die Selbstbeurteilung erforderlichen Benchmark-Texte. Es handelt sich dabei um drei authentische Texte von PH-Studierenden, die unterschiedliche Qualitätsstufen des wissenschaftlichen Schreibens (stark, mittel, schwach) repräsentieren und als primäre Referenz bei der Einordnung des eigenen Textes herangezogen werden. Die Benchmark-Texte entstammen dem Korpus des Projekts «LUDIS», in dem die beiden Aufgabenvarianten zu Unterrichtsstörungen und Mobbing in ihrer ursprünglichen Form ein erstes Mal bei Erstsemestrigen der Pädagogischen Hochschule FHNW eingesetzt wurden (Lindauer & Mezger, in Vorbereitung). Der Auswahl der Benchmark-Texte liegt ein mehrstufiges, empirisches Verfahren zugrunde (ausführlicher dazu z.B. Lindauer, 2024): Ausgehend von den angeführten Kriterien erfolgte zunächst die kriteriengeleitete Auswahl eines einzelnen Ankertextes von mittlerer Qualität, der in einem anschliessenden Pre-Rating als Referenz bei der Beurteilung einer Teilstichprobe der Texte diente. Anhand der Werte aus dem Pre-Rating wurden dann drei Benchmark-Texte von hoher, mittlerer und geringer Qualität (ca. M, $M - 2\,SD$ und $M + 2\,SD$) bestimmt und um Deskriptoren in Form von Beschreibungen der Stärken und Schwächen entlang der vordefinierten Kriterien ergänzt. Abbildung 2 zeigt den mittleren Benchmark-Text für die Aufgabenvariante «Unterrichtsstörungen» sowie die Deskriptoren zu den höher gewichteten Kriterien.

Benchmark-Text 2

Herausforderungen für Lehrpersonen im Schulalltag

Lehrpersonen treffen im Schulalltag immer wieder schwierige Situationen an, in welchen sie gezwungen sind, flexibel reagieren zu können. Ein wichtiges Thema sind die Störungen im Unterricht. Diese können sowohl von den Schülern als auch von den Lehrpersonen ausgehen und führen oftmals zu einer Unterbrechung des Unterrichtes. Aufgrund der Unterrichtsstörungen verliert man folglich viel Lernzeit und der Lehr-Lern-Prozess wird unterbrochen. Klar ist, dass die Folgen einer solchen Störung durchaus schwerwiegend sein können.

Damit es nicht so weit kommt, ist es unerlässlich, die Lehrfachkräfte über solch wichtige Thematiken aufzuklären und zu schulen. Wenn die Lehrpersonen über potentiell schwierige Situationen informiert werden, sind sie besser darauf vorbereitet und können demnach in solchen Situationen professioneller handeln. Auch ist die Führung einer Klasse durch die Lehrperson der unter anderem bedeutsamste Faktor des Unterrichtserfolges (Wang et al. 1993, im Überblick Haag & Streber 2012, Helmke 2009). Daher ist es wichtig, dass der/die Lehrende ein Arbeitsklima schafft, in welchem die Schüler ungestört und möglichst konzentriert an einer Lernaufgabe arbeiten können, um die aktive Lernzeit zu gewährleisten und die Unterrichtsstörungen zu minimieren. Ein weiteres wichtiges Thema an einer Fortbildung ist der Umgang mit den Störungen. Für die Schülerinnen und Schüler ist ein transparentes Regelsystem notwendig, um beispielsweise mit den Erwartungen der Lehrperson vertraut zu werden. Somit haben sie einen Leitfaden und ein besseres Verständnis davon, inwiefern die Erwartungen mit den Beobachtungen und dem Feedback der Lehrperson zusammenhängen (Emmer & Stough 2001; Evertson & Harris 1992). Auch wichtig in Bezug auf die Schülerleistungen ist laut Neuenschwander (2006) die Flexibilität und Adaptivität einer Lehrperson. Wenn diese flexibel mit unerwarteten Situationen umgehen kann, kommt mehr Akzeptanz von den Schülern und das Verhältnis zwischen Lehrperson und Schülern wird somit gestärkt.

Zusammengefasst kann man sagen, dass Weiterbildungen der Lehrfachkräfte das Unterrichtsklima positiv beeinflussen können, da sie vertrauter mit gezielten Situationen gemacht werden und ihr Handeln somit anpassen können. Unterrichtsstörungen können durch die genauere Auseinandersetzung mit gewissen Themen verringert werden und es kann ein grösserer Lernerfolg entstehen. (#5293)

Inhalt

1) Der eigene Standpunkt wird **implizit** genannt: Zu Beginn des zweiten Abschnitts werden Aufklärung und Schulung mit Blick auf «solch wichtige Thematiken» genannt, ohne konkreten Bezug zu den Unterrichtsstörungen.
2) Etwas **mehr als die Hälfte** der zentralen Argumente aus den drei QT werden vorgebracht. Erkennbar sind:
   3. Verhalten der Lernenden [QT 1]
   4. Verhalten der Lehrenden [QT1]
   5. Auswirkungen auf Lernzeit [QT1]
   7. effektive Klassenführung [QT2, QT3]
   8. lernunterstützende Unterrichtsbedingungen [QT3]

   Eine Beschreibung des Phänomens fehlt (Definition & Frequenz).
3) Die zentralen Argumente werden **teilweise unpräzise bzw. übertrieben** dargestellt: So wird in den QT nirgends erwähnt, dass die Störungen oftmals zu Unterbrechungen des Unterrichts führen.

Synthese

4) Die Argumente aus den QT sind **nicht** in einem eigenständigen Text zusammengeführt, sondern die QT werden nacheinander abgearbeitet: QT1, QT2, QT3.

Abbildung 2: Benchmark-Text 2 und Deskriptoren (nur höher gewichtete Kriterien, Variante «Unterrichtsstörungen»; QT = Quelltext[e]).

Mit Blick auf die höher gewichteten Kriterien (vgl. Abbildung 2 und Tabelle 1) zeigt sich für den mittleren Benchmark-Text, dass er den eigenen Standpunkt (implizit) und gut die Hälfte der zentralen Argumente umfasst, diese jedoch stellenweise inkorrekt darstellt und in der Reihenfolge der Quelltexte wiedergibt, sodass kein Text mit eigener Makrostruktur gegeben ist. Erst Benchmark-Text 3 stellt eine eigentliche diskursive Synthese dar, bei der die Argumente inhaltlich über die Quelltexte hinweg gruppiert sind. Anzumerken ist, dass Benchmark-Text 3 eine für Erstsemestrige gelungene Realisierung bildet, nicht jedoch eine musterhafte Umsetzung.

Tabelle 1: Höher gewichtete Kriterien und ihre Realisierungen in den Benchmark-Texten 1 bis 3

Kriterien	**Benchmark-Text 1**	**Benchmark-Text 2**	**Benchmark-Text 3**
eigener Standpunkt	ansatzweise	implizit	vorhanden
zentrale Argumente	1 Argument explizit, 3 implizit	5 von 8 Argumenten	7 von 8 Argumenten
inhaltliche Korrektheit	vereinzelt unpräzise	teilweise unpräzise	weitgehend korrekt
Synthese	Zusammenfassung nur eines QT	kein eigenständiger Text: Aneinanderreihung der QT	Argumente werden gemäss Inhalt strukturiert

Anmerkung: QT = Quelltext(e).

Im Rahmen der Vorbereitung auf die Selbstbeurteilung werden die Studierenden dazu aufgefordert, die Kriterien sowie die Benchmark-Texte und deren Deskriptoren aufmerksam zu lesen. Um sicherzustellen, dass die Studierenden das Verfahren verstehen, werden die Benchmark-Texte und die Deskriptoren in Videodemonstrationen erläutert und das Vorgehen beim Vergleich mit den drei Benchmark-Texten und bei der Zuordnung zu einem Benchmark-Text wird an einem Beispiel vorgeführt. In der Folge findet die eigentliche Selbstbeurteilung statt: Die Studierenden lesen zunächst ihren eigenen Text nochmals durch, bevor sie diesen mit den drei Benchmark-Texten einschliesslich der Deskriptoren vergleichen und dem am ähnlichsten erscheinenden Text zuweisen.

Nach der Beurteilung werden den Studierenden Empfehlungen präsentiert, die gemäss Qualitätsstufe ausdifferenziert sind und den Besuch eines Förderangebots an der Pädagogischen Hochschule FHNW fokussieren. Die Empfehlungen sind mit dem Hinweis versehen, dass die Leistung in einer einzelnen Aufgabe immer mit Vorsicht zu behandeln ist und nur sie selbst beurteilen können, ob ihr Text typisch für sie ist oder nicht. Für alle beschriebenen Bestandteile der Selbstbeurteilung stehen den Studierenden insgesamt 30 bis 45 Minuten zur Verfügung.

4.2.3 Teil 3: Fremdbeurteilung

Der *dritte Teil* der Schreibkompetenzabklärung betrifft die *Fremdbeurteilung.* Alle Studierendentexte werden von zwei PH-Mitarbeitenden (dem Autor und einer weiteren Person) unabhängig voneinander mithilfe der gleichen Benchmark-Texte, die auch der Selbstbeurteilung zugrunde liegen, eingeschätzt. Im Unterschied zur Selbstbeurteilung, bei der die Studierenden lediglich eine Zuordnung zu einem der Benchmark-Texte vornehmen müssen, wird bei der Fremdbeurteilung eine breite Rating-Skala verwendet und von den exakten Punktzahlen, die die Benchmark-Texte im durchgeführten Pre-Rating erhalten haben, ausgegangen. Die Punktzahlen der Benchmark-Texte zur Aufgabe «Unterrichtsstörungen» etwa betragen 74, 103 und 124. In den bisherigen zwei Durchführungen der Schreibkompetenzabklärung war die Konsistenz der Urteile der beiden Rater:innen[2] mit einer ICC(C, 2) von .775 bzw. .820 gut (Cicchetti, 1994). Unter dieser Voraussetzung lässt sich als finales Mass der Fremdbeurteilung der Mittelwert aus den zwei Urteilen berechnen. Um die Fremdbeurteilungen mit den Selbstbeurteilungen in Beziehung setzen und an die Studierenden übermitteln zu können, werden die Punktzahlen aus der Fremdbeurteilung in eine dreistufige Skala überführt. Als Cut-off wird dabei die Mitte zwischen den Punktzahlen von Benchmark-Text 1 und 2 bzw. von Benchmark-Text 2 und 3 gewählt.

Das Resultat aus der Fremdbeurteilung erhalten die Studierenden zeitnah nach der Selbstbeurteilung. Zunächst wird ihnen in einem kurzen Video das Vorgehen bei der Fremdbeurteilung erklärt, bevor sie sowohl die Punktzahl als auch die darauf basierende Zuordnung zu einem der Benchmark-Texte erfahren. Letzteres erfolgt über ein einseitiges Formular mit einer Grafik, in der Fremd- und Selbstbeurteilung zusammengefasst und damit direkt vergleichbar sind (vgl. Abbildung 3). Das Formular und die möglichen Rückschlüsse aus der Beurteilung auf die Qualität ihres Textes werden den Studierenden in einem weiteren Video erläutert. Unter Rückgriff auf die Benchmark-Texte wird ihnen etwa aufzeigt, dass Texte auf Stufe 1 und Stufe 2 in der Regel keine oder höchstens ansatzweise eine Synthese in Form eines Textes mit eigener Struktur darstellen und nur einen Teil der relevanten Inhalte einschliessen. Dazu wird ihnen zum Vergleich neben dem bereits bekannten Benchmark-Text 3 auch noch ein Lösungsvorschlag präsentiert, in dem unter anderem die Verschränkung der Inhalte aus den verschiedenen Quelltexten farblich markiert ist. Abschliessend werden die Studierenden dazu aufgefordert, ihren Text mithilfe dieser Informationen mit dem Benchmark-Text der zugewiesenen Qualitätsstufe zu vergleichen und zu ermitteln, wie die Kriterien in ihrem Text umgesetzt sind. Abgesehen von Fremd- und Selbstbeurteilung erhalten die Studierenden anhand des ihnen abgegebenen Formulars auch nochmals die Empfehlungen, wie sie den Studierenden nach der Selbstbeurteilung bereits präsentiert wurden (vgl. Abbildung 3). Die Bearbeitung des letzten Teils der Schreibkompetenzabklärung zur Fremdbeurteilung ist für die Studierenden mit einem zeitlichen Aufwand von circa 10 bis 15 Minuten verbunden.

---

[2] Iva Koncic und Johannes Zbinden sei an dieser Stelle herzlich für ihre Mitarbeit beim Rating gedankt.

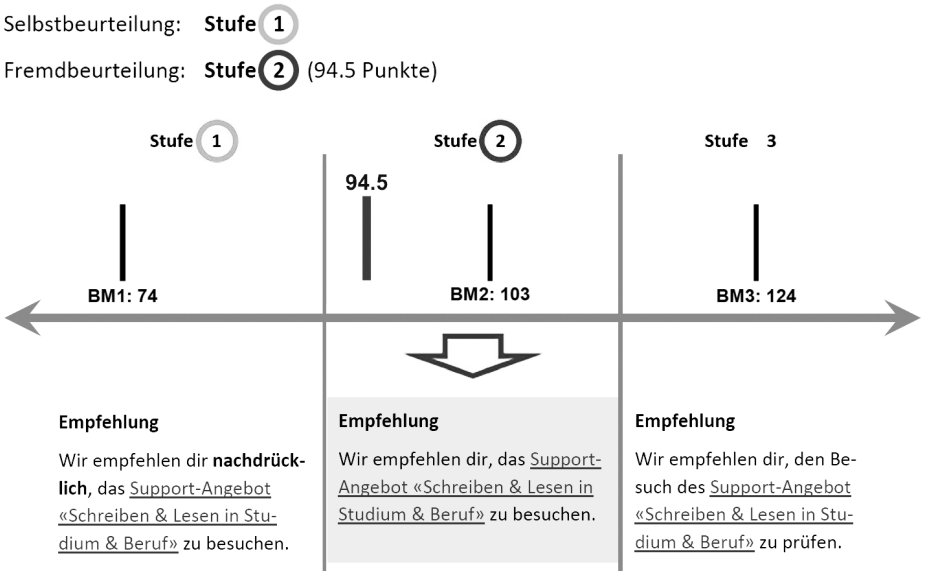


Abbildung 3: Beispiel eines Fremdbeurteilungsformulars aus der Schreibkompetenzabklärung 2025 (Ausschnitt, vereinfacht).

4.2.4 Fragebogen

Am Ende von Teil 2 und 3 wird über einen Fragebogen unter anderem erfasst, wie die Studierenden die Nützlichkeit der Selbst- bzw. Fremdbeurteilung einschätzen. Ihre Einschätzungen müssen die Studierenden über eine elfstufige Likert-Skala abgeben. Bei allen drei Teilen steht zusätzlich ein offenes Antwortfeld zur Verfügung.

4.3 Umsetzung

Die Schreibkompetenzabklärung wird von der Schreibberatung der Pädagogischen Hochschule FHNW verantwortet und durchgeführt, wobei eine enge Zusammenarbeit mit den verschiedenen Studiengangsinstituten besteht. So ist die Schreibkompetenzabklärung in allen Studiengängen in Lehrveranstaltungen integriert, die von den Studierenden üblicherweise im ersten Semester besucht werden und in denen grundlegende Kompetenzen des wissenschaftlichen Arbeitens vermittelt werden. Die Schreibkompetenzabklärung findet in der Selbstlernzeit der Lehrveranstaltungen in vorgegebenen Terminfenstern von rund zehn Tagen statt. Die Online-Durchführung auf der Lernplattform «Moodle» unterstützt diese Durchführungsform, insbesondere auch die terminliche Staffelung und die Steuerung der Bearbeitungszeiten. Da die Schreibkompetenzabklärung für die Studierenden einen obligatorischen Teil des Leistungsnachweises der jeweiligen Lehrveranstaltung darstellt, kann sichergestellt werden, dass alle die Schreibkompetenzabklärung während des Semesters absolvie-

ren. Für Studierende, die Bearbeitungsfristen nicht einhalten, gibt es eine sogenannte Nachzügler:innendurchführung.

4.4 Stichprobe

In den letzten beiden Durchführungsjahren mit dem neu eingeführten Beurteilungsverfahren haben jeweils gegen 700 Erstsemestrige der Pädagogischen Hochschule FHNW an der Schreibkompetenzabklärung teilgenommen.[3] Die Verteilung der Studierenden über die Studiengänge ist in Tabelle 2 dargestellt.

Tabelle 2: Stichprobe

Durchführungsjahr	KU	P	SEK	SP	Gesamt
2024	211 (30.6%)	361 (52.3%)	90 (13.0%)	28 (4.0%)	690 (100%)
2025	202 (30.1%)	361 (53.7%)	93 (13.8%)	16 (2.4%)	672 (100%)

Anmerkungen: KU = Kindergarten/Unterstufe, P = Primarstufe, SEK = Sekundarstufe I, SP = Logopädie & Sonderpädagogik.

Im Rahmen der Schreibkompetenzabklärung werden weder soziodemografische Merkmale noch beispielsweise die Vorbildung der Studierenden erfasst. Zu berücksichtigen ist jedoch, dass Studienanfänger:innen aufgrund ihrer Vorbildung die einführende Lehrveranstaltung, an die die Schreibkompetenzabklärung geknüpft ist, angerechnet werden kann, wodurch diese Studierenden nicht am Assessment teilnehmen.

4.5 Ergebnisse

4.5.1 Fremd- und Selbstbeurteilungen

In Bezug auf Frage 1 nach den Beurteilungen der von den PH-Studierenden im ersten Semester verfassten Texte ist in Tabelle 3 die deskriptive Statistik für die erzielten Fremd- und Selbstbeurteilungen in den Durchführungen 2024 und 2025 dargestellt. Die in der Fremdbeurteilung mit der breiten Intervallskala erzielte durchschnittliche Punktzahl liegt bei 104.2 (SD = 7.55). 69.5 bildet die minimal, 125.5 die maximal erreichte Punktzahl. Für die dreistufige Ordinalskala, in die die Punktzahlen für den Vergleich mit der Selbstbeurteilung überführt wurden (vgl. Abschnitt 4.2.3), resultiert ein Median von 2.00. Von den insgesamt 1362 beurteilten Texten fallen 81 (5.9%) in Qualitätsstufe 1, 1222 (89.7%) in Qualitätsstufe 2 und 59 (4.3%) in Qualitätsstufe 3. Bei der Selbstbeurteilung ist die Zahl an schwach und stark eingestuften Texten mit 165 (12.3%, Qualitätsstufe 1) und 313 (23.3%, Qualitätsstufe 3) deutlich höher. Mit 866 (64.4%)

[3] Für die vorliegende Studie werden die anonymisierten Daten aus dem obligatorischen Assessment (Fremd- und Selbstbeurteilung, Vergleich von Fremd- und Selbstbeurteilung sowie Teile des Fragebogens) genutzt. Die Studierenden nehmen automatisch an dieser Studie teil, ohne dass sie während des Assessments eine Beteiligung an- oder abwählen konnten.

am meisten Texte liegen aber auch hier auf der mittleren Qualitätsstufe. Entsprechend beträgt auch bei der Selbstbeurteilung der Median 2.00.

Tabelle 3: Deskriptive Statistik für die Fremd- und Selbstbeurteilung

	Mdn	*M*	*SD*
Fremdbeurteilung, breite Intervallskala (*N* = 1362)	105.00	104.20	7.55
Fremdbeurteilung, dreistufige Skala[a] (*N* = 1362)	2.00	1.98	0.32
Selbstbeurteilung, dreistufige Skala[a] (*N* = 1344[b])	2.00	2.11	0.59

Anmerkungen: [a] 1 = Qualitätsstufe 1, 2 = Qualitätsstufe 2, 3 = Qualitätsstufe 3. [b] Die Differenz zu *N* bei der Fremdbeurteilung ergibt sich dadurch, dass nicht alle Studierenden nach Teil 1 mit der Lese-/Schreibaufgabe auch Teil 2 mit der Selbstbeurteilung absolviert haben.

Zur Klärung der Frage nach dem Zusammenhang von Selbst- und Fremdbeurteilungen wurden die ordinalen Daten einer Rangkorrelationsanalyse nach Spearman unterzogen. Es zeigt sich eine signifikante, aber schwache Korrelation zwischen den Urteilen der Studierenden und der PH-Fachpersonen ($r = .189$, $p < .001$). Eine höhere Einstufung durch die Studierenden geht somit nur bedingt auch mit einer höheren Einstufung durch die PH-Fachpersonen einher und umgekehrt.

4.5.2 Einschätzung der Beurteilungsarten durch die Studierenden

Am Ende von Teil 3 zur Fremdbeurteilung wurden die Studierenden in einem Fragebogen unter anderem danach gefragt, wie nützlich sie Selbst- und Fremdbeurteilung im Vergleich einschätzen. Ihre Einschätzungen mussten die Studierenden auf einer elfstufigen Likert-Skala vornehmen: 0 = 0:100 (Fremd:Selbst), 5 = 50:50 (Fremd:Selbst), 10 = 100:0 (Fremd:Selbst). Der Median beträgt 6.00, der Mittelwert 5.70 mit einer Standardabweichung von 1.50. Mit 51.8% wählte gut die Hälfte der Studierenden einen Wert > 5 und stuft die Fremdbeurteilung damit als nützlicher ein als die Selbstbeurteilung. Ungefähr ein Drittel (34.2%) der Studierenden entschied sich für den Wert 5 und erachtet folglich die Selbst- und die Fremdbeurteilung als äquivalent. Lediglich 14.0% der Studierenden klickten einen Wert < 5 an und ziehen somit die Selbstbeurteilung der Fremdbeurteilung vor.

4.6 Diskussion

Die Anliegen des vorliegenden Beitrags bestanden darin, einerseits darzustellen, wie ein formatives Assessment, das aus einer Selbst- und einer Fremdbeurteilung besteht und auf einem holistischen Benchmark-Rating basiert, gestaltet werden kann, und andererseits zentrale Befunde aus den bisherigen zwei Durchführungen an der Pädagogischen Hochschule FHNW zu berichten (vgl. Abschnitt 1). Mit Blick auf das erste Anliegen

wurde in Abschnitt 4.2 der Aufbau der Schreibkompetenzabklärung bestehend aus den drei Teilen «Lese-/Schreibaufgabe», «Selbstbeurteilung» und «Fremdbeurteilung» mit Fokus auf die Umsetzung des holistischen Benchmark-Ratings detailliert beschrieben und begründet. Die im Rahmen des zweiten Anliegens fokussierten zentralen Befunde orientierten sich an drei Fragestellungen (vgl. Abschnitt 4.1), die im Folgenden anhand der dazu gewonnenen Ergebnisse diskutiert werden.

Die erste Fragestellung bezog sich darauf, wie die Beurteilungen der von den PH-Studierenden im ersten Semester verfassten Texte ausfallen. Dabei zeigte sich, dass den Texten sowohl in den Selbst- als auch in den Fremdbeurteilungen überwiegend eine mittlere Qualität attestiert wird. Mit Bezugnahme auf den in Abschnitt 4.2.2 dargestellten Benchmark-Text 2 als Repräsentant der mittleren Qualitätsstufe weist dieses Ergebnis darauf hin, dass die Studierenden grösstenteils Texte produzieren, die etwas mehr als die Hälfte der zentralen Argumente beinhalten, punktuell inhaltliche Unschärfen aufweisen und entlang der Quelltexte organisiert sind, das heisst, eine Aneinanderreihung von Zusammenfassungen der drei Quelltexte und in dem Sinne keine diskursive Synthese mit eigener Makrostruktur darstellen. Diese Beobachtung reiht sich in die Befunde bisheriger Untersuchungen mit Studierenden ein, in denen Texte in Form eigentlicher diskursiver Synthesen mit einem hohen Mass an sogenannter Integration bzw. intertextueller Verknüpfung ebenfalls vergleichsweise wenig verzeichnet wurden (Boscolo et al., 2007; Granado-Peinado et al., 2023; Luna et al., 2022; Luna et al., 2020; Mateos et al., 2018; Rickert & Philipp, 2022). Anzumerken ist, dass in der letzten Durchführung der Schreibkompetenzabklärung 2025 vermehrt Texte auftraten, die Hinweise auf eine Anfertigung mithilfe von KI enthalten. Diese Texte sind in der Regel durch eine höhere inhaltliche Vollständigkeit gekennzeichnet, stellen strukturell aber keine Synthesen, sondern in erster Linie aneinandergereihte Zusammenfassungen der Quelltexte dar.

In Bezug auf die zweite Fragestellung nach dem Zusammenhang von Selbst- und Fremdbeurteilung zeigte sich, dass die von den Studierenden und PH-Fachpersonen vorgenommenen Einstufungen – trotz ähnlicher Verteilung über die drei differenzierten Qualitätsstufen hinweg – nur sehr schwach korrelieren. Ein von den Studierenden als qualitativ hoch beurteilter Text wird von den PH-Fachpersonen also nur bedingt gleich eingeschätzt und umgekehrt. Dieser Befund weist darauf hin, dass eine Selbstbeurteilung nicht ausreicht, sondern zusätzlich eine Fremdbeurteilung notwendig ist, da die Studierenden ihre eigenen Schreibfähigkeiten nur bedingt adäquat einzuschätzen vermögen, wie auch aus anderen Studien hervorgeht (León, Panadero & García-Martínez, 2023).

Frage 3 zielte darauf, wie nützlich die Studierenden Selbst- und Fremdbeurteilung einschätzen. Wie an den Antworten auf die Frage nach dem Verhältnis von Selbst- und Fremdbeurteilung punkto Nützlichkeit erkennbar wird, erachten die Studierenden im Durchschnitt die Fremdbeurteilung als etwas hilfreicher als die Selbstbeurteilung.

Dieser Befund könnte darauf zurückgehen, dass sie aus dem schulischen Kontext in erster Linie Fremd- statt Selbstbeurteilungen gewohnt und in der Regel auch noch nie mit holistischen Verfahren, wie dem hier eingesetzten, in Kontakt gekommen sind (vgl. Abschnitt 3.2). Die mittels Benchmark-Ratings vorzunehmende Selbstbeurteilung dürfte für die Studierenden folglich mit Unsicherheiten verbunden gewesen sein, worauf auch der lediglich geringe Zusammenhang zwischen den beiden Beurteilungen hindeutet.

5 Förderangebote zum Schreiben diskursiver Synthesen an der Pädagogischen Hochschule FHNW

Wie es für formative Assessments charakteristisch ist (vgl. Abschnitt 3.1), zielt das im Rahmen der Schreibkompetenzabklärung eingesetzte Beurteilungsverfahren darauf, zu erfassen, wo die Studierenden im Erwerb zentraler akademischer Lese- und Schreibkompetenzen stehen, um auf dieser Grundlage darauf abgestimmte Fördermassnahmen bereitstellen bzw. auswählen zu können. Im Folgenden werden vier entsprechende Förderangebote an der Pädagogischen Hochschule FHNW kurz vorgestellt.

Eine erste Fördermassnahme, die direkt an die Schreibkompetenzabklärung anschliesst, ist ein in verschiedenen Formaten angebotener Kurs zum wissenschaftlichen Lesen und Schreiben. Wie vorangehend dargestellt, erhalten die Studierenden im Anschluss an die Selbst- und Fremdbeurteilung eine Empfehlung zum Besuch dieses Kurses und können ihn im folgenden Semester freiwillig absolvieren. In diesem Kurs haben die Studierenden die Möglichkeit, ihren im Rahmen der Schreibkompetenzabklärung verfassten Text und die dazugehörige Beurteilung individuell mit der Kursleitung zu besprechen. Ansonsten umfassen die Inhalte des Kurses neben den für das Verfassen von diskursiven Synthesen benötigten Lese- und Schreibkompetenzen weitere zentrale wissenschaftliche Kompetenzen und darauf abgestimmte praktische Hilfsmittel, etwa für das Erarbeiten einer Fragestellung oder für den Umgang mit einem Literaturverwaltungsprogramm. Dabei wird beispielsweise auch der unterstützende Einsatz von KI-Tools thematisiert. Bei der definitiven Ausgestaltung des Kursprogramms werden die Vorerfahrungen und die individuellen Lernziele der Teilnehmenden berücksichtigt.

Mit «LUDIS» steht an der Pädagogische Hochschule FHNW ein weiteres Angebot in Form eines Online-Programms zur Förderung des argumentierenden Schreibens zu mehreren Quellen zur Verfügung (Lindauer & Mezger, in Vorbereitung), das auf der Arbeit von Luna et al. (2020) aufbaut. Es besteht aus sechs Lerneinheiten, in denen wesentliche Kompetenzen für das Verfassen wissenschaftlicher diskursiver Synthesen aufgebaut werden. Dabei werden sowohl das Schreibprodukt als auch der Schreibprozess adressiert. Konkret befassen sich die Einheiten mit den folgenden Themen: Definition von argumentativen Synthesen, Argumentationsformen und Typen von Argumenten (komplementär, konträr), Vorgehensweise beim Lesen von multiplen

Quellen, Vorgehensweise beim Schreiben diskursiver Synthesen sowie Analyse des eigenen Lernprozesses. Das Lernprogramm wurde in einer Interventionsstudie hinsichtlich seiner Wirksamkeit mit Blick auf die Qualität der diskursiven Synthesen positiv evaluiert (Lindauer & Mezger, in Vorbereitung) und in der Folge in verschiedenen Studiengängen der Pädagogischen Hochschule FHNW ins Curriculum integriert.

Ein drittes wichtiges Förderangebot bildet die Website der Schreibberatung der Pädagogischen Hochschule FHNW, auf der die Studierenden und Dozierenden eine breite Auswahl an Materialien, Vorlagen und Erklärvideos zu allen Aspekten des wissenschaftlichen Lesens und Schreibens finden. Darüber hinaus können sich die Studierenden mit ihren spezifischen Fragen direkt an die Schreibberatung wenden.

Schliesslich wird in den einzelnen Studiengängen das wissenschaftliche Lesen und Schreiben in verschiedenen Kontexten thematisiert – sei dies in Lehrveranstaltungen oder in der Begleitung von schriftlichen Arbeiten. Das Ergebnis, dass die Fremdbeurteilung im Vergleich mit der Selbstbeurteilung als nützlicher wahrgenommen wird, spricht dafür, in diesen Kontexten auch Feedback-Formen zu etablieren, im Rahmen derer den Lernenden eine aktive Rolle zukommt. Dies wird in neueren theoretischen Konzeptionen von formativer Rückmeldung als zentral erachtet (Scherer et al., 2024) und ist in der hier vorgestellten Selbstbeurteilung angelegt.

6 Fazit und Ausblick

Mit der gestellten Aufgabe und dem gewählten Beurteilungsverfahren werden die mit der Schreibkompetenzabklärung verfolgten Ziele, den Studierenden ein erstes Feedback zu ihren akademischen Lese- und Schreibkompetenzen zukommen zu lassen und ihnen erste Erfahrungen mit der Textsorte der diskursiven Synthese zu ermöglichen, erreicht. Damit ergibt sich eine gute Ausgangslage für die Förderangebote zum Schreiben diskursiver Synthesen.

In Bezug auf das Feedback gilt es anzumerken, dass das Beurteilungsverfahren ausschliesslich schreibproduktbasiert erfolgt, das heisst, anhand der von den Studierenden verfassten diskursiven Synthesen Rückschlüsse auf (nicht) vorhandene akademische Lese- und Schreibkompetenzen gezogen und darauf basierende Rückmeldungen gegeben werden. Ein noch präziseres Bild des aktuellen Lernstands und ein umfangreicheres Feedback dazu an die Studierenden liessen sich über eine Erfassung der Lese- und Schreibprozesse während der Aufgabenbearbeitung erreichen. Mit Blick auf zukünftige Durchführungen der Schreibkompetenzabklärung, für die das Beurteilungsverfahren beständig weiterentwickelt werden soll, ist eine entsprechende Erweiterung in Erwägung zu ziehen. Ein weiterer zu klärender Aspekt bezieht sich etwa auf den Einsatz von KI-Assistenzsystemen beim Bearbeiten der Lese-/Schreibaufgabe. Hier stellt sich einerseits die Frage, wie im Rahmen der Schreibkompetenzabklärung mit KI-

generierten Texten umgegangen werden soll. Andererseits verdeutlicht der Befund, dass mit entsprechend produzierten Texten nur die mittlere Qualitätsstufe erreicht wird, die Notwendigkeit, die Studierenden in den Förderangeboten beim Verarbeiten von Quelltexten bzw. Verfassen von diskursiven Synthesen mit KI-Tools anzuleiten.

Literatur

Boscolo, P., Arfé, B. & Quarisa, M. (2007). Improving the quality of students' academic writing: An intervention study. *Studies in Higher Education, 32* (4), 419–438.

Bouwer, R., Koster, M. & van den Bergh, H. (2023). Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. *Assessment in Education: Principles, Policy & Practice, 30* (3–4), 302–319.

Bouwer, R., Lesterhuis, M., Smedt, F. D., Keer, H. V. & De Maeyer, S. (2024). Comparative approaches to the assessment of writing: Reliability and validity of benchmark rating and comparative judgement. *Journal of Writing Research, 15* (3), 497–518.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6* (4), 284–290.

Decker, L. & Siebert-Ott, G. (2019). Schreibend an fachlichen Diskursen partizipieren: Ergebnisse einer Interventionsstudie zur Förderung der Textkompetenzen von Lehramtsstudierenden (Wissenschaftskommunikation). In M. Steinseifer, H. Feilke & K. Lehnen (Hrsg.), *Eristische Literalität* (S. 299–326). Heidelberg: Synchron.

Feenstra, H. (2014). *Assessing writing ability in primary education. On the evaluation of text quality and text complexity* [Dissertation]. Enschede: Universität Twente.

Granado-Peinado, M., Cuevas, I., Olmos, R., Martín, E., Casado-Ledesma, L. & Mateos, M. (2023). Collaborative writing of argumentative syntheses by low-performing undergraduate writers: Explicit instruction and practice. *Reading and Writing, 36* (4), 909–936.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77* (1), 81–112.

Lenhard, W. & Lenhard, A. (2014–2022). *Berechnung des Lesbarkeitsindex LIX nach Björnson.* Dettelbach: Psychometrica. http://www.psychometrica.de/lix.html

León, S. P., Panadero, E. & García-Martínez, I. (2023). How accurate are our students? A meta-analytic systematic review on self-assessment scoring accuracy. *Educational Psychology Review, 35* (4), Artikel 106.

Lindauer, N. (2024). Effektivität und Effizienz von holistischen Benchmarkratings. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Texte schreiben in allen Unterrichtsfächern. Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 91–111). Münster: Waxmann.

Lindauer, N. & Mezger, R. (in Vorbereitung). *Lernumgebung zu diskursiven Synthesen im Studium – ein Online-Programm zur Förderung des argumentierenden Schreibens zu mehreren Quellen.*

Luna, M., Villalón, R., Martínez-Álvarez, I. & Mateos, M. (2022). Online interventions to help college students to improve the degree of integration of their argumentative synthesis. *Reading and Writing, 36* (4), 937–963.

Luna, M., Villalón, R., Mateos, M. & Martín, E. (2020). Improving university argumentative writing through online training. *Journal of Writing Research, 12* (1), 233–262.

Mateos, M., Martín, E., Cuevas, I., Villalón, R., Martínez, I. & González-Lamas, J. (2018). Improving written argumentative synthesis by teaching the integration of conflicting information from multiple sources. *Cognition and Instruction, 36* (2), 119–138.

Moroni, S., Reusser, K. & Weil, M. (2021). Wissenschaftliche Abschlussarbeiten in der schweizerischen Lehrerinnen- und Lehrerbildung: Ein Überblick. *Beiträge zur Lehrerinnen- und Lehrerbildung, 39* (1), 117–140.

Nelson, N. & King, J. R. (2023). Discourse synthesis: Textual transformations in writing from sources. *Reading and Writing, 36* (4), 769–808.
Raedts, M., Van Steendam, E., Grez, L. D., Hendrickx, J. & Masui, C. (2017). The effects of different types of video modelling on undergraduate students' motivation and learning in an academic writing course. *Journal of Writing Research, 8* (3), 399–435.
Rickert, A. & Philipp, M. (2022). Synthesen schreiben – eine textbasierte Längsschnittuntersuchung studentischer Lese- und Schreibperformanz. In M. Philipp & S. Jambor-Fahlen (Hrsg.), *Lesen: Prozess- und Produktperspektiven von der Wortebene bis zu multiplen Texten* (S. 100–127). Weinheim: Beltz.
Scherer, S., Graham, S. & Busse, V. (2024). How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. *Learning and Instruction, 93,* Artikel 101961, 1–17.
Schoor, C., Rouet, J.-F., Artelt, C., Mahlow, N., Hahnel, C., Kroehne, U. & Goldhammer, F. (2021). Readers' perceived task demands and their relation to multiple document comprehension strategies and outcome. *Learning and Individual Differences, 88,* Artikel 102018.
Sturm, A. (2009). Schreibkompetenzen und Selbsteinschätzung bei angehenden Lehrpersonen. *Zeitschrift Schreiben, 4,* 1–8.
Sturm, A. (2021). Multiple Dokumente verstehen und verarbeiten: Anforderungen und Förderansätze. *Beiträge zur Lehrerinnen- und Lehrerbildung, 39* (2), 247–259.
Sturm, A., Lindauer, N. & Sommer, T. (2023). *Lernstandserfassung «Schreiben als soziale Praxis» (7. und 8./9. Klasse).* Brugg: Pädagogische Hochschule FHNW.
Sturm, A. & Mezger, R. (2013). Wissenschaftliches Schreiben im Studium: Herausforderungen und Enkulturation. *Beiträge zur Lehrerbildung, 31* (3), 358–363.
Sturm, A. & Mezger, R. (2016). Schriftliche Quellenverarbeitung mit formativem Feedback. *Journal für LehrerInnenbildung, 16* (3), 25–29.
Vandermeulen, N., De Maeyer, S., Van Steendam, E., Lesterhuis, M., van den Bergh, H. & Rijlaarsdam, G. (2020). Mapping synthesis writing in various levels of Dutch upper-secondary education. A national baseline study on text quality, writing process and students' perspectives on writing. *Pedagogische Studiën, 97* (3), 187–236.
Vandermeulen, N., Van Steendam, E., De Maeyer, S., Lesterhuis, M. & Rijlaarsdam, G. (2024). Learning to write syntheses: The effect of process feedback and of observing models on performance and process behaviors. *Reading and Writing, 37* (6), 1375–1405.
Van Steendam, E., Vandermeulen, N., De Maeyer, S., Lesterhuis, M., Van den Bergh, H. & Rijlaarsdam, G. (2022). How students perform synthesis tasks: An empirical study into dynamic process configurations. *Journal of Educational Psychology, 114* (8), 1773–1800.
Weigle, S. C. (2002). *Assessing writing.* Cambridge: Cambridge University Press.

Autor und Autorin

Res Mezger, lic. phil., Pädagogische Hochschule der Fachhochschule Nordwestschweiz, res.mezger@fhnw.ch
Nadja Lindauer, Dr., Pädagogische Hochschule der Fachhochschule Nordwestschweiz, nadja.lindauer@fhnw.ch