

Alois Buholzer
Dorothee Brovelli
(Hrsg.)

Formatives Assessment

Perspektiven für Unterricht und
Lehrerinnen- und Lehrerbildung

Alois Buholzer, Dorothee Brovelli (Hrsg.)

Formatives Assessment

Perspektiven für Unterricht und
Lehrerinnen- und Lehrerbildung



Waxmann 2023

Münster • New York

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Print-ISBN 978-3-8309-4749-3
E-Book-ISBN 978-3-8309-9749-8
<https://doi.org/10.31244/9783830997498>

Waxmann Verlag GmbH, Münster 2023
Steinfurter Straße 555, 48159 Münster
www.waxmann.com
info@waxmann.com

Umschlaggestaltung: Anne Breitenbach, Münster
Satz: Mario Moths, Marl

Dieses E-Book ist open access unter der Creative-Commons-Lizenz CC-BY-NC-SA 4.0 verfügbar.



Diese Lizenz gilt nur für das Originalmaterial. Alle gekennzeichneten Fremdinhalte (z.B. Abbildungen, Fotos, Zitate etc.) sind von der CC-Lizenz ausgenommen und für deren Wiederverwendung ist es ggf. erforderlich, weitere Nutzungsgenehmigungen beim jeweiligen Rechteinhaber einzuholen.

Inhalt

<i>Alois Buholzer Dorothee Brovelli</i> Einleitung	7
<i>Laura Graewert Annika Neitemeier Yannick Dohmen Daniel Thurm Stephan Hußmann Bärbel Barzel</i> Potenziale und Hindernisse beim Einsatz eines digitalen Selbst-Assessment-Tools zur Förderung des arithmetischen Basiswissens.....	17
<i>Florian Hofmann</i> Formative Leistungsmessung zu grammatikalischen Grundkompetenzen im Deutschunterricht Wie müssen effektive Rückmeldungen bei datenbasierten Förderkonzepten gestaltet werden?	35
<i>Uwe Maier Christian Klotz</i> Formatives Feedback ist nicht immer hilfreich: Feedbackrezeption und Lernzuwachs in einer Web-App zur Übung von Grammatik und Rechtschreibung.....	63
<i>Mathias Ropohl Hilda Weimar Julia Arnold Svenja Boegel Nele Kampa Maria Opfermann</i> Der Prozess formativen Assessments beim Experimentieren Welche Rolle spielt die kognitive Belastung?.....	83
<i>Merle Ruelmann Urs Grob Alois Buholzer Sandra Zulliger</i> Formatives Assessment in Arbeitsphasen Eine explorative Analyse der Lernunterstützung in Interaktionen zwischen der Lehrperson und Schüler:innen mit verschiedenen kognitiv-motivationalen Profilen	101
<i>Robbert Smit Kurt Hess Patricia Bachmann Heidi Dober</i> Formatives Feedback und Rubrics fördern mathematisches Argumentieren in der Primarschule	129
<i>Sandra Zulliger Merle Ruelmann Hanni Lötscher</i> Unterrichtsgestaltung und Qualität von Peer-Assessment im alltäglichen Unterricht.....	145

<i>Inga Glogger-Frey Anika Bürgermeister Henrik Saalbach</i> Ein digitales Tool zur Unterstützung von formativem Peer-Feedback bei Lehramtsstudierenden.....	171
<i>Claudia Schreiner Fiona MacKay-Falls Christian Kraller</i> Rekonstruktion mentaler Modelle Lehramtsstudierender zur pädagogischen Diagnostik	185
<i>Laura Venitz Ilonca Hardy Mirjam Steffensky Anika Bürgermeister Katharina Junge Henrik Saalbach Miriam Leuchter</i> Einstellungen pädagogischer Fachkräfte in der KiTa zur Bedeutung von formativem Assessment im Kontext der Naturwissenschaften.....	203
<i>Verena Zucker Nicola Meschede Miriam Leuchter</i> Formatives Assessment im naturwissenschaftlichen Grundschulunterricht Was wissen Studierende über Möglichkeiten der Diagnostik und der Rückmeldung und wie hängt dieses Wissen mit ihren Lehr-Lern-Überzeugungen zusammen?.....	221
Autorinnen und Autoren	235

Einleitung

Alois Buholzer | Dorothee Brovelli

Das Thema „Formatives Assessment“ hat in den letzten Jahren zunehmend in den Diskurs zur pädagogischen Diagnostik und darüber hinaus in die Diskussion rund um Unterrichtsqualität Einzug gehalten. Befasste sich die pädagogische Diagnostik bislang schwerpunktmäßig mit Fragen der summativen Leistungsmessung und -bewertung oder mit der Urteilsakkuratheit von Lehrpersonen (Praetorius & Südkamp, 2017), so wird der Fokus mit dem formativen Assessment neuerdings stärker auf die lernprozessbegleitende Diagnostik und die Nutzung ihrer Ergebnisse für die Optimierung von Lehr- und Lernprozessen gelenkt. Diese Neuausrichtung der pädagogischen Diagnostik wird zunehmend durch unterrichtsnahe und fachdidaktisch fundierte Forschungsstudien begleitet und vorangetrieben. Im schweizerischen Kontext beispielsweise ergeben sich verbunden mit den Empfehlungen zur Umsetzung des in der Deutschschweiz maßgeblichen kompetenzorientierten Lehrplans 21, die explizit formative und summative Beurteilungsformen im Unterricht vorsehen, neue Akzente in Bezug auf die Rückmeldung zu Lernprozessen und schulischen Leistungen wie auch in Bezug auf deren Einschätzung (D-EDK, 2014). Auch die deutsche Kultusministerkonferenz (KMK) weist in ihren Empfehlungen zum „Lehren und Lernen in einer digitalisierten Welt“ auf die Wichtigkeit von formativem Feedback zur Unterstützung individueller, personalisierter und kooperativer Lernprozesse hin (KMK, 2021). Der vorliegende Sammelband trägt diesem Perspektivenwechsel Rechnung und vermittelt im ersten Teil anhand von sieben Beiträgen einen Einblick in die gegenwärtige Unterrichtspraxis, der aufzeigt, wie formative Beurteilungsprozesse im Kindergarten, in der Primarschule sowie auf den Sekundarstufen durchgeführt werden. Der zweite Teil des Sammelbands beinhaltet vier Beiträge, die die Anforderungen des formativen Assessments an Lehrpersonen thematisieren, zum Beispiel hinsichtlich ihres Professionswissens und ihrer berufsbezogenen Überzeugungen. Die Beiträge zeigen außerdem auf, wie Lehramtsstudierende für diese Thematik sensibilisiert und qualifiziert werden können.

Der aus dem angelsächsischen Sprachraum stammende Begriff des formativen Assessments wird nachfolgend als „Klammerbegriff“ zur Bezeichnung des Forschungsgegenstands verwendet, mit dem sich die im vorliegenden Band publizierten Beiträge befassen. Der Begriff „Assessment“ lässt sich nicht eins zu eins übersetzen, da im deutschen Sprachraum im Zusammenhang mit Lernprozessen und Leistungsergebnissen oftmals zwischen Erhebung, Beurteilung (Bewertung) und Rückmeldung unterschieden wird, während es sich im Englischen um einen Sammelbegriff handelt (vgl. Bürgermeister & Saalbach, 2018). Infolgedessen wird auch der Begriff „formatives Assessment“ je nach Kontext unterschiedlich übersetzt bzw. verwendet. So spricht man in der Fachliteratur etwa von „formativer Diagnostik“, „prozessbegleitender Beurteilung“ oder „Lernverlaufdiagnostik“, während sich im Bereich der Heil- und Sonderpädagogik der

Begriff „Förderdiagnostik“ etabliert hat. In der Schulpraxis wiederum sind die Begriffe „formative Beurteilung“ oder „fördernde Beurteilung“ gängig. Um diese vielfältigen begrifflichen Ausprägungen von formativen Beurteilungsprozessen gleichermaßen berücksichtigen zu können, wurde der Begriff des formativen Assessments im Titel des vorliegenden Sammelbands als eingedeutschter „Klammerbegriff“ verwendet, der die Beiträge thematisch zusammenhält.

In der Praxis umfasst formatives Assessment unterschiedliche Methoden und Verfahren zur Beurteilung von Lernprozessen und -ergebnissen (davon berichten insbesondere die Studien aus dem ersten Teil des Sammelbands). Trotz der Variabilität des Vorgehens besteht jedoch weitgehende Übereinstimmung in Bezug auf die Funktionen des formativen Assessments, die im Wesentlichen darin bestehen, dass diagnostische Informationen zum Lernen und seinen Ergebnissen mit dem Ziel erfasst werden, die Lehr- und Lernprozesse zu optimieren (Schütze et al., 2018). Während die Schülerinnen und Schüler anhand der gewonnenen Erkenntnisse das eigene Lernen verbessern können, erhält die Lehrperson eine aussagekräftige Ausgangslage für eine adaptive Unterrichtsgestaltung und zur individuellen Lernunterstützung (Helmke, 2008). Formatives Assessment legt den Fokus somit wie einleitend bereits festgehalten auf die prozessbegleitende diagnostische Erfassung von Lernprozessen und die adaptive Nutzung dieser Informationen für den Unterricht und die Steuerung des Lernens. Mit diesem Fokus grenzt es sich von selektiven, auf Notengebung fokussierten Verfahren der Leistungsbewertung ab, die unter dem Begriff des summativen Assessments subsumiert werden können. Bei summativen Assessments werden Lernergebnisse in der Regel am Ende einer Lerneinheit abschließend bewertet und nicht mit Blick auf das weitere Lernen einbezogen.

Formatives Assessment zur Förderung des Lernerfolgs der Schülerinnen und Schüler kann nicht unabhängig von der übergeordneten Ebene der Unterrichtsqualität erfolgen, da es als zentraler Teil von Unterricht seine Wirkung nur im Verbund mit qualitativem Unterricht entfalten kann (Buholzer et al., 2023). Ansätze des formativen Assessment sind deshalb immer auch darauf zu hinterfragen, auf welchem Lehr- und Lernverständnis sie beruhen. Gemäß dem heute bestehenden weitgehenden Konsens, dass Lernen im Aufbauen und Anwenden von Wissens- und Handlungsstrukturen besteht (Reusser, 2006), basieren Ansätze zum formativen Assessment häufig auf einem konstruktivistischen Verständnis von Lernen, Lehren und Diagnostizieren. Für die Durchführung von formativem Assessment bedeutet dies, dass die Schülerinnen und Schüler bewusst in diagnostisch ausgerichtete Überprüfungen einzubeziehen sind, damit sie Einblick in die Prozesse und den Erfolg ihres fachlichen Strukturaufbaus gewinnen können (Smit, 2008), aber auch, um ihnen überfachliche Kompetenzen zu vermitteln, mit deren Hilfe sie ihr Lernen und Problemlösen zunehmend selbst zu steuern vermögen. Formatives Assessment wird somit *nicht für*, sondern *mit* den Lernenden vorgenommen: „Expert FA [Formative Assessment] is a dialogic process that allows students’ ideas to guide instruction and learning“ (Gotwals et al., 2015, S. 407).

Der aktive Einbezug der Schülerinnen und Schüler in die Beurteilung und in die Reflexion von Lernprozessen und -ergebnissen bildet gemeinsam mit der Rolle der Lehrperson auch in der häufig rezipierten Konzeption des formativen Assessments von Black und Wiliam (2018) eine der beiden zentralen Dimensionen. Hinzu kommt die zweite Dimension, die sich an den drei Phasen eines Lernprozesses bzw. an den folgenden drei Leitfragen orientiert: „Wo stehen die Lernenden?“ (aktueller Lernstand), „Wohin soll es gehen?“ (Lernziel) und „Welche Schritte müssen getan werden, um das Ziel zu erreichen?“ (weiteres Vorgehen). Die Konzeption von Black und Wiliam (2018) basiert auf der Kreuzung der auf das Lernen bezogenen Prozessdimensionen mit der Dimension der beteiligten Akteurinnen und Akteure (Lehrpersonen, Mitschülerinnen und Mitschüler sowie Lernende). Die daraus resultierenden Handlungsfelder lassen sich in fünf „Schlüsselstrategien“ verdichten (Abb. 1).

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1 Clarifying learning intentions and criteria for success	2 Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	3 Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	4 Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	5 Activating students as the owners of their own learning	

Abbildung 1: Schlüsselstrategien des formativen Assessments (Black & Wiliam, 2018, S. 560)

Abbildung 1 ist zu entnehmen, dass sich die fünf Schlüsselstrategien des formativen Assessments darin manifestieren, dass Lehrpersonen (1) Lernziele und Beurteilungskriterien mit den Lernenden klären, teilen und verständlich machen, (2) regelmäßig, zum Beispiel durch anregende Fragen und formative Aufgaben, Informationen zum Lernstand einholen (Eliciting Evidence), (3) aufbauend auf diesen diagnostischen Informationen lernförderliche Rückmeldungen geben (Feedback) und die Lernenden durch (4) Peer-Assessments sowie (5) Self-Assessments aktiv in Beurteilungsprozesse einbeziehen.

Hinsichtlich der Umsetzung dieser Schlüsselstrategien können nach Bell und Cowie (2001) verschiedene Strukturierungsgrade unterschieden werden. Formatives Assessment kann entweder geplant sein (formal) oder aus der Situation heraus erfolgen (informal). Das formale formative Assessment dient der Lehrperson als Grundlage für die Unterrichtsgestaltung auf der Makroebene, indem zum Beispiel durch vorab festgelegte, nicht auf Notengebung ausgerichtete Lernkontrollen Informationen zum Lern-

stand der Lernenden eingeholt werden. Das informale formative Assessment bezieht sich demgegenüber auf spontane, sich aus dem Unterrichtsgeschehen heraus ergebende Interaktionen. Bei diesem Ansatz erfasst die Lehrperson situativ Informationen über die Lernprozesse von Schülerinnen und Schülern, die eine Grundlage für individuell-adaptives Feedback darstellen und somit im Gegensatz zum formalen Vorgehen zur Mikroadaptation des Unterrichts beitragen (Bell & Cowie, 2001).

Zur Einleitung in den Sammelband sollen nachfolgend die insgesamt sieben Beiträge des ersten Teils, die sich auf die Anwendung des formativen Assessments im Unterricht beziehen, im Modell von Black und Wiliam (2018) verortet werden. Aus den Ergebnissen soll zudem jeweils eine zentrale Gelingensbedingung bzw. ein Qualitätsmerkmal für lernwirksames formatives Assessment abgeleitet werden. Die Gelingensbedingungen sind im Hinblick auf eine etwaige Generalisierung allerdings sehr vorsichtig zu interpretieren, da sie auf Studien mit sehr unterschiedlichem methodischem Vorgehen und unterschiedlicher Stichprobengröße beruhen.

Der Beitrag *„Potenziale und Hindernisse beim Einsatz eines digitalen Selbst-Assessment-Tools zur Förderung des arithmetischen Basiswissens“* von Laura Graewert, Annika Neitemeier, Yannick Dohmen, Daniel Thurm, Stephan Hußmann und Bärbel Barzel befasst sich mit der Selbst-Diagnose von Schülerinnen und Schülern in Bezug auf mathematische Lernprozesse. Der im Beitrag beschriebene Ansatz des formativen Assessments aktiviert Schülerinnen und Schüler als Verantwortliche des eigenen Lernens und kann daher Schlüsselstrategie 5 „Self-Assessment“ im Modell von Black und Wiliam (2018) zugeordnet werden. Der Beitrag zeigt auf, wie die Durchführung von Self-Assessments mithilfe digitaler Technologien ressourcenschonend unterstützt werden kann. Als bedeutsame Gelingensbedingung für lernwirksame Self-Assessments hat sich in diesem Zusammenhang unter anderem die Bereitstellung einer Musterlösung herausgestellt, mit der die Lernenden ihre eigenen Ergebnisse vergleichen können.

Florian Hofmanns Beitrag *„Formative Leistungsmessung zu grammatikalischen Grundkompetenzen im Deutschunterricht – Wie müssen effektive Rückmeldungen bei datenbasierten Förderkonzepten gestaltet werden?“* fokussiert Schlüsselstrategie 3. Die Ergebnisse können zugleich als Gelingensbedingung gewertet werden: Die in die Untersuchung einbezogenen Schülerinnen und Schüler konnten die größten Kompetenzsteigerungen erzielen, wenn sie von den Lehrpersonen Rückmeldungen mit individuellen Hinweisen zu Fehlern sowie Förderempfehlungen erhielten. Aber auch bereits die Kennzeichnung und die Verbesserung falsch gelöster Aufgaben führte zu lernwirksamen Effekten.

Der Beitrag *„Formatives Feedback ist nicht immer hilfreich: Feedbackrezeption und Lernzuwachs in einer Web-App zur Übung von Grammatik und Rechtschreibung“* von Uwe Maier und Christian Klotz befasst sich ebenfalls mit Schlüsselstrategie 3 nach Black und Wiliam (2018). Die Befunde der Studie legen nahe, dass der Lerninhalt bzw. die Komplexität und der Anspruchsgehalt des Lerninhalts die Art des Feedbacks beeinflussen, was sich wiederum darauf auswirkt, wie formative Feedbacks, die in der Studie mit dem Computer generiert wurden, von den Schülerinnen und Schülern rezipiert

werden. Mit Blick auf die Gelingensbedingung ist daher zu vermuten, dass formative Feedbacks bei komplexen Lerninhalten so formuliert werden sollten, dass sie von den Lernenden auch tatsächlich für die nächsten Lernschritte genutzt werden können.

Der Beitrag *„Der Prozess formativen Assessments beim Experimentieren. Welche Rolle spielt die kognitive Belastung?“* von Mathias Ropohl, Hilda Weimar, Julia Arnold, Svenja Boegel, Nele Kampa und Maria Opfermann setzt sich mit der Frage auseinander, inwiefern die Kompetenz des Experimentierens durch formatives Assessment erhöht werden kann. Im zugrunde liegenden Modell wird insbesondere der Beschäftigung mit dem Lernziel als Grundlage für die Planung des Experiments wie auch den Rückmeldungen zur Umsetzung des Experiments Bedeutung beigemessen, weshalb der Beitrag schwerpunktmäßig den Schlüsselstrategien 1 und 3 zugeordnet werden kann. Die beobachtete höhere Lernwirksamkeit von stark lehrendengelenktem forschendem Lernen verweist auf Gelingensbedingungen auch für die Gestaltung experimenteller Lernumgebungen, in denen Schülerinnen und Schüler Unterstützung in Form von Rückmeldungen erhalten.

Merle Ruelmann, Urs Grob, Alois Buholzer und Sandra Zulliger analysieren in ihrem Beitrag *„Formatives Assessment in Arbeitsphasen: Eine explorative Analyse der Lernunterstützung in Interaktionen zwischen der Lehrperson und Schüler:innen mit verschiedenen kognitiv-motivationalen Profilen“* Interaktionen zwischen Lehrenden und Lernenden in mathematischen Arbeitsphasen in Primarschulklassen. Sie konzentrieren sich dabei vor allem auf die zentralen Unterstützungsstrategien von Lehrpersonen (Schlüsselstrategien 2 und 3) sowie auf das Zusammenspiel der Äußerungen von Lehrpersonen einerseits und Schülerinnen und Schülern andererseits. Aus den diesbezüglichen Ergebnissen kann als Gelingensbedingung der konsequente Einbezug von auf die Schülerinnen und Schüler bezogenen Faktoren (verbale Äußerungen; kognitive und motivationale Voraussetzungen) bei der Lernunterstützung abgeleitet werden.

Der Beitrag *„Formatives Feedback und Rubrics fördern mathematisches Argumentieren in der Primarschule“* von Robbert Smit, Kurt Hess, Patricia Bachmann und Heidi Dober untersucht die Frage, ob mit formativem Feedback, gegebenenfalls unterstützt mit einem Rubric, die Wirksamkeit der Lernbegleitung erhöht werden kann. Mit diesem Erkenntnisinteresse lässt sich der Beitrag im Modell von Black und Wiliam (2018) bei Schlüsselstrategie 3 verorten. Gemäß den präsentierten Studienergebnissen verbesserten sich die Schülerinnen und Schüler auf Klassenebene im Schnitt, wenn sie beim mathematischen Argumentieren durch Rubrics ergänzte Rückmeldungen erhielten. Sowohl das fachspezifische Feedback als auch der damit verbundene Einsatz von Rubrics können daher im Grundsatz als bedeutsame Gelingensbedingungen für lernwirksames formatives Assessment im Kontext des mathematischen Argumentierens aufgefasst werden.

In ihrem Beitrag *„Unterrichtsgestaltung und Qualität von Peer-Assessment im alltäglichen Unterricht“* befassen sich Sandra Zulliger, Merle Ruelmann und Hanni Lötscher am Schluss des ersten Teils mit Schlüsselstrategie 4 nach Black und Wiliam (2018), indem sie die Qualitätsmerkmale der Gestaltung von Peer-Assessments im Mathema-

tikunterricht beschreiben und Zusammenhänge zwischen der Umsetzung einer hohen Peer-Assessment-Qualität und der Unterrichtsgestaltung (Oberflächenstruktur) sowie der Unterrichtsqualität herausarbeiten. Als Gelingensbedingungen für qualitativ hochwertiges Peer-Assessment verweisen sie auf die Bedeutung der Durchführung von eigenständigen Aktivitäten der Schülerinnen und Schüler mit anschließenden lehrpersonenmoderierten Sequenzen sowie auf die Wichtigkeit eines umfassenden Unterrichtsverständnisses, das formative Assessmentstrategien gezielt als Lehr- und Lernstrategie nutzt.

Wie sich den sieben Beiträgen des ersten Teils entnehmen lässt, üben Lehrpersonen bei einer lernwirksamen Durchführung von formativem Assessment eine verantwortungreiche Rolle aus, denn sie bestimmen nicht nur die fachliche Ausrichtung von formativem Assessment und dessen methodische Umsetzung, sondern sie haben auch Einfluss darauf, in welcher Qualität die Reflexions- und Beurteilungsprozesse der Lernenden erfolgen und für die nachfolgenden Lernprozesse genutzt werden (Andrade & Valtcheva, 2009; Harris & Brown, 2013). Es liegen allerdings Befunde vor allem aus Beobachtungsstudien vor, die aufzeigen, dass formatives Assessment im Schulalltag von Lehrpersonen oftmals mit eher geringer Qualität umgesetzt wird (Gotwals et al., 2015; Oswalt, 2013). Da die Durchführung von formativem Assessment aufseiten der Lehrpersonen hohe Anforderungen an deren unterrichtliche Kompetenzen und ihr Professionswissen zu stellen scheint (Stein et al., 2008), könnte die Befundlage darauf hindeuten, dass Lehrpersonen in ihrer Ausbildung häufig noch nicht ausreichend auf ihre anspruchsvolle Aufgabe vorbereitet werden (Buholzer et al., 2020; Bürgermeister et al., 2014; Hebbecker & Souvignier, 2018). Vor diesem Hintergrund erscheint die Frage von hoher Relevanz, wie Lehrpersonen in Aus- und Weiterbildung unterstützt werden können, damit sie in der Lage sind, formatives Assessment in ihrem Unterricht mit hoher Qualität umsetzen. Mit dieser nach wie vor erst ansatzweise erforschten Frage befassen sich die vier Beiträge im zweiten Teil des vorliegenden Bands, die nachfolgend ebenfalls vorgestellt und im Modell von Black und Wiliam (2018) verortet werden.

Inga Glogger-Frey, Anika Bürgermeister und Henrik Saalbach stellen in ihrem Beitrag *„Ein digitales Tool zur Unterstützung von formativem Peer-Feedback bei Lehramtsstudierenden“* ein neu konzipiertes Instrument und seine Implementierung in einem Seminarkonzept vor und setzen damit Schlüsselstrategie 4 wirksam in der Hochschullehre um. In der Studie wurden Lehramtsstudierende dabei unterstützt, beim Schreiben von Lerntagebüchern lernförderliches Feedback zum Einsatz von Lernstrategien zu geben. Dies führte zu positiven Effekten auf ihre Selbstwirksamkeit hinsichtlich des Verfassens von Feedback und auf die Qualität des Feedbacks.

Der Beitrag *„Rekonstruktion mentaler Modelle Lehramtsstudierender zur pädagogischen Diagnostik“* von Claudia Schreiner, Fiona MacKay-Falls und Christian Kraler bedient sich eines empirisch-rekonstruierenden Zugangs. Die anhand von Freewriting-Mikroartikeln erhaltenen Einblicke in die subjektiven Konzepte Studierender verdeutlichen mit Bezug auf das Modell von Black und Wiliam (2018), dass den beteiligten

angehenden Lehrpersonen vor allem die Schlüsselstrategien 2 und 3 geläufig waren und sie in ihren Texten auch den aktiven Einbezug von Lernenden in Beurteilungsprozesse (Schlüsselstrategie 5) nannten, die Schlüsselstrategien 2 und 4 jedoch nicht ansprachen.

Im Beitrag *„Einstellungen pädagogischer Fachkräfte in der KiTa zur Bedeutung von formativem Assessment im Kontext der Naturwissenschaften“* beschäftigen sich Laura Venitz, Ilonca Hardy, Mirjam Steffensky, Anika Bürgermeister, Katharina Junge, Henrik Saalbach und Miriam Leuchter mit Überzeugungen pädagogischer Fachkräfte zur förderbezogenen Diagnostik. Während die befragten pädagogischen Fachkräfte naturwissenschaftliche Bildung zwar als wichtig erachten und diesbezüglich hohe Selbstwirksamkeitsüberzeugungen aufweisen, scheint prozessbegleitende Diagnostik im Bildungsbereich der Naturwissenschaften für sie kaum eine Rolle zu spielen, weshalb sie nur selten genutzt wird.

Zum Abschluss des zweiten Teils zeigen Verena Zucker, Nicola Meschede und Miriam Leuchter im Beitrag *„Formatives Assessment im naturwissenschaftlichen Grundschulunterricht – Was wissen Studierende über Möglichkeiten der Diagnostik und der Rückmeldung und wie hängt dieses Wissen mit ihren Lehr-Lern-Überzeugungen zusammen?“*, dass gemäß ihrer quantitativen Studie bei Studierenden nach dem Besuch einer Vorlesung bereits Wissen zu Möglichkeiten der Diagnostik und der Rückmeldung vorhanden ist, dieses Wissen aber noch nicht ausdifferenziert ist und außerdem nicht immer formativ ausgerichtet zu sein scheint. Diesbezüglich besonders hervorzuheben sind die seltene Nennung des Self- und des Peer-Assessments (Schlüsselstrategien 4 und 5), die fehlende Übertragung auf den naturwissenschaftlichen Grundschulunterricht sowie der Zusammenhang zwischen dem Wissen zur Diagnostik und den konstruktivistischen Lehr-Lern-Überzeugungen.

In ihrer Gesamtheit zeigen die vier Beiträge zur Lehrerinnen- und Lehrerbildung exemplarisch auf, dass verschiedene Zugänge gewählt werden können, um professionelles formativ-diagnostisches Handeln bei (angehenden) Lehrpersonen aufzubauen und zu fördern. Zum einen können Studierende dazu angeregt werden, ihre Kenntnisse und Überzeugungen zu (formativer) pädagogischer Diagnostik zu verbalisieren und zu reflektieren. Zum anderen können in der Hochschullehre Settings geschaffen werden, in denen die Studierenden im Sinn eines „pädagogischen Doppeldeckers“ (Wahl, 2013) formative Beurteilungen selbst erleben, diese reflektieren und die auf diese Weise gewonnenen Erkenntnisse auf die Gestaltung der eigenen schulischen Praxis übertragen.

Die Schlüsselstrategien von Black und Wiliam (2018) erweisen sich allerdings nicht nur im Klassenzimmer und an Institutionen der Lehrerinnen- und Lehrerbildung als zielführend, sondern haben – gewissermaßen auf der Metaebene – auch maßgeblich zur wissenschaftlichen Qualität des vorliegenden Sammelbands beigetragen. Die Autorinnen und Autoren waren in einem „Call for Papers“ dazu aufgefordert worden, entweder einen Beitrag einzureichen, der auf empirischer Grundlage über verschiedene Zugänge von formativen Beurteilungsprozessen im Kindergarten, in der Primarschule oder auf der Sekundarstufe berichtet, oder einen Beitrag zu verfassen, der die Anforderungen

des formativen Assessments an Lehrpersonen thematisiert. Von den vierzehn zugesagten Manuskripten wurden zwölf eingereicht und anschließend einem Double-blind-Review-Verfahren unterzogen. Wie die im Folgenden abgedruckten Beiträge zeigen, konnte insbesondere Schlüsselstrategie 4 auch bei der Vorbereitung des Sammelbands überaus gewinnbringend eingesetzt werden.

Wir bedanken uns an dieser Stelle bei den Autorinnen und Autoren für das Verfassen der Beiträge und der Reviews bzw. „Peer-Assessments“. Besonders danken möchten wir zudem Priska Aregger für die umsichtige Organisation und die sorgfältige administrative Unterstützung bei der Zusammenstellung des Bands.

Literatur

- Andrade, H. L. & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12–19. <https://doi.org/10.1080/00405840802577544>
- Bell, B. & Cowie, B. (2001). *Formative assessment and science education*. Kluwer. <https://doi.org/10.1002/sce.1022>
- Black, P. & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Buholzer, A., Baer, M., Ruelmann, M. & Zulliger, S. (2023). Formatives Assessment – eine vernachlässigte Dimension von Unterrichtsqualität? *Bildung und Erziehung*, 76(1), 43–60. <https://doi.org/10.13109/buer.2023.76.1.43>
- Buholzer, A., Baer, M., Zulliger, S., Torchetti, L., Ruelmann, M., Häfliger, A. & Lötscher, H. (2020). Formatives Assessment im alltäglichen Mathematikunterricht von Primarlehrpersonen: Häufigkeit, Dauer und Qualität. *Unterrichtswissenschaft*, 48(4), 629–661. <https://doi.org/10.1007/s42010-020-00083-7>
- Bürgermeister, A., Klieme, E., Rakoczy, K., Harks, B. & Blum, W. (2014). Formative Leistungsbeurteilung im Unterricht: Konzepte, Praxisberichte und ein neues Diagnoseinstrument für das Fach Mathematik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik: Tests und Trends* (S. 41–60). Hogrefe.
- Bürgermeister, A. & Saalbach, H. (2018). Formatives Assessment: Ein Ansatz zur Förderung individueller Lernprozesse. *Psychologie in Erziehung und Unterricht*, 65(3), 194–205. <http://dx.doi.org/10.2378/peu2018.art11d>
- D-EDK. (2014). *Lehrplan 21*. Deutschschweizer Erziehungsdirektoren-Konferenz.
- Gotwals, A. W., Philhower, J., Cisterna, D. & Bennett, S. (2015). Using video to examine formative assessment practices as measures of expertise for mathematics and science teachers. *International Journal of Science and Mathematics Education*, 13(2), 405–423. <https://doi.org/10.1007/s10763-015-9623-8>
- Harris, L. R. & Brown, G. T. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101–111. <https://doi.org/10.1016/j.tate.2013.07.008>
- Hebbecke, K. & Souvignier, E. (2018). Formatives Assessment im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts. *Zeitschrift für Erziehungswissenschaft*, 21(4), 735–765. <https://doi.org/10.1007/s11618-018-0834-y>

- Helmke, A. (2008). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Klett-Kallmeyer.
- KMK. (2021). *Lehren und Lernen in der digitalen Welt: Ergänzung zur Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt“*. Beschluss der Kultusministerkonferenz vom 09.12.2021. Berlin: Sekretariat der Kultusministerkonferenz. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_12_09-Lehren-und-Lernen-Digi.pdf
- Oswalt, S. G. (2013). *Identifying formative assessment in classroom instruction*. Dissertation. Boise State University.
- 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.
- Reusser, K. (2006). Konstruktivismus – vom epistemologischen Leitbegriff zur Erneuerung der didaktischen Kultur. In M. Baer, M. Fuchs, P. Füglistner, K. Reusser & H. Wyss (Hrsg.), *Didaktik auf psychologischer Grundlage: Von Hans Aebli's kognitionspsychologischer Didaktik zur modernen Lehr- und Lernforschung* (S. 151–168). hep.
- Schütze, B., Souvignier, E. & Hasselhorn, M. (2018). Stichwort – Formatives Assessment. *Zeitschrift für Erziehungswissenschaft*, 21(4), 697–715. <https://doi.org/10.1007/s11618-018-0838-7>
- Smit, R. (2008). Formative Beurteilung im kompetenz- und standardorientierten Unterricht. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 26(3), 383–392. <https://doi.org/10.25656/01:13688>
- Stein, M. K., Engle, R. A., Smith, M. S. & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, 10(4), 313–340. <https://doi.org/10.1080/10986060802229675>
- Wahl, D. (2013). *Lernumgebungen erfolgreich gestalten*. Klinkhardt.

Potenziale und Hindernisse beim Einsatz eines digitalen Selbst-Assessment-Tools zur Förderung des arithmetischen Basiswissens

Laura Graewert¹ | Annika Neitemeier² | Yannick Dohmen³ | Daniel Thurm¹ |
Stephan Hußmann² | Bärbel Barzel³

¹Universität Siegen | ²Technische Universität Dortmund | ³Universität Duisburg-Essen

Abstract: Eine individuelle, diagnosegeleitete Förderung von Lernenden („Formatives Assessment“) ist ein entscheidender Faktor für eine tragfähige Gestaltung von mathematischen Lernprozessen. Digitale Technologien bieten durch dynamische und interaktive Visualisierungen, Adaptivität und automatisches Feedback vielfältige Möglichkeiten, gezielt zu diagnostizieren und zu fördern. Jedoch zeigen sich bei vielen formativen Assessment-Tools Herausforderungen hinsichtlich der diagnostischen Tiefe und der aktiven Einbindung von Lernenden. Im BASE-Projekt wird daher ein verstehensorientiertes, digitales Selbst-Assessment-Tool (BASE-Tool) zum arithmetischen Basiswissen entwickelt und beforscht. Im Rahmen einer Pilotstudie wurden Potenziale und Hindernisse herausgearbeitet, die sich für Lernende beim Arbeiten mit dem BASE-Tool ergeben. Die Bereitstellung einer Musterlösung zur Durchführung des Selbst-Assessments hat sich dabei als lernförderlich herausgestellt. Bedarfe zur Weiterentwicklung gibt es in Bezug auf die intuitive Bedienbarkeit des BASE-Tools sowie die inhaltliche und visuelle Gestaltung des Selbst-Assessments und des automatischen Feedbacks zum Selbst-Assessment.

1 Einführung

Die zunehmende Relevanz digitaler Technologien im Alltag stellt neue Anforderungen an das Bildungssystem und damit auch an den Mathematikunterricht (SWK, 2021). Dabei rücken insbesondere die Möglichkeiten für eine individuelle, diagnosegeleitete Förderung („Formatives Assessment“) in den Fokus, denn Formatives Assessment gilt „als eines der wirksamsten Rahmenkonzepte zur Förderung schulischen Lernens“ (Schütze et al., 2018, S. 697). Durch den Einsatz von digitalen Technologien ergeben sich zum Beispiel durch dynamische und interaktive Darstellungen oder automatische Datenanalysen neue Potenziale, formatives Assessment zu unterstützen. Fachdidaktische Analysen bereits bestehender Angebote zur digitalen Diagnose und Förderung im deutschsprachigen Raum zeigen jedoch, dass diese überwiegend prozedurales und weniger konzeptuelles Wissen fokussieren (Thurm & Graewert, 2022). Das bedeutet, es werden vorrangig Aufgaben implementiert, zu deren Lösung „Wissen bzgl. des Ausführens von mathematischen Verfahren“ benötigt wird statt „Wissen zur mathematischen Begriffen und zur Begründung der Prozeduren“ (Ruwisch, 2020, S. 1273). Zudem lässt

sich feststellen, dass Diagnosen meist auf Basis von Lösungsquoten erfolgen und die Eigenverantwortung der Lernenden – beispielsweise durch Selbst-Assessments – nicht gefördert wird (Thurm & Graewert, 2022). Damit ist ein Nachholbedarf in Bezug auf die forschungsbasierte Entwicklung digitaler, formativer und verstehensorientierter Selbst-Assessmentangebote für das Fach Mathematik zu verzeichnen. Dieser Bedarf wird im Projekt BASE aufgegriffen, in dessen Rahmen ein digitales Selbst-Assessment-Tool (BASE-Tool) zum arithmetischen Basiswissen entwickelt und beforscht wird.

In dem vorliegenden Beitrag sollen das Konzept des BASE-Tools sowie erste Ergebnisse einer explorativen Pilotstudie vorgestellt werden. Dazu werden im nächsten Abschnitt zunächst die theoretischen Grundlagen erläutert. Im 3. Abschnitt wird die Forschungsfrage dargestellt, bevor im 4. Abschnitt die übergeordneten Designprinzipien des BASE-Tools dargelegt werden. Es folgt im 5. Abschnitt die Ergebnisdarstellung der Pilotstudie. Im 6. Abschnitt werden die Ergebnisse zusammengefasst und die nächsten Schritte hinsichtlich der Weiterentwicklung des BASE-Tools aufgezeigt.

2 Theoretische Grundlagen

2.1 Formatives Assessment

Formatives Assessment kann als Prozess verstanden werden, in dessen Rahmen Informationen zu den fachlichen Lernvoraussetzungen von Schüler:innen erhoben werden, um auf dieser Basis das Lehren und Lernen zu verbessern:

„... formative assessment refers to the collaborative processes engaged in by educators and students for the purpose of understanding the students’ learning and conceptual organization, identification of strengths, diagnosis of weaknesses, areas for improvement, and as a source of information that teachers can use in instructional planning and students can use in deepening their understandings and improving their achievement“ (Cizek, 2010, S. 6–7).

Dabei sind insgesamt fünf Schlüsselstrategien formativen Assessments zu unterscheiden (William & Thompson, 2008): (1) Klären und Teilen von Lernzielen und Erfolgskriterien, (2) Erfassen des Verständnisses von Lernenden, (3) Bereitstellen von lernförderlichem Feedback, (4) Aktivierung der Lernenden als instruktionale Ressource füreinander (Peer-Assessment) und (5) Aktivierung der Lernenden als Verantwortliche für ihr eigenes Lernen (Selbst-Assessment).

Nicht alle Strategien sind notwendig, um ein Assessment als formativ zu bezeichnen. Jede der Strategien kann jedoch dazu beitragen, Lernende beim Erreichen von gesetzten Lernzielen zu unterstützen (Cizek, 2010). Häufig wird formatives Assessment jedoch lehrer:innenzentriert umgesetzt und Lernende werden nicht aktiv in den Assessmentprozess einbezogen. So zeigten etwa Kippers et al. (2018) in einer Befragung von Lehr-

kräften, dass Selbst-Assessments nur sporadisch oder überhaupt nicht im Unterricht implementiert werden. In einer von Buholzer et al. (2020) durchgeführten Videostudie zur Häufigkeit, Dauer und Qualität der Umsetzung von formativem Assessment im Mathematikunterricht stellte sich zudem heraus, dass Selbst-Assessments nur einen Anteil von 1 % der mathematischen Unterrichtszeit ausmachen, wenn diese durchgeführt werden. Daher wird gefordert, formatives Assessment stärker lerner:innenzentriert zu gestalten (Cizek, 2010).

2.2 Selbst-Assessment

Im vorliegenden Beitrag wird Selbst-Assessment als ein Prozess aufgefasst, bei dem Lernende über ihre eigenen Fähigkeiten, Lernprozesse oder Lernprodukte reflektieren und/oder diese beurteilen, um ihr individuelles Lernen zu fördern und ihre schulischen Leistungen zu verbessern (Andrade & Boulay, 2003; Brown & Harris, 2013; Panadero et al., 2016; Andrade, 2019). Selbst-Assessment hat somit per Definition einen formativen Charakter: „if there is no opportunity for adjustment and correction, self-assessment is almost pointless“ (Andrade, 2019, S. 2).

Zur Strukturierung von Selbst-Assessments können vorab definierte Standards genutzt werden, beispielsweise in Form von Checklisten. Einerseits können sich die Checklisten auf allgemeine Kompetenzen beziehen, wie etwa auf das Darstellen von Multiplikationen in Bildern. Andererseits können sich die Checklisten auf konkrete Eigenschaften einer Aufgabenlösung beziehen. Für eine Aufgabe, bei der eine Multiplikation durch Sprünge am Zahlenstrahl dargestellt werden soll, kann eine entsprechende Checklisten-Aussage beispielsweise wie folgt lauten: „Das Ergebnis der Rechnung entspricht der Gesamtlänge aller meiner Sprünge“.

Für die fachliche Auseinandersetzung mit den eigenen Fähigkeiten, Lernprozessen und Lernprodukten im Rahmen des Selbst-Assessments sind metakognitive Fähigkeiten notwendig, wie etwa die Fähigkeit zur Selbstbeobachtung oder Selbstkontrolle (Zimmerman, 2002). Die Aktivierung und Nutzung metakognitiver Fähigkeiten und die Übernahme von mehr Eigenverantwortung, zum Beispiel im Rahmen von Selbst-Assessments, kann dazu beitragen, dass sich die schulischen Leistungen von Lernenden verbessern (Brown & Harris, 2013). Selbst-Assessments bergen jedoch die Gefahr, dass Lernende ihre Leistungen nicht korrekt einschätzen und folglich aus ihren Selbsteinschätzungen und der Reflektion der eigenen Arbeit fachlich nicht-tragfähige Schlüsse ziehen (Dunning et al., 2004).

2.3 Formatives Feedback

Eine wichtige Strategie beim formativen (Selbst-)Assessment ist die Rückmeldung von diagnostischen Informationen an die Lernenden in Form von Feedback (William & Thompson, 2008). In Anlehnung an Shute (2008) kann formatives Feedback als Information definiert werden, die Lernenden mit der Absicht zur Verfügung gestellt wird,

Veränderungen in ihrem Denken und Handeln anzuregen, um das Lernen zu verbessern. Dabei kann zwischen verschiedenen Feedbacktypen unterschieden werden. So erhalten Lernende bei dem „*Knowledge of result/response-Feedback*“ (KR) Informationen zur Korrektheit ihrer Antwort. Beim „*Knowledge of the correct response-Feedback*“ (KCR) werden Informationen zur richtigen Lösung bereitgestellt, zum Beispiel in Form einer Musterlösung. Beim „*Multiple-try-Feedback*“ (MTF) wird den Lernenden zurückgemeldet, ob ihre Antwort korrekt oder inkorrekt ist. Im Fall einer inkorrekten Antwort erhalten sie die Möglichkeit zur Überarbeitung ihrer Lösung. Weiterführende Informationen werden in Form von elaborierten Feedbacks zur Verfügung gestellt, beispielsweise als „*Knowledge about concepts-Feedback*“ (KC). Dieses Feedback enthält Erklärungen oder Hinweise, die das konzeptuelle Verständnis der Lernenden unterstützen sollen (Narciss, 2008).

Die Unterstützung der Lernenden kann durch die Bereitstellung von Feedback jedoch nicht gewährleistet werden. Zwar ist die Formulierung und Bereitstellung von Feedback immer mit dem Ziel verbunden, das Lernen zu verbessern, jedoch wird die Bedeutung des Feedbacks nicht nur durch dessen Inhalt und Design bestimmt. Die Rolle der Lernenden ist ebenso entscheidend (Esterhazy & Damşa, 2019), denn inwiefern das Feedback an die Lernenden tatsächlich Veränderungen im Denken und Handeln anregt, hängt im Wesentlichen davon ab, ob und wie das Feedback von den Lernenden wahrgenommen, interpretiert und genutzt wird (Semmer & Jacobshagen, 2010; Rezat, 2021). Daher ist es umso wichtiger, Feedback in verschiedenen Designzyklen zu entwickeln und innerhalb dieser zu analysieren, wie das Feedback den Lernprozess von Schüler:innen beeinflusst (Rezat, 2021). Ausgehend von den Analyseergebnissen kann eine Weiterentwicklung des Feedbacks erfolgen, um eine möglichst gute Passung zwischen der intendierten und der tatsächlichen Bedeutung des Feedbacks für die Lernenden zu erzielen.

2.4 Potenziale digitaler Technologien zur Unterstützung formativen Assessments

Digitale Technologien eröffnen zahlreiche Potenziale, um formatives Assessment im Mathematikunterricht zu unterstützen (Drijvers et al., 2016). Diese Potenziale ergeben sich beispielsweise durch dynamische und interaktive Aufgabenformate, die zur Erfassung individueller Konzepte und Schüler:innenvorstellungen genutzt werden können oder durch die automatisierte Auswertung von Lernendendaten (Yerushalmy & Olsher, 2020). Die automatisierte Auswertung ermöglicht eine unmittelbare Bereitstellung von Feedback (Bokhove & Drijvers, 2012) sowie eine Individualisierung von Lernwegen (Shute et al., 2008).

In Bezug auf technologiegestützte Selbst-Assessments und damit einer stärkeren Einbindung von Lernenden besteht jedoch Bedarf für mehr forschungsbasierte Entwicklung (Ruchniewicz, 2022).

3 Forschungsziel und Forschungsfrage

Das zentrale Ziel des BASE-Projektes ist die Entwicklung eines digitalen Tools (BASE-Tool), welches konzeptuelles Wissen fokussiert und Lernende durch Selbst-Assessments stärker in den formativen Assessmentprozess einbezieht. Die individualisierte Wiederholung des arithmetischen Basiswissens soll den Lernenden den Übergang von der Grundschule in die weiterführenden Schulen erleichtern.

Das Projekt verfolgt dabei einen Design-Based-Research-Ansatz (Gravemeijer & Cobb, 2006; Hußmann et al., 2013), in dessen Rahmen Interventionen mit kleinen Lerngruppen in mehreren Designzyklen durchgeführt werden. Die Interventionen folgen spezifischen Designprinzipien und dienen dazu, unter spezifischen Forschungsfragen Lernprozesse zu verstehen und Lehr-/Lernumgebungen Schritt für Schritt weiterzuentwickeln.

Im Projekt BASE wurden erste Prototypen des BASE-Tools exemplarisch für den Inhaltsbereich der Multiplikation entwickelt. Im Rahmen einer ersten Pilotstudie wurden die Prototypen in zwei Designzyklen mit jeweils acht Lernenden einer Grundschule sowie einer weiterführenden Schule erprobt. Ziel der Pilotstudie war es, Erkenntnisse zu den Bearbeitungsweisen der Lernenden zu erhalten und herauszuarbeiten, welche Potenziale und Hindernisse sich für Lernende ergeben. Somit wurde die folgende Forschungsfrage fokussiert:

Forschungsfrage: *Welche Potenziale und Hindernisse ergeben sich für die Lernenden beim Arbeiten mit dem BASE-Tool und was lässt sich daraus für die Gestaltung des BASE-Tools ableiten?*

4 Designprinzipien des BASE-Tools

Als Grundlage für die Entwicklung des BASE-Tools dienen die im Rahmen des Projekts KOSIMA entwickelten und beforschten „Rechenbausteine“ zur individualisierten und diagnosegeleiteten Aufbereitung des arithmetischen Basiswissens (Hußmann et al., 2011). Die zentrale Idee der Rechenbausteine ist die diagnosegeleitete, eigenaktive Aufbereitung des konzeptuellen Wissens mit Hilfe von adaptiven und gestuften Förderaufgaben. Wesentlich sind die Designprinzipien (D1) Verstehensorientierung, (D2) Selbst-Assessment, (D3) Adaptive Lernwege und (D4) Diagnosegeleitete Förderung. Diese Designprinzipien der Rechenbausteine werden im BASE-Tool durch zwei weitere Designprinzipien ergänzt: (D5) Dynamische und interaktive Musterlösung und (D6) Automatisches Feedback zum Selbst-Assessment (Thurm, 2021; Graewert & Thurm, 2022) (Tab. 1). Die Designprinzipien werden in Abschnitt 5.1 an einer Aufgabe des BASE-Tools konkretisiert.

Tabelle 1: Designprinzipien des BASE-Tools

(D1) Verstehens- orientierung	<ul style="list-style-type: none"> • Nachhaltiges Lernen durch Aufbau von inhaltlichem Verständnis • Differenzierte Verstehensangebote • Gegebenenfalls Rückschritte, wenn Verstehensgrundlagen fehlen • Stütze durch Darstellungen
(D2) Selbst-Assessment	<ul style="list-style-type: none"> • Eigenständige Auswertung der bearbeiteten Diagnoseaufgabe entlang aufgabenbezogener, inhaltlicher Kriterien, z. B. „Das Ergebnis der Rechnung entspricht der Gesamtlänge aller Sprünge“ • Stärkung von metakognitiven und selbstregulativen Fähigkeiten
(D3) Adaptive Lernwege	<ul style="list-style-type: none"> • Automatische Zuweisung von Aufgaben in Abhängigkeit von Aufgabenbearbeitung und Selbst-Assessment • Individualisierung von Lernwegen
(D4) Diagnosegeleitete Förderung	<ul style="list-style-type: none"> • Erhebung der Kenntnisse und Vorstellungen der Lernenden • Gezielte Förderung auf Grundlage der Diagnose • Gestufte Förderaufgaben mit unterschiedlichem Anforderungsniveau
(D5) Dynamische und inter- aktive Musterlösung	<ul style="list-style-type: none"> • Bereitstellung einer dynamischen und interaktiven Musterlösung z. B. mit Abspielfunktion bei zeitlich sukzessiven Darstellungen und optional auswählbaren Visualisierungshilfen (<i>Knowledge of the correct response-Feedback, Knowledge about concepts-Feedback</i>)
(D6) Automatisches Feedback zum Selbst-Assessment	<ul style="list-style-type: none"> • Automatisches Feedback zur Korrektheit des Selbst-Assessments (<i>Knowledge of result/response-Feedback</i>)

Die Struktur des BASE-Tools ist in Abbildung 1 dargestellt. Beim Arbeiten mit dem BASE-Tool führen die Lernenden zunächst eine Prä-Diagnostik durch. Dabei bearbeiten sie drei bis vier interaktive Diagnoseaufgaben (Feld „Diagnoseaufgaben“ in Abb. 1). Nach Bearbeitung jeder Aufgabe führen die Lernenden ein Selbst-Assessment ihrer Aufgabenlösung durch (Feld „Selbst-Assessment“ in Abb. 1). Hierzu bekommen sie eine Checkliste mit aufgabenbezogenen Selbst-Assessment-Aussagen, welche eine korrekte Lösung charakterisieren. Zudem wird den Lernenden eine Musterlösung zur Verfügung gestellt, welche abhängig vom Aufgabentyp dynamische und interaktive Elemente enthalten kann und welche die Lernenden bei der Durchführung des Selbst-Assessments unterstützen soll. Über die interaktiven Elemente können visuelle Hinweise in der Musterlösung ausgewählt werden, welche der Hervorhebung und Erklärung bestimmter Eigenschaften der Musterlösung dienen. Nach Durchführung des Selbst-Assessments erhalten die Lernenden automatisch eine Rückmeldung zu

diesem. Den Lernenden werden Informationen darüber zur Verfügung gestellt, hinsichtlich welcher Selbst-Assessment-Aussagen sie ihre eigene Lösung korrekt bzw. inkorrekt eingeschätzt haben.

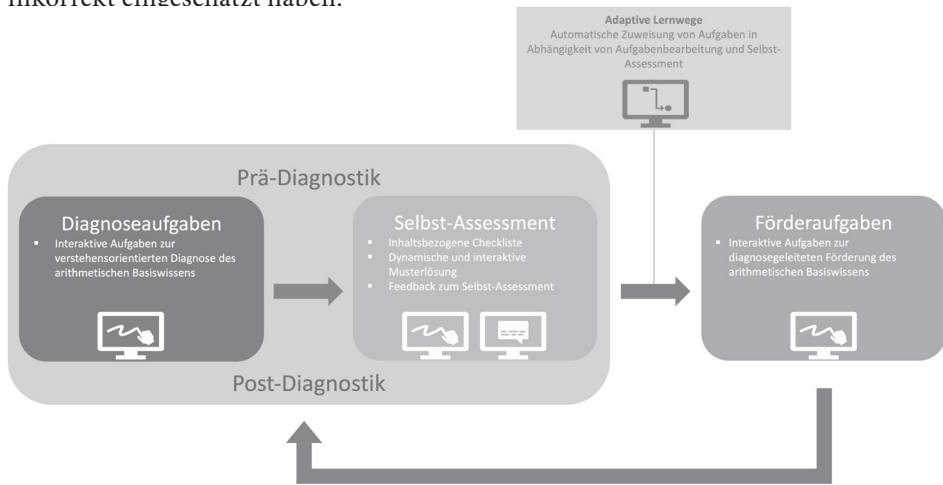


Abbildung 1: Die Struktur des BASE-Tools

Als nächstes werden auf Grundlage der Aufgabenlösung und des Selbst-Assessments bedarfsgerechte Förderaufgaben adaptiv zugewiesen (Feld „Förderaufgaben“ in Abb. 1). Nach der Bearbeitung der Förderaufgaben erfolgt eine Post-Diagnostik, im Rahmen derer die Lernenden die Diagnoseaufgaben (mit veränderten Zahlenwerten) und die zugehörigen Selbst-Assessments nochmals bearbeiten.

Im vorliegenden Beitrag wird der Fokus auf die *Diagnostik* gelegt (Felder „Diagnoseaufgaben“ und „Selbst-Assessment“ in Abb. 1), das heißt, die Bearbeitung der Diagnoseaufgaben und der Selbst-Assessments stehen im Vordergrund.

5 Pilotstudie

Im Folgenden werden die Entwicklung und die Erprobung einer Diagnostik am Beispiel einer Aufgabe dargestellt, die auf das Verständnis der Multiplikation als wiederholte Addition abzielt.

5.1 Erster Designzyklus

Design des Prototyps

Die Aufgabe bestand darin, zu einer vorgegebenen Multiplikationsaufgabe ein passendes Bild in den Zahlenstrahl einzuzeichnen (D1-Verstehensorientierung). Nach Abgabe der Aufgabenbearbeitung erhielten die Lernenden eine Checkliste mit inhaltsbezogenen Selbst-Assessment-Aussagen (rechte Seite in Abb. 2). Mit deren Hilfe sollten die Lernenden

überprüfen, welche der in den Aussagen genannten Eigenschaften in ihrer eigenen Lösung erfüllt bzw. nicht erfüllt sind (D2-Selbst-Assessment¹). War die Beantwortung dieser Selbst-Assessment-Aussagen fehlerhaft, dann wurde dies den Lernenden zurückgemeldet und sie erhielten die Möglichkeit zur einmaligen Überarbeitung (*Multiple-try-Feedback*).

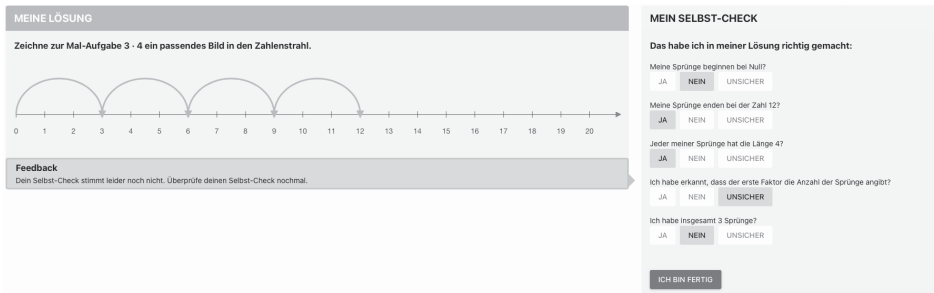


Abbildung 2: Erstes Feedback – Korrektheit der Antworten & Überarbeitung

Traten nach der Überarbeitung weiterhin Fehler in der Beantwortung der Aussagen auf, wurde dies den Lernenden angezeigt und eine dynamische und interaktive Musterlösung bereitgestellt (D5-Dynamische und interaktive Musterlösung, linke Seite in Abb. 3).

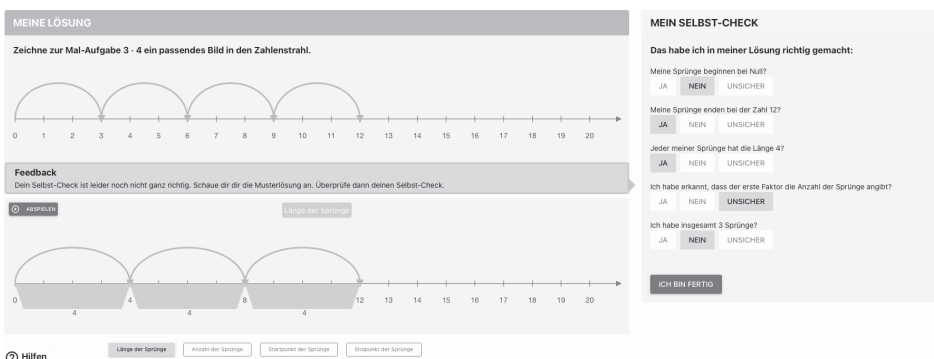


Abbildung 3: Zweites Feedback – Musterlösung & Überarbeitung

Mithilfe der Musterlösung konnten die Antworten auf die Selbst-Assessment-Aussagen erneut überarbeitet werden (*Multiple-try-Feedback*), bevor im letzten Schritt eine abschließende Auswertung der Aussagen erfolgte (D6-Automatisches Feedback zum Selbst-Assessment, rechte Seite in Abb. 4).

Wurde bei einer der in der Liste enthaltenen Aussagen „Richtig“ angezeigt, bedeutete dies, dass sich die lernende Person hinsichtlich dieser Aussage korrekt eingeschätzt hat.

1 Da der Begriff „Selbst-Assessment“ für die Lernenden unverständlich ist, wurde im ersten Designzyklus der Begriff „Selbst-Check“ und im zweiten Designzyklus der Begriff „Selbst-Überprüfung“ genutzt.

Falls eine Fehleinschätzung vorlag, wurde dies über die Formulierung „Richtig wäre ‚Ja‘/‚Nein‘“ zurückgemeldet.

The screenshot displays a digital learning interface. On the left, under the heading 'MEINE LÖSUNG' (My Solution), a task asks to draw three arcs on a number line from 0 to 20. The student's solution shows three arcs starting at 0, 4, and 8, each ending at 4, 8, and 12 respectively. Below the solution is a 'Feedback' section with a 'Schau, was in deinem Selbst-Check richtig war.' (Look, what was right in your self-check) message and an 'ABRUFEN' (View) button. At the bottom, there are buttons for 'Hilfen' (Help) and 'Länge der Sprünge' (Length of jumps).

On the right, the 'MEIN SELBST-CHECK' (My Self-Check) panel contains several questions with 'Richtig' (Correct) and 'Falsch' (Incorrect) buttons. The questions and their status are:

- 'Meine Sprünge beginnen bei Null?' (My jumps start at zero?) - 'Richtig wäre "JA"' (Correct would be "Yes")
- 'Meine Sprünge enden bei der Zahl 12?' (My jumps end at the number 12?) - 'Richtig' (Correct)
- 'Jeder meiner Sprünge hat die Länge 4?' (Every one of my jumps has a length of 4?) - 'Richtig wäre "NEIN"' (Correct would be "No")
- 'Ich habe erkannt, dass der erste Faktor die Anzahl der Sprünge angibt?' (I recognized that the first factor indicates the number of jumps?) - 'Richtig wäre "NEIN"' (Correct would be "No")
- 'Ich habe insgesamt 3 Sprünge?' (I have a total of 3 jumps?) - 'Richtig' (Correct)

 At the bottom of the self-check panel is a button labeled 'WEITER ZUR NÄCHSTEN AUFGABE' (Next to the next task).

Abbildung 4: Drittes Feedback – Korrektheit der Antworten

Erprobung

Die Erprobung im ersten Designzyklus des BASE-Tools fand mit acht Lernenden einer vierten Klasse statt. Eine detaillierte Auswertung findet sich in Thurm und Kühn (2022). Während der Erprobung wurden die Lernenden gebeten, die Methode des „Lauten Denkens“ anzuwenden (Konrad, 2020). Tätigten die Lernenden keine oder unverständliche Aussagen, so wurde die Bitte geäußert, das Vorgehen zu erklären oder es wurden an dieser Stelle Nachfragen gestellt. Im Anschluss an den Bearbeitungsprozess wurde mit den Lernenden ein halbstrukturiertes Interview geführt, bei dem die Lernenden danach gefragt wurden, wie sie die verschiedenen Designelemente des BASE-Tools (z. B. die visuellen Hilfen in der Musterlösung) wahrgenommen und inwiefern diese sie im Lernprozess unterstützt haben. Die Bearbeitungsprozesse sowie die Interviews wurden videografiert.

Ergebnisse

Viele der Lernenden nutzten zwar das erste Feedback zur Korrektheit der ausgewählten Antworten (Abb. 2), um ihre Auswahl noch einmal zu überprüfen. Jedoch führte die Überprüfung selten zu einer Identifikation von Fehlern oder zu einer Überarbeitung.

In Bezug auf die Musterlösung (linke Seite in Abb. 3) wurden von vielen Lernenden die dynamischen und interaktiven Elemente (Abspielfunktion, visuelle Hinweise) verwendet und verstanden. Doch nur wenige Lernende stellten einen Bezug zwischen der eigenen Lösung und der Musterlösung her und keine bzw. keiner der Lernenden setzte die eigene Lösung, die Musterlösung *und* die Selbst-Assessment-Aussagen in Beziehung zueinander.

Beim abschließenden Feedback zur Korrektheit der ausgewählten Antworten auf die Selbst-Assessment-Aussagen (Abb. 4) war es für die Lernenden teilweise unklar, ob sich die Rückmeldung des BASE-Tools auf die eigene Lösung, die Musterlösung oder die Aussagen zum Selbst-Assessment bezog.

5.2 Zweiter Designzyklus

Design des Prototyps

Auf Grundlage der Ergebnisse aus dem ersten Designzyklus wurde die Struktur des BASE-Tools auf Ebene des Designs und des Inhalts im zweiten Designzyklus grundlegend überarbeitet. Dabei zielte die Überarbeitung insbesondere auf die Gestaltung des Selbst-Assessments ab.

Da das erste Feedback zur Korrektheit der Schüler:innenantworten im ersten Designzyklus (Abb. 2) nur selten zu einer Identifikation von Fehlern und einer Überarbeitung der Antworten geführt hatte, wurde dieses Feedback entfernt. Stattdessen wurde den Lernenden unmittelbar nach der Aufgabenbearbeitung die Musterlösung angezeigt. Die Musterlösung konnte so direkt zur Unterstützung des Selbst-Assessments herangezogen werden (Abb. 5).

The screenshot displays a digital learning environment. On the left, under 'Meine Lösung' (My Solution), a number line from 0 to 30 is shown with four large, hand-drawn arcs representing jumps of length 4, starting at 0, 6, 12, and 18. Below this is a 'Mögliche Lösung' (Possible Solution) section with a play button and a similar number line, but with smaller, more frequent jumps. At the bottom of the left panel are three input fields: 'Anzahl der Sprünge' (Number of jumps), 'Länge der einzelnen Sprünge' (Length of individual jumps), and 'Gesamtlänge aller Sprünge' (Total length of all jumps). On the right, the 'Selbst-Überprüfung' (Self-check) section contains several questions with 'JA' (Yes), 'NEIN' (No), and 'UNSICHER' (Uncertain) buttons. The questions are: 1) 'Meine Lösung erfüllt die folgenden Eigenschaften der möglichen Lösung:' (My solution meets the following characteristics of the possible solution:), 2) 'Der erste Faktor in der Rechnung entspricht der Anzahl meiner Sprünge.' (The first factor in the calculation corresponds to the number of my jumps.), 3) 'Der zweite Faktor in der Rechnung entspricht der Länge meiner einzelnen Sprünge.' (The second factor in the calculation corresponds to the length of my individual jumps.), 4) 'Das Ergebnis der Rechnung $6 \cdot 4$ entspricht der Gesamtlänge aller meiner Sprünge.' (The result of the calculation $6 \cdot 4$ corresponds to the total length of all my jumps.), 5) 'Ich habe noch Fragen zu der Aufgabe oder der Lösung.' (I still have questions about the task or the solution.). At the bottom of the right panel is a button labeled 'ICH BIN FERTIG' (I am finished).

Abbildung 5: Erstes Feedback – Musterlösung

Da die visuellen Hinweise in der Musterlösung (Abb. 3) für die meisten Lernenden bereits im ersten Designzyklus nachvollziehbar und verständlich waren, wurden diesbezüglich keine designbezogenen oder inhaltlichen Änderungen vorgenommen.

Für einen Großteil der Lernenden war es jedoch schwierig, einen Zusammenhang zwischen der eigenen Lösung, der Musterlösung und den Selbst-Assessment-Aussagen herzustellen. Um die Lernenden zur Nutzung aller drei Elemente (eigene Lösung, Musterlösung, Selbst-Assessment-Aussagen) anzuregen, wurde in den Selbst-Assessment-Aussagen der Bezug zwischen der symbolischen und ikonischen Darstellungsebene stärker fokussiert. Beispielsweise wurde die Aussage „Jeder meiner Sprünge hat die Länge 4“ durch die Aussage „Der zweite Faktor in der Rechnung entspricht der Länge meiner einzelnen Sprünge“ ersetzt. Die Veränderung sollte dazu beitragen, dass die Lernenden ihre eigene Lösung mithilfe der Musterlösung und den Aussagen zum Selbst-Assessment reflektieren und die verschiedenen Informationen miteinander in

Verbindung bringen. Um eine kognitive Überforderung aufgrund der inhaltlich anspruchsvollen Aussagen zu vermeiden, wurden die Selbst-Assessment-Aussagen in ihrer Anzahl auf weniger Aussagen beschränkt (Abb. 5, rechte Seite).

Aufgrund der Tatsache, dass es im ersten Designzyklus für einige Lernende nicht verständlich war, worauf sich das abschließende Feedback zur Korrektheit der ausgewählten Antworten bezog (Abb. 4), wurde auch dieses im zweiten Designzyklus überarbeitet. Statt Informationen darüber zur Verfügung zu stellen, welche Antwortoption richtig war oder richtig gewesen wäre (z. B. „Richtig wäre ‚Ja‘“), wurde im zweiten Designzyklus für die Lernenden explizit gemacht, ob sie sich hinsichtlich einer bestimmten Eigenschaft richtig eingeschätzt haben und ob diese Eigenschaft in ihrer Lösung erfüllt ist oder nicht, zum Beispiel „Du hast erkannt, dass deine Lösung diese Eigenschaft nicht erfüllt“ (Abb. 6 und 7, rechte Seite).

In Bezug auf die Darstellung und die Übermittlung des Feedbacks wurden hierbei zwei verschiedene Design-Varianten erprobt: Einige der Lernenden erhielten die erste Variante (Abb. 6), bei welcher nach Durchführung des Selbst-Assessments die ausgewählten Antworten farblich entweder in grün oder rot markiert und schriftliche Rückmeldungen unmittelbar zur Verfügung gestellt wurden. Eine grüne Hervorhebung der Antwort entsprach dabei einer korrekten, eine rote Hervorhebung einer fehlerhaften Selbsteinschätzung.

Abbildung 6: Zweites Feedback – Korrektheit der Antworten (1. Variante)

Die schriftlichen Rückmeldungen wurden auch in der zweiten Variante genutzt (Abb. 7), diese konnten sich die Lernenden hier jedoch optional über einen Infobutton anzeigen lassen. Statt die Antworten der Lernenden farblich zu markieren, wurde diese einer automatischen Überprüfung gegenübergestellt. Eine korrekte Selbsteinschätzung war in dieser Version daran zu erkennen, dass die Antwort der lernenden Person mit der automatischen Überprüfung übereinstimmte.

The screenshot displays a digital learning interface with two main sections: 'Meine Lösung' (My Solution) and 'Mögliche Lösung' (Possible Solution). The 'Meine Lösung' section shows a number line from 0 to 30 with four large jumps of length 6, starting at 0 and ending at 24. Below it, the 'Mögliche Lösung' section features a button labeled 'Abspielen' (Play) and a number line with eight smaller jumps of length 4, also starting at 0 and ending at 24. At the bottom, there are three input fields: 'Anzahl der Sprünge' (Number of jumps), 'Länge der einzelnen Sprünge' (Length of individual jumps), and 'Gesamtlänge aller Sprünge' (Total length of all jumps).

On the right side, there is a 'Selbst-Überprüfung' (Self-Check) panel. It contains the text: 'Meine Lösung erfüllt die folgenden Eigenschaften der möglichen Lösung:'. Below this, there are two columns of questions with 'JA' (Yes) and 'NEIN' (No) buttons, and an information icon (i). The questions are:

- Der erste Faktor in der Rechnung entspricht der Anzahl meiner Sprünge. (Buttons: JA, NEIN)
- Der zweite Faktor in der Rechnung entspricht der Länge meiner einzelnen Sprünge. (Buttons: NEIN, NEIN)
- Das Ergebnis der Rechnung 6 · 4 entspricht der Gesamtlänge aller meiner Sprünge. (Buttons: JA, JA)
- Ich habe noch Fragen zu der Aufgabe oder der Lösung. (Button: UNSICHER)

 At the bottom of the panel, there is a button 'Aufgabe abschließen' (Finish task) with a right-pointing arrow.

Abbildung 7: Zweites Feedback – Korrektheit der Antworten (2. Variante)

Erprobung

Die zweite Erprobung wurde mit acht Lernenden einer fünften Gymnasialklasse durchgeführt. Die Methodik des Interviews und die Aufgaben wurden aus dem ersten Designzyklus übernommen. Allerdings sollten diesmal nicht die Aufgaben $3 \cdot 4$ und $3 \cdot 3$, sondern die Aufgaben $6 \cdot 4$ und $5 \cdot 6$ am Zahlenstrahl dargestellt werden.

Ergebnisse

Die Beobachtung der Lernenden bei der Nutzung des BASE-Tools und die Analyse ihrer Vorgehensweisen im zweiten Designzyklus zeigte, dass ein Großteil der Lernenden die Musterlösung (Abb. 5) zur Durchführung des Selbst-Assessments einsetzte. Auch wurde bei der Begründung der Antworten auf die Selbst-Assessment-Aussagen explizit auf die Musterlösung sowie auf die eigene Lösung Bezug genommen. Die dynamischen und interaktiven Elemente der Musterlösung blieben jedoch häufig ungenutzt: Die Lernenden ließen sich selten die Musterlösung abspielen und nur vereinzelt wurden die visuellen Hinweise zur Anzahl, Länge und der Gesamtlänge der Sprünge ausgewählt. Auf Grundlage der beobachteten Bearbeitungsprozesse und Interviews lässt sich dieses Verhalten auf zwei Gründe zurückführen: (1) Die Musterlösung war für die Lernenden auch ohne die visuellen Hinweise beispielsweise zur Sprunganzahl oder -länge nachzuvollziehen. (2) Die Möglichkeit des dynamischen Abspielens und des Auswählens visueller Hinweise war für die Lernenden nicht ersichtlicher.

Hinsichtlich der beiden Varianten für das zweite Feedback (Abb. 6 und 7) zeigten sich keine Differenzen: In beiden Varianten traten insofern Schwierigkeiten auf, dass die ergänzenden Erläuterungen nicht näher betrachtet wurden und falls doch, die Unterschiede zwischen den Formulierungen „erkannt/nicht erkannt“ bzw. „erfüllt/nicht erfüllt“ von den Lernenden häufig unbemerkt blieben.

6 Zusammenfassung und Ausblick

Ziel des Projektes ist die Entwicklung eines digitalen, verstehensorientierten Selbst-Assessment-Tools zur Diagnose und Förderung des arithmetischen Basiswissens von Lernenden. Für diesen Zweck wurde im Sinne eines Design-Based-Research-Ansatzes eine Pilotstudie mit insgesamt 16 Lernenden durchgeführt und erste Prototypen des BASE-Tools in zwei Designzyklen erprobt.

Eine Limitation der vorliegenden Pilotstudie besteht darin, dass im Rahmen der ersten Erprobung Lernende einer vierten Grundschulklasse und bei der zweiten Erprobung Lernende einer fünften Gymnasialklasse mit dem BASE-Tool arbeiteten. Das bedeutet, dass sich die Ergebnisse beim zweiten Designzyklus nicht eindeutig auf die Überarbeitung des BASE-Tools zurückführen lassen. Die Ergebnisse könnten auch in der Tatsache begründet sein, dass die Gymnasialschüler:innen älter und in ihrem Lernstand vermutlich weiter fortgeschritten waren als die Grundschüler:innen. Trotz dieser Limitation konnten im Rahmen der Erprobungen Ansatzpunkte zur Weiterentwicklung des digitalen Selbst-Assessment-Tools herausgearbeitet werden. Daraus lassen sich Schlussfolgerungen ableiten, die bei der zukünftigen Weiterentwicklung des Tools berücksichtigt werden können.

So zeigte sich im ersten Designzyklus, dass das erste Feedback zur Korrektheit der Antworten auf die Selbst-Assessment-Aussagen die Lernenden dazu anregte, ihre eigene Lösung zu überprüfen. Selten führte die Bereitstellung des ersten Feedbacks jedoch zu einer Identifikation von Fehlern und einer Änderung der Eingaben. Für die Weiterentwicklung des Prototyps im zweiten Designzyklus wurde es daher als sinnvoll erachtet, den Lernenden die Musterlösung unmittelbar nach der Bearbeitung der Diagnoseaufgabe bereitzustellen, sodass die Einschätzung der eigenen Lösung direkt in Bezug auf eine konkrete Aufgabenlösung erfolgen konnte. Es zeigte sich, dass die Lernenden durch diese Veränderungen im Tool-Design dabei unterstützt werden konnten, einen Bezug zwischen ihrer eigenen Lösung, der Musterlösung und den Selbst-Assessment-Aussagen herzustellen und sich intensiv mit ihren Aufgabenbearbeitungen auseinanderzusetzen. So nutzten beispielsweise einige der Lernenden die Musterlösung, um die Selbst-Assessment-Aussagen zu beantworten und zu begründen.

Sofern die dynamischen und interaktiven Elemente (Abspielfunktion, visuelle Hilfen z. B. zur Anzahl der Sprünge) in der Musterlösung genutzt wurden, stellten diese sich als verständlich für die Lernenden heraus, sodass eine Überarbeitung ihres Inhalts und ihres Designs in beiden Zyklen nicht notwendig war. Um die Nutzung der dynamischen und interaktiven Elemente für die Lernenden etwas intuitiver zu gestalten, sollen die Anordnung und das Design der Buttons verändert und Hinweise zur deren Nutzung in das Tool implementiert werden.

Während die abschließende Rückmeldung zur Korrektheit der Antworten auf die Selbst-Assessment-Aussagen im ersten Designzyklus für die meisten Lernenden unverständlich war, konnten die ergänzenden Informationen im zweiten Designzyklus nicht dazu beitragen, diesen Verständnisschwierigkeiten entgegenzuwirken. Hingegen stellte sich heraus, dass die Art der Darstellung – abgesehen von den ergänzenden Informationen – sowohl bei der ersten Variante als auch bei der zweiten Variante für die Lernenden nachvollziehbar ist. Um weniger eine korrekte oder inkorrekte Selbsteinschätzung zu betonen (Abb. 6: 1. Variante) und auf auftretende Konflikte zwischen der Selbst-Überprüfung und der automatischen Überprüfung „urteilsfrei“ hinzuweisen, soll das Feedback daher zukünftig wie in der zweiten Variante (Abb. 7) gestaltet werden, wobei die ergänzenden Informationen entfernt werden sollen. Auf Konflikte zwischen der Selbst-Überprüfung und der automatischen Überprüfung soll durch farbliche und symbolische Hervorhebungen stärker aufmerksam gemacht werden (Graewert et al., 2023a).

Neben dem automatischen Feedback zum Selbst-Assessment sollen die Lernenden zudem eine explizite Rückmeldung über die Korrektheit ihrer Aufgabenlösung erhalten. Das bedeutet, das Designprinzip (D6) soll um ein automatisches Feedback zur *Korrektheit der Aufgabenlösung* ergänzt werden (Graewert et al., 2023b).

Zusammenfassend lässt sich festhalten, dass die Einbindung von Lernenden in den Assessmentprozess komplex ist und nicht immer davon ausgegangen werden kann, dass die beabsichtigten Effekte erzielt werden. Dies betont die Notwendigkeit, Feedback nicht nur als Information zu betrachten, welche Lernenden zur Unterstützung des Lernprozesses bereitgestellt wird. Vielmehr sollte es von Interesse sein, Feedback als einen Prozess zu verstehen und im Rahmen dessen zu untersuchen, ob und wie Lernende die ihnen zur Verfügung gestellten Informationen wahrnehmen, interpretieren und nutzen (vgl. Rezat, 2021; Leighton, 2019). Mit Blick darauf erscheint es sinnvoll, die konkrete Ausgestaltung von Feedback im Rahmen von Designzyklen zu entwickeln und zu erforschen (Rezat, 2021).

Neben der Weiterentwicklung des Prototypens soll es daher das Ziel anknüpfender Studien sein, über die Hindernisse und Potenziale hinaus zu untersuchen, inwiefern Feedback im Rahmen von Selbst-Assessment-Prozessen das multiplikative Verständnis und die Metakognition von Lernenden beeinflusst (Graewert et al., 2023c).

Literatur

- Andrade H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, 1–13. <https://doi.org/10.3389/educ.2019.00087>
- Andrade, H. G. & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research*, 97(1), 21–34. <https://doi.org/10.1080/00220670309596625>

- Bokhove, C. & Drijvers, P. (2012). Effects of feedback in an online algebra intervention. *Technology, Knowledge, Learning*, 17, 43–59. <https://doi.org/10.1007/s10758-012-9191-8>
- Brown, G. T. & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Hrsg.), *SAGE Handbook of Research on Classroom Assessment* (S. 367–393). SAGE Publications. <https://doi.org/10.4135/9781452218649.n21>
- Buholzer, A., Baer, M., Zulliger, S., Torchetti, L., Ruelmann, M., Häfliger, A. & Lötscher, H. (2020). Formatives Assessment im alltäglichen Mathematikunterricht von Primarlehrpersonen: Häufigkeit, Dauer und Qualität. *Unterrichtswissenschaft*, 48, 629–661. <https://doi.org/10.1007/s42010-020-00083-7>
- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of Formative Assessment*, (S. 3–17). Routledge. <https://doi.org/10.4324/9781315166933-1>
- Drijvers, P., Ball, L., Barzel, B., Heid, M. K., Cao, Y. & Maschietto, M. (2016). *Uses of technology in lower secondary mathematics education. A concise topical survey*. Springer. https://doi.org/10.1007/978-3-319-33666-4_1
- Dunning, D., Heath C. & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Esterhazy, R. & Damşa, C. (2019). Unpacking the feedback process: An analysis of undergraduate students' interactional meaning-making of feedback comments. *Studies in Higher Education*, 44(2), 260–274. <https://doi.org/10.1080/03075079.2017.1359249>
- Graewert, L., Thurm, D., Neitemeier, A., Hußmann, S., Barzel, B. & Dohmen, Y. (2023a). BASE – Formatives Selbst-Assessment mit digitalen Medien. In IDMI-Primar Goethe-Universität Frankfurt (Hrsg.), *Beiträge zum Mathematikunterricht 2022. 56. Jahrestagung der Gesellschaft für Didaktik der Mathematik* (S. 769–772). WTM. <http://doi.org/10.17877/DE290R-23624>
- Graewert, L., Thurm, D., Hußmann, S. & Barzel, B. (2023b). Digital formative self-assessment [Manuskript eingereicht zur Publikation]. *Proceedings of the 16th International Conference on Technology in Mathematics Teaching (ICTMT 16)*.
- Graewert, L., Thurm, D. & Barzel, B. (2023c). Automatic feedback in self-assessed multiplicative learning [Manuskript eingereicht zur Publikation]. *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME 13)*.
- Graewert, L. & Thurm, D. (2022). Using digital technology to foster students' mathematical and self-assessment competencies. In C. Fernández, S. Llinares, A. Gutiérrez & N. Planas (Hrsg.), *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (S. 219). PME.
- Gravemeijer, K. & Cobb, P. (2006). Design research from a learning design perspective. In J. Van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Hrsg.), *Educational Design Research* (S. 45–85). Routledge.
- Hußmann, S., Thiele, J., Hinz, R., Prediger, S. & Ralle, B. (2013). Gegenstandsorientierte Unterrichtsdesigns entwickeln und erforschen - Fachdidaktische Entwicklungsforschung im Dortmunder Modell. In M. Komorek & S. Prediger (Hrsg.), *Der lange Weg zum Unterrichtsdesign: Zur Begründung und Umsetzung genuin fachdidaktischer Forschungs- und Entwicklungsprogramme* (S. 25–42). Waxmann.
- Hußmann, S., Prediger, S., Barzel, B. & Leuders, T. (Hrsg.). (2011). *Mathewerkstatt 5, Rechenbausteine, Selbsttest*. Cornelsen.
- Kippers, W. B., Wolterinck, C. H., Schildkamp, K., Poortman, C. L. & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision ma-

- king in classroom practice. *Teaching and Teacher Education*, 75, 199–213. <https://doi.org/10.1016/j.tate.2018.06.015>
- Konrad, K. (2020). Lautes Denken. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 1–21). Springer. https://doi.org/10.1007/978-3-658-18387-5_41-2
- Leighton, J. P. (2019). Students' interpretation of formative assessment feedback: Three claims for why we know so little about something so important. *Journal of Educational Measurement*, 56(4), 793–814. <https://doi.org/10.1111/jedm.12237>
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer & M. P. Driscoll (Hrsg.), *Handbook of Research on Educational Communications and Technology* (S. 125–144). Routledge.
- Panadero, E., Brown, G. L. & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28, 803–830. <https://doi.org/10.1007/s10648-015-9350-2>
- Rezat, S. (2021). How automated feedback from a digital mathematics textbook affects primary students' conceptual development: Two case studies. *ZDM – Mathematics Education*, 53(1), 1433–1455. <https://doi.org/10.1007/s11858-021-01263-0>
- Ruchniewicz, H. (2022). *Sich selbst diagnostizieren und fördern mit digitalen Medien. Forschungsbasierte Entwicklung eines Tools zum formativen Selbst-Assessment funktionalen Denkens*. Springer. <https://doi.org/10.1007/978-3-658-35611-8>
- Ruwisch, S. (2020). Konzeptuelles und prozedurales Wissen beim Längenverständnis – eine Schulbuchanalyse. In Siller, H.-S., Weigel, W. & Wörler, J. F. (Hrsg.), *Beiträge zum Mathematikunterricht 2020*. WTM-Verlag. <https://doi.org/10.37626/GA9783959871402.0>
- Schütze, B., Souvignier, E. & Hasselhorn, M. (2018). Stichwort – Formatives Assessment. *Zeitschrift für Erziehungswissenschaft*, 21, 697–715. <https://doi.org/10.1007/s11618-018-0838-7>
- Semmer, N. & Jacobshagen, N. (2010). Feedback im Arbeitsleben – eine Selbstwert-Perspektive. *Gruppendynamik und Organisationsberatung*, 41, 39–55. <https://doi.org/10.1007/s11612-010-0104-9>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Shute, V. J., Hansen, E. & Almond, R. (2008). You can't fatten a hog by weighing it – Or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education*, 18(4), 289–316.
- SWK (Ständige wissenschaftliche Kommission der Kultusministerkonferenz) (2021). *Stellungnahme zur Weiterentwicklung der KMK-Strategie „Bildung in der digitalen Welt“*.
- Thurm D. & Graewert, L. A. (2022). *Digitale Mathematik-Lernplattformen in Deutschland*. Springer. <https://doi.org/10.1007/978-3-658-37520-1>
- Thurm D. & Kühn, P. (2022). Mehr als richtig oder falsch? – Digitales formatives Selbst-Assessment verstehensorientiert gestalten. In B. Brandt, L. Bröll & H. Dausend (Hrsg.), *Digitales Lernen in der Grundschule III – Fachdidaktiken in der Diskussion* (S. 367–380). Waxmann Verlag.
- Thurm, D. (2021). Digital technology supporting formative self-assessment. In M. Inprasitha, N. Changsri & N. Boonsena (Hrsg.), *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education* (S. 185). PME.
- William, D. & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Hrsg.), *The future of assessment: Shaping teaching and learning* (S. 53–82). Lawrence Erlbaum. <https://doi.org/10.4324/9781315086545-3>

- Yerushalmy, M. & Olsher, S. (2020). Online assessment of students' reasoning when solving example-eliciting tasks: using conjunction and disjunction to increase the power of examples. *ZDM – Mathematics Education*, 52, 1033–1049. <https://doi.org/10.1007/s11858-020-01134-0>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Formative Leistungsmessung zu grammatikalischen Grundkompetenzen im Deutschunterricht

Wie müssen effektive Rückmeldungen bei datenbasierten Förderkonzepten gestaltet werden?

Florian Hofmann

Friedrich-Alexander-Universität Erlangen-Nürnberg

Abstract: Im bildungs- und schulbezogenen Diskurs nehmen Schlagwörter wie „Umgang mit Heterogenität“ oder „Differenzierung“ seit mehreren Jahren eine Schlüsselstellung ein. Um jedoch individuelle Lernstände zur Optimierung von Lehr-Lernprozessen zu erfassen, ist ein verändertes Verständnis von Pädagogischer Diagnostik notwendig. Unter dem Oberbegriff formatives Assessment sammeln sich zahlreiche Konzepte, die alle die Intention eint, Lernen anhand nicht-selektiver Teststrukturen zu verbessern. Die vorliegende Studie basiert auf dieser Grundüberlegung und versucht Hinweise zu formulieren, wie effektive Rückmeldungen zu formativen Leistungsmessungen konzipiert werden sollten. Hierzu wurden Schülerinnen und Schüler der Sekundarstufe I nach Kompetenztests zu grammatikalischen Grundfertigkeiten im Deutschunterricht Untersuchungsgruppen mit unterschiedlichen Rückmeldungen zugeteilt. Es zeigte sich, dass Lernende, die eine Korrektur durch eine Lehrkraft sowie weitere individuelle Hinweise bzw. Förderempfehlungen erhielten, die größten Kompetenzsteigerungen erzielten. Im Beitrag werden Befunde, Interpretationsansätze sowie mögliche Konsequenzen für die Unterrichtspraxis dargestellt und diskutiert.

1 Einführung

Die gravierend veränderten Lebens- und Arbeitsbedingungen in unserer Gesellschaft haben unter anderem zur Folge, dass die Diversität bei Schülerinnen und Schülern weiter zunimmt und die Bildungssysteme auf diese Herausforderung adäquat reagieren müssen (Schroeder, 2016). Schlagwörter wie „Umgang mit Heterogenität“ oder „Differenzierung“ nehmen daher im aktuellen bildungs- und schulbezogenen Diskurs eine Schlüsselstellung ein (z. B. Ludin, 2021). Die Pädagogische Diagnostik nimmt bei der Reaktion auf die skizzierte Situation eine zentrale Rolle ein, da eine Adaption der unterrichtlichen Aktivitäten an profunde Kenntnisse von individuellen Lernvoraussetzungen und Lernständen der Schülerinnen und Schüler gebunden ist (Moser Opitz, 2022). Um diese zu erfassen und zur Optimierung von Lehr-Lernprozessen zu nutzen, sind innovative Methoden und Messinstrumente notwendig (Winter, 2015; Tröster, 2018). Unter dem Oberbegriff formatives Assessment haben sich in den letzten Jahrzehnten

zahlreiche Konzepte und Ansätze gesammelt, die alle die Intention eint, Lernen anhand einer nicht-selektiv gerichteten Teststruktur zu verbessern (McDonald & Boud, 2003). Die Potentiale derartiger testbasierter Förderkonzepte gelten mittlerweile als empirisch belegt (z. B. Kingston & Nash, 2011; Hattie, 2012); positive Resultate mehrerer aktueller domänenspezifischer Untersuchungen (z. B. Rakoczy et al., 2017) stützen diese Einschätzung. Allerdings ist die Befundlage keineswegs eindeutig und die mittleren Effektstärken für formatives Assessment variieren stark (Dunn & Mulvenon, 2009; Kingston & Nash, 2011). Diese häufig deutlich voneinander abweichenden Effekte formativen Assessments lassen sich unter anderem mit unterschiedlichen Konzepten, Domänen, Kompetenzbereichen, Zielgruppen oder Korrektur- bzw. Rückmeldeoptionen erklären (Bennett, 2011; Klauer, 2014; Schmidt, 2020). Eine Konsequenz dieser inkonsistenten Befundlage muss es daher sein, konkrete Konzepte formativen Assessments zu erproben und für die Implementation in der Schulpraxis vorzuschlagen (z. B. Maier et al., 2012; Schmidt, 2020). An dieser Stelle knüpft das Projekt TÜV (*Testen-Üben-Verstehen*) an: Mit Hilfe von regelmäßig angesetzten formativen Assessments und unterschiedlichen Rückmeldeformaten soll ein Förderkonzept umgesetzt werden, welches Kompetenzsteigerungen ermöglicht. Neben der Konzeption einer im Schulalltag tatsächlich nutzbaren Unterrichtssequenz mit regelmäßig integrierten formativen Assessments sollen durch die Auswertung und eine nachfolgende Evaluation situationsspezifische Merkmale für derartige Förderkonzepte identifiziert und Hinweise zur Entwicklung und Durchführung von formativen Assessments im Unterrichtsalltag formuliert werden. Im folgenden Beitrag werden der erste Teil des Pilotprojekts TÜV-D (*Testen-Üben-Verstehen* im Deutschunterricht), welcher sich auf die Erprobung unterschiedlicher Rückmeldeversionen formativer Assessments im Rahmen eines datenbasierten Förderkonzepts im Deutschunterricht konzentriert, sowie erste Befunde kurz vorgestellt.

2 Theoretische Grundlagen und Stand der Forschung

2.1 Formatives Assessment

Der Terminus des formativen Assessments wurde im Zuge der Forschungshistorie aus unterschiedlichen Perspektiven und mit diversen Schwerpunkten verortet: So existieren beispielsweise Definitionen anhand diagnostischer Strategien bzw. Dimensionen (z. B. Pawlik, 1976) oder aus dem Blickwinkel der Evaluationsforschung (z. B. Scriven, 1967). Diverse Abhandlungen, auch im deutschsprachigen Bereich (z. B. Klauer, 2014; Maier, 2010; Schütze et al., 2018; Bürgermeister & Saalbach, 2018), haben sich den verschiedenen Definitionen, Hintergründen und Weiterentwicklungen gewidmet (z. B. Dunn & Mulvenon, 2009). Sehr häufig wird nach wie vor als Ausgangspunkt für weitere Gedanken auf die Verortung von Black und Wiliam (2009, S. 9) zurückgegriffen, die in erster Linie hinsichtlich der *Intention*, also der diagnostischen Ausrichtung, unterscheidet:

„Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited.“

Primärintention eines formativen Assessments ist gemäß jener Definition das (zusätzliche) Einholen von Informationen für alle am Lehr-Lernprozess beteiligten Akteure. Diese so gewonnenen Informationen sollen Lernprozesse sowie Lernen optimieren aber keine Grundlage für Selektionsentscheidungen sein (McDonald & Boud, 2003). Daraus kann abgeleitet werden, dass formative Assessments nur dann sinnvoll sind, wenn sie während eines Lehr-Lernprozesses stattfinden (Pawlik, 1976).

Souvignier und Haselhorn (2018, S. 694) bilanzieren aus diesen definierenden Attributen, dass „der Grundgedanke formativen Assessments bestechend einfach ist“ und allen Ansätzen einen Namen gibt, bei denen „systematisch erhobene Leistungsinformationen genutzt werden, um pädagogische Entscheidungen mit dem Ziel einer Optimierung des Lehr-Lernprozesses zu treffen.“ Diese lediglich einen Rahmen bildende Verortung hat zur Folge, dass der Terminus wenig trennscharf genutzt wird (Dunn & Mulvenon, 2009) und eine Vielzahl an völlig unterschiedlichen, oftmals kaum miteinander zu vergleichender Maßnahmen das Label formatives Assessment trägt (Bennett, 2011; Bürgermeister & Saalbach, 2018). Klieme und Kollegen (2010) stellen in diesem Kontext fest, dass sich die Bandbreite von Ansätzen und Konzepten, die unter dem Oberbegriff des formativen Assessments gelistet werden, von einer adaptiven Unterrichtssteuerung bis hin zum „Curriculum embedded assessment“ erstreckt (Klieme et al., 2010) – ein Ansatz, der auf stark formalisierten Vorgaben, höherwertigen Verfahren der Testkonstruktion und auf eine inhaltliche sowie strukturelle Verknüpfung mit summativen Erhebungen basiert. Diese enorme Bandbreite, welche sich unter dem Mantel des formativen Assessment bündelt, hat zur Folge, dass man nach Klauer (2014) eher von (unterschiedlichen) *Konzepten* formativen Assessments sprechen sollte und nicht von formativen Assessments im *Allgemeinen*. Aus dieser Konstellation ergibt sich mit Blick auf die skizzierte Befundlage weiter, dass formative Assessments zwar große Effekte auf Kompetenzsteigerungen erzielen *können*, ausgewiesen durch mehrere Übersichtsartikel (z. B. Black & Wiliam, 1998; Bürgermeister & Saalbach, 2018) bzw. geprüft durch diverse Metaanalysen (z. B. Kingston & Nash, 2011; Hattie, 2012), aber nicht von einem *generellen* Beleg für die Wirksamkeit dieses Ansatzes gesprochen werden kann (Bennett, 2011). Eine entscheidende Konsequenz aus dieser Befundlage ist, dass prinzipielle Gelingensbedingungen oder allgemeingültige Kriterien für wirksames formatives Assessments nur bedingt angeführt werden können. Auch für die Operationalisierung in der Schulpraxis bilanziert Schmidt (2020, S. 25), dass „bislang keine einheitliche Vorstellung davon existiert, welche konkreten Strategien und Praktiken formatives Assessment auf der Handlungsebene kennzeichnen und wie diese miteinander in Verbindung stehen.“ Allerdings haben sich in den letzten Jahren die Komponenten (1) (formative) *Diagnostik*, (2) *Feedback* sowie die Ableitung von

(3) *Fördermaßnahmen* als zentrale Merkmale nahezu aller Umsetzungskonzepte von formativem Assessment etabliert (Hebbecke & Souvignier, 2018). Diese übergeordneten Komponenten wiederum können durch Handlungsschritte, die sich in mehreren Operationalisierungsoptionen bewährt haben, konkretisiert werden (z. B. Maier, 2010). Daher hat es sich als erfolgsversprechend gezeigt, nicht die prinzipielle Wirksamkeit von formativem Assessment belegen zu wollen, sondern konkrete Konzepte, basierend auf der angeführten Grundintention und Struktur, zu entwickeln und zu prüfen, um anschließend *kontextualisierte* Umsetzungshinweise vorzuschlagen (z. B. Hasselhorn et al., 2014; Käter et al., 2016; Hofmann, 2017).

Das im Artikel vorgestellte Projekt TÜV fußt auf der vorgestellten übergreifenden Grundintention formativen Assessments (z. B. Souvignier & Haselhorn, 2018); strukturiert sich über die drei Komponenten Diagnose, Feedback sowie Förderung (Hebbecke & Souvignier, 2018); konkretisiert sich dann jedoch gemäß den angeführten Überlegungen anhand weiterer etablierter Attribute gängiger Förderkonzepte und stellt somit eine mögliche Umsetzungsvariante formativen Assessments dar.

2.2 Feedbackformate

Als empirisch belegt können positive Effekte von Rückmeldungen auf den Wissenserwerb und den Kompetenzaufbau betrachtet werden (z. B. Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Wie angeführt, stellt Feedback eine der konzeptübergreifenden Komponenten nahezu aller Ansätze von formativem Assessment dar (Hebbecke & Souvignier, 2018). Diesen werden nicht zuletzt deshalb große Potentiale zugeschrieben, da die Schülerinnen und Schüler durch die regelmäßigen (zusätzlichen) Lernstandsüberprüfungen kontinuierlich Feedback zu ihrem aktuellen Kompetenzstand erhalten (z. B. Hebbecke & Souvignier, 2018). Angeführt werden muss in diesem Zusammenhang jedoch, dass hinsichtlich der Feedbackwirkungen große Unterschiede festzustellen sind, was neben Domänen- und individuellen Personenmerkmalen vor allem auf formale sowie inhaltliche Attribute der Rückmeldungen zurückzuführen ist (Deci et al., 1999). So spielen vor allem die Variationen des Feedbacks eine große Rolle, wenn es um potentielle Effekte geht (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). In der Feedbackforschung existieren daher mehrere Modelle, welche hinsichtlich Feedbackformen, -dimensionen und -ebenen differenzieren. Eine der am häufigsten zum Einsatz kommenden Klassifikationen stammt von Narciss (2006), in welcher zwischen einfachen und elaborierten Feedbackformen unterschieden wird: Die Bandbreite der Rückmeldungen reicht hier von einer dichotomen Information, ob die angeführte Antwort richtig oder falsch ist (KR, „Knowledge of result/response“), bis zu adaptiven Formen, bei denen z. B. weitere Hinweise, Beispiele oder Lösungsstrategien (KH, „Knowledge on how to proceed“) bzw. gar „Hinweise auf metakognitive Strategien“ (KMC, „Knowledge on meta-cognition“) angeführt werden (Narciss, 2006, S. 23). Nach Mory (2004) sind mit dem Einsatz elaborierter Feedbackformen höhere Effekte zu erwarten – wobei prinzipiell Informationen auf mehreren Ebenen (z. B. Arbeitsverhalten, Arbeitsphasen,

Ergebnis) größere Erfolge versprechen (Hattie & Timperley, 2007). Einschränkend muss jedoch angeführt werden, dass das angebotene Feedback zum einen überhaupt genutzt werden muss (Jónsson et al., 2018), wenn entsprechende positive Effekte erzielt werden möchten, und zum anderen einen bestimmten zielgruppenabhängigen Zuschnitt aufweisen sollte (Narciss, 2014). *Schulisches* Feedback muss daher stets auch unter der Prämisse bereitgestellt werden, dass Schülerinnen und Schüler besonders gestaltete Rückmeldungen benötigen (Zierer et al., 2015). Dies beginnt schon damit, dass es je nach Entwicklungsstufe der Lernenden einen gewissen Umfang nicht überschreiten sollte (Narciss, 2014) – auch wenn prinzipiell weitere Informationen hilfreich sein könnten (Kluger & DeNisi, 1996). Schulisches Feedback muss sich daher stets beschränken und auf Primärinformationen reduzieren (Zierer et al., 2015). Im Idealfall berücksichtigt Feedback individuelle Persönlichkeitsmerkmale und bezieht sich auf individuelle Lerngewohnheiten (Harris et al., 2014; van der Kleij, 2019); bedacht werden muss in diesem Falle jedoch, dass Lehrkräfte derartige Rückmeldungen als belastend und als nicht im täglichen Unterrichtsgeschehen umsetzbar einschätzen (Fischer, 2019). Dennoch sollte das gewählte Format in jedem Falle mehr Informationen als die Gesamtzahl an Fehlern oder Bewertungseinheiten offerieren: Lernförderliche Rückmeldungen im schulischen Kontext sind wertschätzend, zeigen einen Kompetenzstand nachvollziehbar auf und geben Hinweise, wie potentielle Verbesserungen erreicht werden können (Winter, 2015). Rückmeldungen sind jedoch nicht nur in die Richtung der Lernenden, sondern auch in jene der Lehrenden gerichtet: Regelmäßige formative Assessments liefern auch den Lehrkräften die notwendigen Informationen zur Adaption des Unterrichts sowie zur Einleitung von (individuellen) Fördermaßnahmen (Black & Wiliam, 2009). Diese Erkenntnisse der Feedbackforschung, unter besonderer Berücksichtigung der Befunde im Kontext von schulischen Lehr-Lernprozessen, gilt es folglich zu berücksichtigen, wenn Förderkonzepte, die auf formativem Assessment basieren, konzipiert werden. Da Feedback eines der Kernelemente aller Konzepte formativen Assessments darstellt (Schildkamp et al., 2020), wird in der ersten Phase des Projekts TÜV ein Vergleich unterschiedlicher Feedbackformate zu formativen Assessments vorgenommen und untersucht, welche Effekte die verschiedenen Rückmeldeformen auf die Kompetenzentwicklung besitzen.

2.3 Datenbasiertes Fördern

Ein Konzept formativen Assessments, bei welchem die drei Basiskomponenten (1) Diagnostik, (2) Feedback und (3) Förderung (Hebbecker & Souvignier, 2018) nach sehr konkreten Vorgaben umgesetzt werden, ist die Lernverlaufsdagnostik. Dem Ansatz werden große Potentiale bei der Überprüfung von Fördermaßnahmen und bei der Steigerung von Lern- und Kompetenzständen zugeschrieben (Hattie, 2012). Im deutschsprachigen Raum wird in diesem Kontext häufig die terminologische Verortung von Klauer (2011, S. 207) herangezogen:

„Bei der Lernverlaufsdagnostik geht es darum, statt einer punktuellen Testerhebung regelmäßig in relativ kurzfristigen Abständen die Entwicklung einer Kompetenz über einen längeren Zeitpunkt hinweg durch geeignete Tests zu verfolgen.“

Einer der Erfolgsgaranten dieses Konzepts formativen Assessments, die strengen Vorgaben an die psychometrische Güte der eingesetzten Instrumente (Förster et al., 2017), muss mit Blick auf den praktischen Einsatz im Schulalltag gleichermaßen als Hürde gesehen werden: Der bei Lernverlaufsdagnostiken hohe methodologische Anspruch (z. B. Deno, 2003) hat zur Folge, dass Lehrkräfte schon beim Anwenden bzw. Auswerten von existierenden Konzepten in der Regel vor große, partiell unlösbare Herausforderungen gestellt werden (Kopmann, 2021). Die sogar noch weitergehende Idee, dass Tests zu formativen Zwecken idealerweise von den Lehrkräften selbst entwickelt werden, da diese aktuelle Unterrichts- und Kompetenzentwicklungen am besten einschätzen können (Klauer, 2014), ist mit diesem Konzept in der Schulpraxis nur schwer umsetzbar (Börnert, 2011). Blumenthal und Kollegen bilanzieren daher in diesem Kontext zu Recht, dass mittlerweile der tatsächlichen Nutz- und Anwendbarkeit von Lernverlaufsdagnostiken eine besonders herausragende Stellung zukommen muss (Blumenthal et al., 2022). Nicht weniger bedeutsam für die tatsächliche Nutzung ist mittlerweile die Frage, wie entsprechende Konzepte in den Schulalltag implementiert werden können (Souvignier et al., 2016). Ansätze mit schulpraktischem Bezug betonen daher als Konsequenz aus der skizzierten Problematik, dass weniger die *Verfahren* formativen Assessments als der *Kontext* bzw. das *Ziel* entscheidend seien und folglich der verwendeten Konzeptbezeichnung eine eher untergeordnete Rolle zukäme (Breucker & Kuhl, 2022). So existieren mittlerweile diverse Konzepte, die analog zum ursprünglichen Ansatz der Lernverlaufsdagnostik geplant und regelmäßig Grundkompetenzen über eine längere Zeitspanne hinweg immer wieder erfassen, erhobene Daten systematisch aufzeichnen und auswerten sowie den Unterricht anhand der gewonnen Informationen adaptieren bzw. (individuelle) Fördermaßnahmen konzipieren (z. B. Maier et al., 2012) – die Anforderungen an die gewählten Verfahren und die eingesetzten Instrumente sind jedoch mit jenen anderer formativer bzw. summativer Erhebungen im Schulalltag vergleichbar und besitzen nicht den Anspruch, in die Nähe formaler Testverfahren zu gelangen (Balt et al., 2017a; Kopmann, 2021). Die Bezeichnungen für derartige Umsetzungen sind jedoch nicht stringent: Es existieren von Schulpraktikerinnen und -praktikern entwickelte (Alltags-)Konzepte, welche die strengen Vorgaben an die eingesetzten Instrumente nicht oder nicht vollständig erfüllen und dennoch unter dem Label einer Lernverlaufsdagnostik geführt werden (Gebhardt et al., 2021). Weiter werden in diesem Kontext Begriffe wie *testbasiertes Fördern* bzw. *testbasiertes Lernen* (z. B. Fischer et al., 2015), *diagnosebasiertes Fördern* (Fischer, 2022) oder die ursprünglich der Förderdiagnostik entstammenden Bezeichnung *datenbasiertes Fördern* (Urton et al., 2021) genutzt. Partiiell verbergen sich hinter den Begriffen und Konzepten „echte“ Lernverlaufsdagnostiken, die tatsächlich formale Verfahren der Testkonstruktion einsetzen (Klauer, 2014; Gebhardt et al., 2021), häufig wird aber ein anderer

Begriff gewählt, um diese Assoziation nicht zu erzeugen. Betont werden soll jedoch stets, dass diese Umsetzung des formativen Assessments regelmäßig Daten als Grundlage für eine längerfristig ausgerichtete förderorientierte Unterrichtsplanung generiert und die entsprechenden Entwicklungen auch dokumentiert (Rosidin et al., 2020). Der hier verwendete Begriff des *datenbasierten Förderns* kann, so wie er im vorliegenden Projekt konkret belegt ist, in die Liste der Bezeichnungen eingeordnet werden, die zwar strukturelle und intentionale Merkmale der Lernverlaufdiagnostik umsetzen, bei der Testkonstruktion in der Regel aber auf informelle Verfahren zurückgreift. Diese Vorgehensweise muss aber nicht zwingend als Limitation verstanden werden, da so die über aktuelle Problemfelder im Unterricht am besten informierten Akteure, die Lehrerinnen und Lehrer, die für die Diagnose notwendigen Instrumente selbst oder mitentwickeln können (Klauer, 2014). Zudem werden die meisten formativen Assessments mit informellen Verfahren umgesetzt (Ruiz-Primo & Furtak, 2007). Derartige formative Assessments mit datenbasierten Förderkonzepten erweitern damit das Spektrum der „Lernverlaufdiagnostiken“ dahingehend, dass die gewonnenen Daten andere Informationen liefern und folglich auch anders für die Planung der nächsten Lehr-Lernschritte genutzt werden. Zudem wird schon in der Bezeichnung zumeist betont, dass der Schwerpunkt bei diesen Konzepten nicht zwingend auf dem eingesetzten diagnostischen Verfahren liegt, sondern eher auf der Konzeption geeigneter Fördermaßnahmen bzw. welche *Fördermethoden* angewandt werden sollten (Förster et al., 2017). Einige Umsetzungen dieses Ansatzes stellen daher auch vermehrt heraus, dass der diagnostische Vorgang und die so erhobenen Daten „nur“ die Grundlage für den eigentlichen Mehrwert des Verfahrens bilden: Pädagogische Förderentscheidungen mit Hilfe von systematischen Förderkonzepten umsetzen, um individuelle Entwicklungen zu unterstützen (Balt et al., 2017b). Essentiell bei diesen Umsetzungsansätzen ist jedoch, dass die eingesetzten Instrumente direkt mit Material bzw. Hinweisen zur Förderung verknüpft sind (Gebhardt et al., 2021) – Instrumente und Fördermaterialien sind also jeweils Teil eines inhaltlich, methodisch und strukturell zusammengehörenden Konzepts (Kopmann, 2021).

2.4 Förderpotential im Deutschunterricht – Das Projekt TÜV

Mehrere internationale Schulleistungsstudien haben bei deutschen Schülerinnen und Schülern schularten- und jahrgangsstufenübergreifend Defizite in den unterschiedlichen Kompetenzbereichen des Deutschunterrichts festgestellt. So lagen deutsche Schülerinnen und Schüler beispielsweise bei der PISA-Erhebung im Jahre 2018 im Bereich der Lesekompetenz lediglich auf Platz 15; die 15-jährigen Schülerinnen und Schüler aus Österreich (Platz 23) und der Schweiz (Platz 24) schnitten ähnlich schlecht ab (OECD, 2019). Aktuelle Befunde lassen den Rückschluss zu, dass u.a. die Auswirkungen der Corona-Pandemie diese Situation noch zusätzlich verschärft (z. B. Stanat, 2022).

Die skizzierte Problematik verlangt neue Lehr-Lernkonzepte sowie individuelle Förderung (Hußmann & Schröter, 2022). Jene ist jedoch nur möglich, wenn aktuelle Lern- und Kompetenzstände bekannt sind. Es zeigte sich, dass regelmäßig und systematisch eingesetzte formative Assessments signifikante Kompetenzsteigerungen (mit-)bewirken können (Schmidt, 2020; Blumenthal, 2022). Gerade im Deutschunterricht werden mittlerweile Konzepte formativen Assessments, bei denen in unterschiedlichen Variationen regelmäßig Grundkompetenzen erhoben und Ergebnisse dokumentiert werden häufig eingesetzt (z. B. Gebhardt et al., 2021).

Das Projekt TÜV-D (*Testen-Üben-Verstehen im Deutschunterricht*) ging aus einem Projektseminar der universitären Lehrerinnen- und Lehrerbildung mit dem Schwerpunkt „Formative Leistungsmessungen im Schulalltag“ hervor. Die Lehrveranstaltung setzte sich zur einen Hälfte aus Studierenden und zur anderen Hälfte aus Lehrkräften zusammen. Jeweils ein/e Studierende/r und eine Lehrkraft bildeten ein Tandem, welches gemeinsam einen formativen Test konzipierte, einsetzte und auswertete. In der universitären Lehrveranstaltung wurde somit ein informelles Instrument entwickelt, das nach den gängigen an Schulen vorherrschenden Grundsätzen erstellt wurde (Schrader & Helmke, 2014). Die Studierenden sollten auf diese Weise unter anderem einen Eindruck gewinnen, wie Lehrkräfte einen formativ genutzten Test konzipieren und welche Problemfelder sich daraus ergeben (Hofmann et al., 2019). Weiter sollten die angehenden Lehrerinnen und Lehrer bei der auf die erhobenen Daten basierenden Unterrichtsplanung beteiligt sein (Maier, 2015).

Ausgangspunkte für das nachfolgende Projekt TÜV-D waren das prinzipielle Interesse aller beteiligten Akteure, die Potentiale formativen Assessments gemeinsam mit Studierenden näher zu untersuchen sowie das notwendige Reagieren auf nach Ansicht der Lehrkräfte bestehende Defizit bei grundlegenden Kompetenzen im Deutschunterricht. Das Projekt setzt somit nicht bei der Einführung bzw. erstmaligen Erarbeitung von Kompetenzen an, sondern beabsichtigt eigentlich bereits ausgeprägte Basiskompetenzen zu stärken bzw. immanente Probleme durch (zusätzliche) Fördermaßnahmen zu beseitigen (Streber, 2016). Das als Projekttitel gewählte Akronym TÜV gibt hierbei den grundlegenden Ablauf der Fördermaßnahme bereits an:

- *T – Testen*: Den Einstieg in den Förderprozess stellt das formative Assessment dar. Das bedeutet, dass gemäß etablierter Gelingensbedingungen für Fördermaßnahmen und Übungsphasen (Lerche, 2019) bzw. definierender Handlungsschritte formativen Assessments (Maier, 2010) zuerst geklärt wird, welchen Kompetenzstand die Schülerinnen und Schüler haben und welche Lücken möglicherweise vorhanden sind (Hattie & Timperley, 2007; Black & Wiliam, 2009; Maier, 2015). Der nachfolgend noch näher erläuterte Kompetenztest „Sprache untersuchen“ diagnostiziert den aktuellen Lern- bzw. Kompetenzstand in dieser Domäne und fokussiert (in der ersten Phase des Projekts) auf Bereiche und Niveaustufen, die mit Blick auf die Curriculumchronologie schon ausgeprägt sein sollten und essentielle Grundlage für folgende Kompetenzerweiterungen darstellen. Die Aufgaben hierzu wurden nun nicht mehr wie in der Seminarveranstaltung von den

beteiligten Lehrkräften in Zusammenarbeit mit den Studierenden entwickelt, sondern einem zugänglichen Aufgabenpool des Instituts zur Qualitätsentwicklung im Bildungswesen Berlin (IQB) entnommen (siehe 4.5).

- *Ü – Üben:* Die Analyse des formativ genutzten Kompetenztests ermöglicht es, „substanzielle Aussagen zur Ausgangsproblematik“ zu treffen (Jürgens, 2022, S. 557), um gezielte (individuelle) Fördermaßnahmen zu formulieren (Streber, 2016) und Lehr-Lernsequenzen entsprechend zu adaptieren (Maier, 2015). Im Projekt wird in den nachfolgenden Unterrichtsstunden konkret auf die Kompetenzbereiche eingegangen, die durch das formative Eingangsassessments als (noch) nicht vollständig ausgeprägt identifiziert wurden.
- *V – Verstehen:* Ziel der konkret ausgerichteten (zusätzlichen) Übungsaufgaben und -phasen ist es, zu einem „Verständnis von Inhalten und Anwendungen“ zu gelangen (Lerche, 2019, S. 212). Im Projekt TÜV wird an dieser Stelle nicht davon ausgegangen, dass die Förder- und Übungsmaßnahmen (Schritt 2 „Üben“) bei allen Schülerinnen und Schülern in kürzester Zeit die intendierten Effekte erzielen. Daher ist das Projekt in Anlehnung an diverse Grundsätze des datenbasierten Förderns (siehe 2.3) angelegt und konstituiert sich durch regelmäßige Kompetenztests, mit welchen der durch die zurückliegende Intervention tatsächlich erreichte aktuelle Kompetenzstand ermittelt werden kann. Die Projektstruktur orientiert sich damit an einer methodisch-didaktischen Planungsspirale (Meyer, 2014), das bedeutet, dass das real erworbene Kompetenzniveau regelmäßig diagnostiziert wird und Ausgangspunkt für die nächste Übungs- und Förderphase darstellt (Klauer, 2014).

Bei der Konzeption und Operationalisierung des Projekts wurde zudem darauf geachtet, die Perspektive der Lehrerinnen und Lehrer im Schulalltag gebührend zu berücksichtigen (z. B. Kopmann, 2021). Die an der Planung und Umsetzung beteiligten Lehrkräfte betonten vor allem, dass sie einen permanenten Zeitdruck spürten und keine aufwendigen zusätzlichen Aufgaben übernehmen könnten. Diese Einschätzungen decken sich mit bekannten Befunden aus der Lehrerbelastungsforschung: So geben viele Lehrkräfte als Ursache für Überlastung und Krankheit unter anderem enorme Vorbereitungs- und Korrekturbelastungen an (z. B. Sprenger, 2019). Die Umsetzung formativer Assessments wird als besonders fordernd bewertet (Bennett, 2011). Die Lehrkräfte im Projekt betonten unter Rückgriff auf die angeführten Thesen, dass zeitaufwendige und korrekturintensive Förderkonzepte ihrer Meinung nach unter den augenblicklichen Rahmenbedingungen im Schulalltag nicht ohne Weiteres umzusetzen seien. Daher sollten Tests entwickelt werden, die auf basale Kompetenzen fokussieren, nur wenig Zeit bei der Durchführung und Korrektur in Anspruch nehmen und über einen längeren Zeitraum Lernstände erfassen (Maier et al., 2012).

Da unter anderem die DESI-Studie als mögliche Begründung für ungenügende Lese- und Schreibleistungen sowie eine nur partiell ausreichend ausgeprägte Kompetenz im Bereich der Textkonstruktion eine mangelnde Bewusstheit für grammatikalische und

stilistische Phänomene identifiziert (z. B. Helmke & Klieme, 2008), wurde dieser Kompetenzbereich für den ersten Teil des Projekts ausgewählt. Zudem nimmt der Bereich „Sprache und Sprachgebrauch untersuchen“ innerhalb der vier Kompetenzbereiche (zudem „Sprechen und zuhören“, „Schreiben“ und „Lesen – mit Texten und Medien umgehen“) der KMK-Standards für den Mittleren Schulabschluss eine herausragende Stellung ein, da er „in Beziehung zu jedem der drei anderen Bereiche steht“ (KMK, 2004, S. 7).

2.5 Fragestellungen der vorliegenden Studie

Ziel des Projekts TÜV ist, ein datenbasiertes Förderkonzept für verschiedene Kompetenzbereiche im Deutschunterricht zu konzipieren und zu evaluieren. Neben der Prüfung, ob sich das entwickelte Instrument in der Unterrichtspraxis tatsächlich bewährt, standen bei der Pilotstudie vor allem die Effektivität unterschiedlicher Feedbackformate im Fokus. Folgende Forschungsfragen sollen im ersten Auswertungsschritt beantwortet werden:

1. Welche Kompetenzzuwächse erzielen die Schülerinnen und Schüler mit dem datenbasierten Förderkonzept im Vergleich zu herkömmlich unterrichteten Schülerinnen und Schülern?
2. Lassen sich Unterschiede bei den Lern- und Kompetenzzuwächsen durch unterschiedliche Feedbackformate im datenbasierten Förderkonzept erklären?

3 Methode

3.1 Stichprobe

Die Pilotstudie des Projektes TÜV-D wurde an einer bayerischen Realschule im städtischen Großraumbereich in allen siebten Klassen ($N = 8$) im Schuljahr 2013/2014 durchgeführt. Die Auswahl der Schule und damit der Klassen kann als Gelegenheitsstichprobe bezeichnet werden, da alle vier in der siebten Klasse unterrichtenden Lehrkräfte (jede Lehrkraft unterrichtete in zwei Klassen das Fach Deutsch) am vorausgegangenen universitären Projektseminar „Formative Leistungsmessungen im Schulalltag“ als Tutorinnen und Tutoren teilnahmen (siehe 2.4) und in diesem Zusammenhang ihr Interesse an einer vertieften Kooperation signalisierten. Die Lehrkräfte griffen nicht in den Projektunterricht ein, waren jedoch anwesend (siehe 4.4). Sämtliche für das Projekt relevante Lehr-Lernprozesse im Fach Deutsch sowie die Durchführung der Kompetenztests übernahmen Mitarbeiterinnen und Mitarbeiter der Universität bzw. Studierende. Den Schülerinnen und Schülern wurde das Projekt vor Beginn ausführlich erläutert. Anschließend konnten die Jugendlichen frei entscheiden, ob sie an dem Förderprogramm mit formativen Assessments teilnehmen wollten oder nicht. Zudem konnten die Schülerinnen und Schüler jederzeit aus dem Projekt aussteigen. Alle

Erziehungsberechtigten wurden informiert, die Schulleitung gab ihr Einverständnis, erteilte jedoch strikte Vorgaben hinsichtlich der genehmigten Datenerfassung sowie des zeitlichen Umfangs des Projekts. Schülerinnen und Schüler, die an einem Termin fehlten, wurden nicht berücksichtigt, so dass die Stichprobe letztlich $N = 237$ Schülerinnen und Schüler ($M = 13,64$ Jahre, $SD = 0.84$; 54,43 % weiblich) umfasste.

3.2 Design

Das Forschungsdesign ist in Abbildung eins grafisch dargestellt. Vor dem Projektbeginn wurden alle beteiligten Lehrkräfte ausführlich instruiert und hatten genügend Zeit, Rückfragen zu stellen. Alle Informationen zur Durchführung wurden zudem in einem Lehrerhandbuch, das in Papier und Digitalform vorlag, zusammengefasst. Weiter gab es jeden Freitag ein Projekttreffen, in welchem der weitere Projektablauf sowie mögliche Probleme besprochen wurden. Das Pilotprojekt startete in der zweiten Schulwoche nach den Sommerferien mit einer Einführungsveranstaltung für Schülerinnen und Schüler sowie Lehrkräfte. Im Rahmen dieser Veranstaltung wurde auch die Zuteilung zu den verschiedenen Untersuchungsgruppen vorgenommen (Abb. 1).

Da in allen acht teilnehmenden Klassen nahezu gleich viele Schülerinnen und Schüler untergebracht waren, sah der Untersuchungsplan vor, jeweils zwei Klassen zufällig (per Los) auf die drei Experimentalgruppen sowie die Kontrollgruppe zu verteilen. Folglich waren auch die vier konzipierten Untersuchungsgruppen nahezu gleich groß (Kontrollgruppe, KG, $n_{(KG)} = 58$; Experimentalgruppe A, EG-A, $n_{(EG-A)} = 58$; Experimentalgruppe B, EG-B, $n_{(EG-B)} = 62$; Experimentalgruppe C, EG-C, $n_{(EG-C)} = 59$); bedeutsame Unterschiede bei der Zusammensetzung (Geschlecht, Alter) konnten nicht festgestellt werden; Kompetenz- und Leistungsunterschiede wurden nur domänen- und kompetenzspezifisch durch den Pretest (Kompetenztest 1) und aufgrund der Vorgaben der Schulaufsicht nicht durch Zensuren oder vorherige Schulleistungen erfasst. Auf eine vorab angedachte zufällige Zuteilung der Lernenden innerhalb der Klassen zu einer der vier Untersuchungsgruppen musste aus organisatorischen Gründen verzichtet werden – vor allem der innerhalb eines Klassenverband dann bestehende unterschiedliche Unterrichtsablauf (z. B. Kompetenztests vs. keine Kompetenztests oder Selbstkorrektur vs. Fremdkorrektur) wurde von den Lehrkräften sowie den während des Projekts unterrichtenden Universitätsmitarbeiterinnen abgelehnt. Vorteil der umgesetzten Konstellation ist jedoch, dass der Unterricht im regulären Klassenverband nach gewohnten Abläufen ohne einschneidende Änderungen stattfand, die Intervention ohne großen organisatorischen Aufwand in den Unterrichtsalltag implementiert werden kann und die Ergebnisse des Projekts so direkten Bezug zur Schulpraxis besitzen (z. B. Brown, 1992; Pissarek & Wild, 2018).

Zu Beginn der Durchführungsphase wurde bei allen Lernenden (alle Untersuchungsgruppen, $N = 237$) der aktuelle Kompetenzstand in der Domäne „Sprache untersuchen“ erfasst (Pretest, Kompetenztest 1). In diesem Zusammenhang wurden auch

Posttest & AT6 ⁴		t6		Kompetenztest 6 „Sprache untersuchen“ Abschluss erhebung		Posttest & AT6 ⁴			
Treatment		LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“		
		t5	Kompetenztest 5 „Sprache untersuchen“		X	t5	Kompetenztest 5 „Sprache untersuchen“		
		LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“		
		t4	Kompetenztest 4 „Sprache untersuchen“		X	t4	Kompetenztest 4 „Sprache untersuchen“		
		LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“		
		t3	Kompetenztest 3 „Sprache untersuchen“		X	t3	Kompetenztest 3 „Sprache untersuchen“		
		LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“		
		t2	Kompetenztest 2 „Sprache untersuchen“		X	t2	Kompetenztest 2 „Sprache untersuchen“		
LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“	LL ³	2 h/Woche Unterricht „Sprache untersuchen“	Treatment			
Pretest & ET1 ²		t1		Kompetenztest 1 „Sprache untersuchen“ Einstiegs erhebung				Pretest & ET1 ²	
Experimentalgruppen (EG)		EG-A: Schüler selbstkorrektur ¹		Kontrollgruppe (KG)					
		EG-B: Lehrerkorrektur ¹							
		EG-C: Lehrerkorrektur ¹ mit Hinweisen							
		zufällige Zuordnung							

Anmerkungen: ¹Die Formulierung erfolgte aus Gründen der Reduktion, gemeint sind stets Schülerinnen und Schüler sowie Lehrerinnen und Lehrer; ²ET1 = Fragebogen zu Beginn des Projekts; ³LL = Lehr- Lernprozess Fragebogen zum Abschluss des Projekts; ⁴AT6 = Fragebogen zum Ende des Projekts

Abbildung 1: Design der Vorstudie TÜV-D

soziodemographische Daten sowie bekannte Einflussfaktoren (z. B. Arbeitsverhalten, Motivation, schulische Selbstwirksamkeit) erhoben (Fragebogen ET1). In jeder Klasse wurde anschließend für einen Monat zwei Stunden pro Woche die Kompetenz „Sprache untersuchen“ unterrichtet. Alle Klassen nutzten die gleichen, von den Lehrkräften der teilnehmenden Klassen sowie den Mitarbeiterinnen und Mitarbeitern der Universität erstellten, Materialien. Grundlage dieser waren die vom Institut zur Qualitätsentwick-

lung (IQB) zur Verfügung gestellten Unterlagen und didaktischen Kommentierungen (IQB, 2021) – passgenau zu den gewählten Kompetenzbereichen der Bildungsstandards. Den Unterricht in diesen Stunden übernahmen in allen Klassen zwei Studentinnen, die immer als Tandem agierten. Die jeweiligen Lehrkräfte nahmen beobachtend am Unterricht teil, griffen jedoch nicht ein. Der Stundenablauf wurde vorher durch eine detaillierte Unterrichtsskizze genau festgelegt. Die unterrichtenden Studentinnen wurden im Vorfeld während eines Praxisseminars geschult, hatten daher alle acht Unterrichtsstunden bereits mehrfach gehalten (partiell vor anderen Studierenden sowie Hochschuldozierenden) und erhielten während der Vorbereitungsphase kriteriengeleitete Rückmeldungen. Um einen möglichst gleichen Unterrichtsablauf zu gewährleisten bzw. potentielle Abweichungen zu erfassen, kontrollierten zwei weitere Studierende unabhängig voneinander mit einem Kriterienraster („Check up“) die Durchführung der Unterrichtsstunden. Alle Schülerinnen und Schüler der Experimentalgruppen (EG) schrieben zusätzlich zu den beschriebenen Unterrichtsstunden jeden Freitag einen auf die gleichen Kompetenzbereiche und analog aufgebauten Paralleltest (Kompetenztests 2–5; siehe 4.5), der zehn Minuten in Anspruch nahm. Am darauffolgenden Montag werteten die Lernenden der EG-A den Test anhand einer Musterlösung (Lösungsbogen zum Anlegen) selbst aus; die Schülerinnen und Schüler der EG-B u. EG-C erhielten den von den Lehrkräften anhand der Musterlösung korrigierten Test zurück, wobei die EG-B eine Rückmeldung zum Status (*richtig* vs. *falsch*) und ggf. die korrekte Lösung bekamen, die Probanden der EG-C zusätzlich noch individuelle Fehlerhinweise und Förderempfehlungen. Bestandteil des vorgegebenen Unterrichtsablaufs war bei allen Experimentalgruppen eine Auseinandersetzungphase mit dem korrigierten Text sowie ggf. mit den Rückmeldungen. In dieser Phase konnten die Lernenden auch Verständnisfragen stellen bzw. sich untereinander austauschen. Am Projektende schrieben alle Schülerinnen und Schüler (alle Experimentalgruppen sowie die Kontrollgruppe) einen Abschlusstest (Posttest, Kompetenztest 6). Lernende und Lehrkräfte wurden zu diesem Zeitpunkt zudem mittels eines weiteren Fragebogens (AT6) zum Projekt befragt. Abschließend füllten die Schülerinnen und Schüler einen Lerndokumentationsbogen aus, in dem die Ergebnisse der sechs Kompetenztests, unter Berücksichtigung der diversen Kompetenz- und Komplexitätsbereiche (siehe 4.5) sowie potentielle Entwicklungsziele eingetragen wurden (Winter, 2015). Zudem konnten sich die Schülerinnen und Schüler in einer Diskussionsrunde frei zum Projekt äußern oder ihre Kommentare schriftlich in eine aufgestellte Urne einwerfen. Die ursprünglich geplanten Interviews mit Lernenden und Lehrenden, ein zu Beginn des zweiten Schulhalbjahres (nach 14 Wochen) angedachter Follow-Up-Test (Kompetenztest 7) sowie eine zweite Projektphase am Ende des zweiten Schulhalbjahres, die sich unter anderem auf die Ergebnisse dieser siebten Messung nach einer längeren Unterrichtszeit, in welcher die Kompetenz „Sprache untersuchen“ nicht Gegenstand im Lehr-Lernprozess war, beziehen sollte, blieben leider aus. Hauptgrund hierfür war, dass die Schulleitung keine weiteren Zeitressourcen für ein wissenschaftlich begleitetes Lern- und Förderprogramm während der Regelschulzeit sah.

Der vorliegende Beitrag stellt erste Befunde zur Kompetenzentwicklung in den unterschiedlichen Experimentalgruppen im Vergleich zur Kontrollgruppe vor.

3.3 Instrumente

3.3.1 Fragebogen ET1

Mit dem *Einführungsfragebogen (ET1)* wurden neben soziodemographischen Daten weitere bekannte Einflussfaktoren erhoben:

- *Anstrengung und Ausdauer im Fach Deutsch*: Die Anstrengung/Ausdauer wurde mit drei Items ($\alpha = .68$) über eine adaptierte PISA-Skala erfasst (Kunter et al., 2002); Beispielimitem „Im Fach Deutsch arbeite ich sehr fleißig.“ (Range: 1 = „stimmt gar nicht“ bis 4 = „stimmt ganz genau“).
- *Intrinsische Motivation zur Teilnahme am testbasierten Förderkonzept*: Die Motivation zur Teilnahme wurde mit 12 Items ($\alpha = .79$) über die adaptierte „Kurzskaala intrinsischer Motivation (KIM)“ erfasst (Wilde et al., 2009); Beispielimitem „Ich fand die Teilnahme am Förderunterricht sehr interessant.“ (Range: 1 = „stimme überhaupt nicht zu“ bis 4 = „stimme völlig zu“).
- *Selbstkonzept im Fach Deutsch*: Das Selbstkonzept wurde mit zehn Items ($\alpha = .89$) über eine adaptierte DESI-Skala erfasst (Wagner et al., 2009); Beispielimitem: „Ich bin im Fach Deutsch ganz gut.“ (Range: 1 = „stimme überhaupt nicht zu“ bis 4 = „stimme völlig zu“).
- *Fachbezogene schulische Selbstwirksamkeit in Deutsch*: Die Selbstwirksamkeit wurde mit acht Items ($\alpha = .89$) über eine DESI/PISA-Skala erfasst (Wagner et al., 2009); Beispielimitem: „Ich bin sicher, dass ich auch sehr schwierigen Unterrichtsstoff verstehen kann.“ (Range: 1 = „stimmt gar nicht“ bis 4 = „stimmt ganz genau“).
- *Fachspezifische Lernstrategien in Deutsch*: Die domänenspezifischen Lernstrategien wurden mit vier Items ($\alpha = .78$) erfasst (Baumert et al., 1997); Beispielimitem „Wenn ich neue Dinge für Deutsch lerne, denke ich darüber nach, wie der neue Stoff zum Gelernten passt.“ (Range: 1 = „stimme überhaupt nicht zu“ bis 4 = „stimme völlig zu“).
- *Einstellungen zum Fach Deutsch*: Die Einstellungen wurden mit sechs Items ($\alpha = .79$) über eine DESI-Skala erfasst (Wagner et al., 2009); Beispielimitem: „Mir macht das Fach Deutsch Spaß.“ (Range: 1 = „stimme überhaupt nicht zu“ bis 4 = „stimme völlig zu“).
- *Interesse am Kompetenzbereich „Sprache untersuchen“*: Das Interesse wurde mit vier Items ($\alpha = .82$) über eine adaptierte PISA-Skala erfasst (Kunter et al., 2002); Beispielimitem: „Wie interessant findest du das Thema „Sprache untersuchen?“ (Range: 1 = „überhaupt nicht bis 4 = „sehr“).

Die angegebenen Kennwerte zur Reliabilität beziehen sich jeweils auf die vorliegende Stichprobe der Studie.

3.3.2 Kompetenztest(s) „Sprache untersuchen“

Für die sechs Messzeitpunkte (Abb. 1) wurden sechs Paralleltests zusammengestellt, die alle die gleichen Kompetenzbereiche in den gleichen Abstufungen mit der gleichen Struktur (Anzahl der Aufgaben, Form der Aufgaben) erfassen. Beim Paralleltestverfahren ist der Test folglich bei jedem Messzeitpunkt neu, misst aber bei vergleichbarer Gestaltung sowie ähnlichen Aufgaben stets das gleiche Merkmal und besitzt stets den gleichen Schwierigkeitsgrad (Klauer, 2011).

Anders als bei den Instrumenten, die während der universitären Lehrveranstaltung eingesetzt wurden (siehe 2.4), sind die verwendeten Aufgaben im Projekt TÜV-D nicht von den beteiligten Akteuren selbst konzipiert worden, sondern wurden dem für Lehrkräfte frei zugänglichen Aufgabenuniversum zu den VERA-Vergleichsarbeiten des Instituts für Qualitätsentwicklung im Bildungswesen (IQB) entnommen und erfüllen damit die vom IQB für die VERA-Vergleichsarbeiten angelegten Standards (KMK, 2015). Die Aufgabenentwicklung kann folgendermaßen kurz skizziert werden (IQB, 2021, o. S.):

- „Unter der Führung des IQB und der Betreuung von Wissenschaftlerinnen und Wissenschaftlern an Universitäten und Hochschulen werden die Aufgaben von einem Team erfahrener Lehrkräfte vorgeschlagen.
- Nach einer eingehenden Überprüfung sowie einer psychometrischen Bewertung findet eine Überarbeitung statt.
- Vor dem endgültigen Einsatz durchlaufen alle Aufgaben eine weitere Pilotierungsphase.“

Die Aufgaben werden in Testmodulen sowie Beispiel- und Übungsaufgaben teilweise ohne Beschränkung, teilweise nur Lehrkräften zur Verfügung gestellt; die vollständige Dokumentation der Aufgabenentwicklung ist auf dem Internetauftritt des IQB transparent dargestellt (IQB, 2021).

Analog dem Vorgehen bei der Testkonstruktion für Vergleichsarbeiten wurden auch im Projekt TÜV-D damit lediglich Aufgaben nach einem vorgegebenen Schema zu Tests zusammengestellt. Dieses Grundschema sieht für alle sechs im Projekt TÜV-D eingesetzten Paralleltests vor, zu drei verschiedenen Kompetenzbereichen der Gesamtkompetenz „Sprache untersuchen“ (KMK, 2004) jeweils eine Aufgabe in je einem der drei zur Verfügung stehenden Schwierigkeitsstufen anzubieten. Es wurden stets die sechs zuletzt eingestellten Aufgaben ausgewählt. Vereinzelt fand eine geringfügige kontextuelle Adaption statt – Inhalt, Format und Struktur wurden ohne Einschränkung übernommen. Jeder Paralleltest besteht damit aus 27 Aufgaben.

Die Kompetenz „Sprache untersuchen“ bzw. „Sprachgebrauch“ kann nach der Kompetenzstrukturdarstellung in den Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss in der Domäne „Leistungen von Sätzen und Wortarten kennen und für Sprechen, Schreiben und Textuntersuchung nutzen“ in folgende Kompetenzbereiche untergliedert werden (KMK, 2004, S. 16):

- *Kompetenzbereich 1 „Wörter untersuchen“*: Wortebene, Wortbildung, Wortarten;
- *Kompetenzbereich 2 „Sätze untersuchen“*: Satzstrukturen (z. B. Haupt- u. Nebensatz, Satzglieder), Satzebene (z. B. Satzarten, Satzreihe u. -gefüge);
- *Kompetenzbereich 3 „grammatikalische Kategorien untersuchen“*: Tempus, Modus, Aktiv/Passiv, Genus, Numerus, Kasus, Steigerung.

Jeder Kompetenzbereich ist in Anlehnung an das Kompetenzstufenmodell der KMK-Bildungsstandards (KMK, 2014) in die drei Kompetenzstufen (1) *Mindeststandard*, (2) *Regelstandard*, (3) *Regelstandard plus* unterteilt. Die im ursprünglichen KMK-Kompetenzstufenmodell weiter vorgesehene vierte Stufe (*Optimalstandards*), die „bei ausgezeichneten individuellen Lernvoraussetzungen [...] die Erwartungen der Bildungsstandards übertreffen“, wurde in dieser Projektphase nicht berücksichtigt – nicht zuletzt, da es im Projekt TÜV-D um das Identifizieren von Förderpotential hinsichtlich möglicher bestehenden Defizite und nicht um die Förderung hochleistender oder gar hochbegabter Lernenden ging (z. B. Kopmann, 2021) – was mit Blick auf individuelle Förderung im Umgang mit Heterogenität selbstredend ebenfalls eine bedeutsame Perspektive darstellte (z. B. Fischer, 2022).

3.4 Statistische Analysen

Die Auswertung der Daten erfolgte mithilfe des Datenauswertungsprogrammes SPSS 28. Bei allen Analysen wurde auf die in der Sozialwissenschaft gebräuchliche Signifikanzschwelle von $p < .05$ zurückgegriffen. Wie bereits angeführt, wurden nur vollständige Datensätze in die Analyse eingeschlossen ($N = 237$; $n_{(KG)} = 58$; $n_{(EG-A)} = 58$; $n_{(NG-B)} = 62$; $n_{(NG-C)} = 59$).

Um zu untersuchen, ob das auf formativen Assessments beruhende datenbasierte Förderkonzept (Unterschied Kontrollgruppe vs. Experimentalgruppe ohne Unterscheidung hinsichtlich der Rückmeldeviationen; Forschungsfrage zwei) einen Einfluss auf den Kompetenzzuwachs (Differenz Pre- zu Posttest) in der Domäne „Sprache untersuchen“ besitzt, wurde eine einfaktorielle Kovarianzanalyse (ANCOVA) mit einem zweistufigen messwiederholten Innersubjektfaktor „Messzeitpunkt“ und dem zweistufigen Zwischensubjektfaktor „Stichprobe (Kontrollgruppe vs. Experimentalgruppen)“ gerechnet. Die bei der Vorstellung des Einführungsfragebogens (ET1) vorgestellten bekannten Einflussfaktoren sowie die soziodemographischen Daten „Alter“ und „Geschlecht“ wurden als Kovariaten implementiert. Gemäß der Forschungsfrage eins wurde erwartet, dass ein signifikanter Interaktionseffekt zwischen den unabhängigen Variablen „Messzeitpunkt“ und „Gruppenzugehörigkeit“ vorliegt.

Für die Prüfung der zweiten Forschungsfrage wurde eine einfaktorielle ANCOVA mit dem zweistufigen messwiederholten Innersubjektfaktor „Messzeitpunkt“ und dem vierstufigen Zwischensubjektfaktor „Stichprobe (Kontrollgruppe vs. Experimentalgruppe A vs. Experimentalgruppe B vs. Experimentalgruppe C)“ sowie den vorgestellten Kovariaten gerechnet. Mit Blick auf die Forschungsfrage zwei wurde wiederum

angenommen, dass sich ein statistisch bedeutsamer Interaktionseffekt zwischen den unabhängigen Variablen „Messzeitpunkt“ und „Gruppenzugehörigkeit“ zeigt. Um festzustellen, welche Gruppen sich gegebenenfalls signifikant voneinander unterscheiden, werden Post-Hoc-Tests berechnet.

Im Vorfeld wurden die Voraussetzungen für das gewählte Verfahren geprüft; die Residuen waren gemäß dem Shapiro-Will-Test normalverteilt ($p > .05$), die Homogenität der Varianzen wurde durch den Levene-Test untersucht ($p = .745$) die Homogenität der Regressionssteigungen wurde durch die Bildung der entsprechenden Interaktionsterme nachgewiesen, welche für alle Kovariaten vorlag ($p > .05$).

4 Ergebnisse

Mittelwerte und Standardabweichungen der Kontrollvariablen sowie der abhängigen Variable sind in Tabelle 1 dargestellt. Der Kompetenzstand im Bereich „Sprache untersuchen“ ist zu Beginn des Projekts (Pretest, Kompetenztest 1) über das komplette Sample hinweg eher niedrig; es findet jedoch in allen Gruppen eine deutliche Steigerung bis zum Abschlusstest (Posttest, Kompetenztest 6) statt.

Zur Beantwortung der ersten Forschungsfrage, ob die Teilnahme am testbasierten Förderkonzept Effekte auf die Kompetenzentwicklung besitzt, wurde mit einer ein-faktoriellen Kovarianzanalyse geprüft, inwiefern sich die drei Experimentalgruppen zusammen ($n_{(EG-ges)} = 179$) und die Kontrollgruppe ($n_{(KG)} = 58$) bedeutsam voneinander unterscheiden. Untersucht wurde die Kompetenzentwicklung (in korrekt bearbeiteten Aufgaben/max. 27) vom ersten (Pretest) zum sechsten Messzeitpunkt (Posttest) unter Berücksichtigung des Einflusses der angeführten Kontrollvariablen. Die ANCOVA mit Messwiederholung mit Greenhouse-Geisser-Korrektur zeigt, dass unter Bereinigung des Einflusses der Kontrollvariablen ein hoch signifikanter Haupteffekt für die Interaktion zwischen „Messzeitpunkt“ und „Gruppenzugehörigkeit“ vorliegt ($F(2, 13) = 36.38$; $p = .001$; $\eta^2 = .511$). Folglich konnten in dieser Studie die Schülerinnen und Schüler, die regelmäßig formative Leistungsmessungen im Bereich „Sprache untersuchen“ schreiben (EG_{gesamt} , $n_{(EG-ges)} = 179$), größere Kompetenzsteigerungen erzielen als die Lernenden, welche die Thematik ohne ein datenbasiertes Förderkonzept bearbeiten (KG, $n_{(KG)} = 58$).

Im Zuge der Klärung der zweiten Forschungsfrage wurden die unterschiedlichen Rückmeldeviationen in den drei Experimentalgruppen berücksichtigt. Erwartungsgemäß wird der Test der Zwischensubjekteffekte signifikant ($F(1.83, 86.19) = 70.34$; $p = .001$, $\eta^2 = .646$), damit existieren bedeutsame Unterschiede zwischen den Gruppen unter Berücksichtigung der Kontrollvariablen. Bonferroni-korrigierte Post-Hoc-Tests zeigten nach der Bereinigung um die Kontrollvariablen signifikant höhere Kompetenzsteigerungen in der Experimentalgruppe C (Korrektur durch die Lehrkraft mit individuellen Fehlerhinweisen und Förderempfehlungen) ($p < .001$, $M_{Diff} = 3.18$, 95 % - CI[2.13, 4.89]) und der Experimentalgruppe B (Korrektur richtig/falsch durch die Lehrkraft) ($p < .001$, $M_{Diff} = 5.78$, 95 % - CI[3.64, 7.57]) im Vergleich zur Kontrollgrup-

Tabelle 1: Mittelwerte der Kontrollvariablen und der abhängigen Variablen

Skala	Kontrollgrup-		Experimentalgruppen (EG)		
	Gesamt	pe (KG) (n = 58)	EG-A (n = 58)	EG-B (n = 62)	EG-C (n = 59)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Anstrengung und Ausdauer im Fach Deutsch	2.62 (0.44)	2.49 (0.42)	2.63 (0.38)	2.68 (0.51)	2.64 (0.46)
Intrinsische Motivation zur Teilnahme	3.02 (0.92)	3.11 (1.09)	3.13 (1.15)	2.98 (0.94)	2.91 (0.96)
Selbstkonzept im Fach Deutsch	3.61 (0.56)	3.55 (0.46)	3.48 (0.55)	3.67 (0.51)	3.58 (0.58)
Fachbezogene schulische Selbstwirksamkeit in Deutsch	2.56 (1.02)	2.49 (0.66)	2.71 (0.79)	2.47 (0.60)	2.53 (0.98)
Fachspezifische Lernstrategien in Deutsch	2.31 (0.66)	2.19 (0.59)	2.34 (0.61)	2.48 (0.71)	2.21 (0.63)
Einstellungen zum Fach Deutsch	2.14 (1.17)	2.18 (1.07)	2.30 (1.12)	2.08 (1.11)	2.03 (1.24)
Interesse am Kompetenzbereich „Sprache untersuchen“	2.04 (1.03)	2.01 (1.12)	2.33 (1.09)	1.88 (1.14)	2.09 (1.18)
Bewertung Kompetenztest „Sprache untersuchen“ Pretest ¹	13.13 (5.81)	12.75 (5.93)	14.94 (6.13)	13.23 (6.09)	12.46 (5.79)
Bewertung Kompetenztest „Sprache untersuchen“ Posttest ¹	17.63 (5.45)	16.41 (5.87)	16.44 (7.32)	18.04 (7.09)	19.87 (6.98)

Anmerkungen:¹ maximal 27 korrekt gelöste Aufgaben

pe. In der Experimentalgruppe A (Schülerselbstkorrektur mit Musterlösung) lag eine niedrigere Kompetenzentwicklung als in der Kontrollgruppe vor ($M_{\text{Diff}} = 0.13$, 95 % - CI[0.04, 3.87]), diese wird jedoch nicht signifikant ($p = .610$).

5 Diskussion

Dieser Beitrag stellt Befunde zur Effektivität von Förderkonzepten mit formativen Assessments im Deutschunterricht vor. Im Hinblick auf die erste Forschungsfrage, inwieweit ein Unterrichtskonzept mit regelmäßig eingesetzten formativen Assessments größere Kompetenzzuwächse ermöglicht als ein herkömmlicher Unterricht ohne derartige Verfahren, zeigte sich, dass die Schülerinnen und Schüler, die regelmäßig an formativen Kompetenztests teilnahmen, sich stärker verbesserten als die Lernenden, die nicht mit diesem Ansatz unterrichtet wurden. Allerdings verbesserten sich alle Gruppen vom ersten bis zum sechsten Messzeitpunkt signifikant.

Betrachtet man im Rahmen von Forschungsfrage zwei die Effekte der unterschiedlichen Feedbackinterventionen, so zeigte sich, dass erwartungskonform die Schülerinnen und Schüler, die Rückmeldungen mit individuellen Fehlerhinweisen und Förderempfehlungen von den Lehrkräften erhielten (Experimentalgruppe C), die größten Kompetenzsteigerungen erzielen konnten (vgl. Hattie, 2012); der Unterschied zu den Lernenden, bei denen die Lehrkräfte lediglich falsche Aufgaben kennzeichneten und entsprechend verbesserten (Experimentalgruppe B), jedoch deutlich geringer als angenommen ausfiel (vgl. Narciss, 2006). Überraschend war weiterhin, dass es keinen signifikanten Unterschied zwischen den Lernenden der Kontrollgruppe und jenen der Experimentalgruppe A, welche formative Assessments schrieben und die Korrektur dieser selbst vornahmen, gab. Unterhalb des Signifikanzniveaus verbesserte sich die Kontrollgruppe sogar stärker als die Experimentalgruppe A. Dies erstaunt mit Blick auf die Befundlage durchaus, da mehrere Studien existieren, die sogar formative Assessments ohne weitere Interventionen oder Feedback hohe Effekte zuschrieben (z. B. Jacobs, 1998). Bedacht werden muss in diesem Zusammenhang jedoch, dass die abschließend gemessenen Kompetenzstände zwar durch den Pretest den Kompetenzstand zu Beginn des Projekts berücksichtigen, nicht jedoch möglicherweise existierende methodische Vorerfahrungen bei den Schülerinnen und Schülern (Lipowsky, 2006) oder lernwirksame Effekte, die zwar bereits vor Projektbeginn angebahnt wurden, sich jedoch erst am Ende der Studie zeigten – und folglich in keinem kausalen Zusammenhang zur Intervention stehen (Händel & Ziegler, 2022). Erklärt werden könnte dieses Ergebnis jedoch auch damit, dass es für die Lernenden und deren potentiellen Lernerfolg entscheidend ist, ob die Lehrkraft für die Korrektur verantwortlich ist (externes Feedback) oder der Lernende selbst (internes Feedback) (Narciss, 2020; Riemer, 2022). Diese Erkenntnis geht konform mit bisherigen Befunden, bei welchen Fremdfedback in schulischen Lehr-Lernprozessen ebenfalls höhere Effekte zugesprochen werden (z. B. Wollenschläger et al., 2012). Möglicherweise gehen diese Ergebnisse auch auf den jewei-

ligen Stand selbstgesteuerten Lernens zurück: Schülerinnen und Schüler, die hier eine höhere Stufe erreicht haben können unter Umständen stärker von selbst auszuwertenden Tests profitieren als Lernende, deren Kompetenz in dieser Domäne erst weniger stark ausgebildet ist (vgl. Konrad & Traub, 2018).

Ziel der Studie war es, ein konkretes Konzept in einer bestimmten Domäne zu testen. Dies erfolgte mit Blick auf die bisher häufig geäußerte Kritik an eher allgemeinen Aussagen zu potentiellen Effekten und Konzeptionen (Bennet, 2011). Limitierend muss aber in diesem Zusammenhang angefügt werden, dass das erprobte Konzept zwar als erfolgsversprechend eingestuft werden darf, Verallgemeinerungen aber kaum vorgenommen werden dürfen.

Dies liegt nicht zuletzt daran, dass es sich vor allem im äußerst komplexen Untersuchungsfeld Schule als sehr herausfordernd darstellt, tatsächlich „echte“ experimentelle Designs umzusetzen (Böhm-Kasper & Weishaupt, 2008). So müssen auch im vorliegenden Falle alle bekannten Limitationen derartiger Designs bei der Interpretation der Ergebnisse bedacht werden (Bortz & Döring, 2015). Vor allem erscheint es in der Feldforschung nahezu unmöglich alle potentiell relevanten Ausgangsbedingungen und Störvariablen entsprechend zu berücksichtigen (Händel & Ziegler, 2022). Bei der Pilotstudie TÜV-D muss an dieser Stelle daher angeführt werden, dass keine zuordenbaren Daten zum Unterrichtsverlauf vor dem Beginn des Projekts oder zum Unterrichtsstil der Lehrkräfte vorliegen. Alle Lehrkräfte gaben bei der Projektbewerbung an, formativen Assessments einen hohen Nutzen zuzuschreiben (siehe 4.3) und diese regelmäßig in den Unterricht zu integrieren. Wie, in welcher Form oder wie häufig die Umsetzung jedoch tatsächlich stattfand, ist nicht bekannt. So ist es durchaus plausibel, dass aufgrund der vorherigen Unterrichtserfahrungen zwischen den Schülerinnen und Schülern deutlich Unterschiede im Umgang mit formativen Assessments sowie den zugehörigen Rückmeldungen existieren (Schmidt, 2020). Zudem kann selbstredend nur bedingt kontrolliert werden, wie offerierte individuelle Rückmeldungen tatsächlich von den Schülerinnen und Schülern genutzt wurden (Hagena et al., 2022). Die intensive Auseinandersetzung mit angebotenen Rückmeldungen stellt aber selbstredend eine essentielle Voraussetzung für deren Wirksamkeit dar (Narciss, 2014). Tatsächlich belegt ist im vorliegenden Projekt in diesem Kontext folglich lediglich, dass die Lehrenden die Ergebnisse im Unterricht besprochen haben und die konzipierten Unterrichtsstunden zur Förderung der Kompetenz umgesetzt wurden. Wie bereits skizziert können unterschiedliche schulische Vorerfahrungen, z. B. im Bereich der Unterrichtsgestaltung, Kompetenzentwicklungen bei den Lernenden stark beeinflussen (Lipowsky, 2006).

In diesem Zusammenhang muss weiter limitierend angeführt werden, dass die potentiell bedeutsame Vorerfahrung im Bereich formativen Assessments zwar nicht hinsichtlich möglicher Unterschiede statistisch erfasst wurde, aber bei nahezu allen Schülerinnen und Schülern zumindest vorhanden war. Dies ist bei der Analyse der Stichprobenszusammensetzung in jedem Falle zu berücksichtigen, da das Sample somit eine Positivauswahl darstellt: Alle beteiligten Lehrkräfte sowie nahezu alle Schülerinnen und Schüler hatten im Vorfeld des Projekts im Rahmen der Lehrveranstal-

tung „Formative Leistungsmessungen im Schulalltag“ bereits Kontakt mit Konzepten formativen Assessments. Die meisten Lernenden kannten somit sowohl den Ablauf eines Unterrichts, in welchen regelmäßig formative Assessments und entsprechende Förderphasen integriert sind, als auch den potentiellen Nutzen dieser Verfahren – eine üblicherweise eher ablehnende Haltung gegenüber zusätzlichen Tests war somit kaum gegeben (Bennett, 2011; Hagen et al., 2022).

Weiter muss hinsichtlich der Stichprobe selbstredend angeführt werden, dass diese eher klein ausfällt und ausschließlich aus Schülerinnen und Schülern einer Schule besteht. Diese Problembereiche stellen jedoch keine spezifischen Limitationen der vorliegenden Projektstudie dar (Brown, 1992), sondern müssen häufig als Einschränkungen berücksichtigt werden, wenn es um die „Nutzbarmachung von Ergebnissen aus der Interventionsforschung“ geht (Souvignier, 2022). Im Zuge der Auseinandersetzung mit realistischen Standards in der Interventionsforschung werden bestehende Problembereiche wie beispielsweise die Stichprobengröße oder das Konstanthalten von Untersuchungsbedingungen augenblicklich auch unter der Prämisse diskutiert, ob sich „Interventionsprogramme in der schulischen Praxis unter unterschiedlichen Bedingungen und in der Hand verschiedener Lehrkräfte bewähren müssen“ (Souvignier, 2022, S. 70). Mit Blick auf diese Perspektive muss somit erweiternd angeführt werden, dass die Pilotstudie TÜV-D nicht nur lediglich an einer Schule durchgeführt wurde, sondern auch keine unterschiedlichen Lehrkräfte zum Einsatz kamen – im vorliegenden Fall wurde versucht, die Untersuchungsbedingungen, soweit es im Falle einer Interventionsstudie in der Schulpraxis möglich ist, anzugleichen. Beide angerissene Untersuchungsbedingungen liefern somit bei der Interpretation zu berücksichtigende Limitationen. Für das vorliegende Projekt gilt es daher zu prüfen, ob das entwickelte Förderprogramm TÜV-D auf Grundlage von formativen Assessments auch an anderen Schulen und vor allem auch von anderen Lehrkräften mit ähnlichen Effekten durchgeführt werden könnte.

Im Zuge dieser Überlegungen hinsichtlich eines Einsatzes in der Unterrichtspraxis muss weiter ergänzt werden, dass die Evaluation des Projektes TÜV zwar die Potentiale eines derartigen Förderkonzeptes aufzeigt, Fragen einer Operationalisierung im täglichen Schulalltag jedoch nicht bzw. noch nicht Gegenstand der Erprobung waren. Unter Berücksichtigung aktueller Befunde der Implementationsforschung (z. B. Schrader et al., 2020) ist darauf zu verweisen, dass neben bekannten, übertragbaren Gelingensbedingungen für erfolgreiche Implementationen auch und vor allem kontext- sowie organisationsspezifische Ansätze besonders bedeutsam für die Etablierung innovativer Unterrichtskonzepte, -formen und -methoden sind (Holtappels, 2019). Betrachtet man unter dieser Prämisse Kernelemente des vorgestellten Projektes, so müsste beispielsweise unter anderem thematisiert werden, wie dauerhaft ein möglicher zeitlicher Mehraufwand durch regelmäßige Erhebungs- und Auswertungszeitpunkte kompensiert werden könnte (Hagen et al., 2022) oder wie Lernende über eine längere Periode hinweg für derartige Förderkonzepte motiviert werden können. Lehrkräfte betonen in diesem Kontext zumeist, dass sie bei der Operationalisierung von innovativen Unterrichtskonzepten zum einen in die diversen Planungs- und Umsetzungsphasen

integriert sein wollen (Gräsel, 2010); zum anderen die notwendigen strukturellen und organisatorischen Rahmenbedingungen für eine Implementation vorfinden möchten (Janssen et al., 2013). Auch die Herausforderung, geeignete Instrumente zu entwickeln, die – wenn möglich – höheren Ansprüchen gerecht werden, stellt eine nicht zu unterschätzende Hürde dar (Börnert, 2014; Blumenthal et al., 2022). Allerdings wird diesen Komplikationen mittlerweile durchaus Rechnung getragen und es existieren bereits Erfahrungsberichte bzw. Befunde, die implementationsrelevante Attribute formativen Assessments benennen (z. B. Schütze et al., 2018). Einer der bedeutsamsten Faktoren in diesem Zusammenhang ist, dass Lehrkräfte vom Nutzen formativen Assessments überzeugt sind und daher mögliche Hürden und zusätzlichen Arbeitsaufwand mit Blick auf potentielle Kompetenzsteigerungen bei den Schülerinnen und Schülern akzeptieren (z. B. Heitink et al., 2016). Durch evaluierte Unterrichtskonzepte oder Lernarrangements mit integrierten formativen Assessments konnte zudem mittlerweile nicht nur nachgewiesen werden, dass solche Konzepte tatsächlich in den Schulalltag eingebunden werden können, sondern auch, welche Vorgehensweisen für eine erfolgreiche Implementation bedeutsam werden (z. B. Schmidinger, 2013; Hagena et al., 2022).

Literatur

- Balt, M., Ehlert, A. & Fritz, A. (2017a). Theoriegeleitete Testkonstruktion dargestellt am Beispiel einer Lernverlaufsdagnostik für den mathematischen Anfangsunterricht. *Empirische Sonderpädagogik*, 9(2), 165–183. <https://dx.doi.org/10.25656/01:15012>
- Balt, M., Ehlert, A. & Fritz, A. (2017b). Lernverlaufsdagnostik im mathematischen Anfangsunterricht. In U. Kortenkamp & A. Kuzle (Hrsg.), *Beiträge zum Mathematikunterricht* (S. 1155–1158). WTM-Verlag.
- Baumert, J., Gruehn, S., Heyn, S., Köller, O. & Schnabel, K—U. (1997). *Bildungsverläufe und Psychosoziale Entwicklungen im Jugendalter (BIJU). Dokumentation – Band 1. Skalen Längsschnitt 1. Wellen 1-4*. Max-Planck-Institut für Bildungsforschung. <https://dx.doi.org/10.7477/13:284:1>
- Bennet, R. E. (2011). Formative Assessment: A critical review. *Assessment in Education*, 18(1), 5–25. <https://dx.doi.org/10.1080/0969594X.2010.513678>
- Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–21. <https://dx.doi.org/10.1007/s11092-008-9068-5>
- Blumenthal, S. (2022). Lernverlaufsdagnostik. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 633–648). Universitätsbibliothek. <https://dx.doi.org/10.5283/epub.53149>
- Blumenthal, S., Gebhardt, M., Förster, N. & Souvinier, E. (2022). Internetplattformen zur Diagnostik von Lernverläufen von Schülerinnen und Schülern in Deutschland – Ein Vergleich der Plattformen Lernlinie, Levumi und quop. *Zeitschrift für Heilpädagogik*, 73(4),

- 153–167. <https://dx.doi.org/10.5283/epub.52069>
- Böhm-Kasper, O. & Weishaupt, H. (2008). Quantitative Ansätze und Methoden in der Schulforschung. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 91–123). Springer VS. https://dx.doi.org/10.1007/978-3-531-91095-6_4
- Börnert, M. (2014). Lernverlaufsdiagnostik. Definition, Einsatzbereiche und Perspektiven für die pädagogische Praxis. *Potsdamer Zentrum für empirische Inklusionsforschung (ZEIF)*, 2, 1–8.
- Bortz, J. & Döring, N. (2015). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (5. Aufl.). Springer Medizin. <https://dx.doi.org/10.1007/978-3-642-41089-5>
- Breucker, T. & Kuhl, J. (2022). Formative Diagnostik in der deutschen Sonderpädagogik. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 619–632). Universitätsbibliothek. <https://dx.doi.org/10.5283/epub.53149>
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), 141–178. https://dx.doi.org/10.1207/s15327809jls0202_2
- Bürgermeister, A. & Saalbach, H. (2018). Theoretischer Beitrag: Formatives Assessment: Ein Ansatz zur Förderung individueller Lernprozesse. *Psychologie in Erziehung und Unterricht (PEU)*, 65(3), 194–205. <https://dx.doi.org/10.2378/peu2018.art11d>
- Deci, E. L., Koestner, R. & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 12(6), 627–668. <https://dx.doi.org/10.1037/0033-2909.125.6.627>
- Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, 37(1), 184–192. <https://dx.doi.org/10.1177/00224669030370030801>
- Dunn, K. E. & Mulvenon, S. W. (2009). A critical review of research on formative assessment: the limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14(7), 1–11. <https://dx.doi.org/10.7275/jg4h-rb87>
- Fischer, C. (2022). Diagnosebasierte individuelle Förderung im potenzialorientierten Umgang mit Diversität. In C. Fischer & D. Rott (Hrsg.), *Individuelle Förderung – Heterogenität und Handlungsperspektiven in der Schule* (= utb, Nr. 5919) (S. 27–40). Waxmann. <http://dx.doi.org/10.36198/9783838559193>
- Fischer, C. (2019). *Individuelle Förderung als schulische Herausforderung* (= Schriftenreihe des Netzwerk Bildung). Friedrich-Ebert-Stiftung.
- Fischer, U., Rösch, S., Nuerk, H.-C. & Moeller, K. (2015). Erkennen von Rechenschwäche durch LehrerInnen und Testungen im Klassenverband. *Lernen und Lernstörungen*, 4(4), 269–285. <https://dx.doi.org/10.1024/2235-0977/a000116>
- Förster, N., Kuhn, J.-T. & Souvignier, E. (2017). Normierung von Verfahren zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik*, 9(2), 116–122. <https://dx.doi.org/10.25656/01:14998>
- Gebhardt, M., Jungjohann, J. & Schurig, M. (2021). *Lernverlaufsdiagnostik im förderorientierten Unterricht. Testkonstruktion, Instrumente, Praxis*. Ernst Reinhardt.
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 13(1), 7–20. <https://dx.doi.org/10.1007/s11618-010-0109-8>
- Händel, M. & Ziegler, A. (2022). Bildungspsychologische Forschung. In S. Spiel, T. Götz, P. Wagner, M. Lüftenegger & B. Schober (Hrsg.), *Bildungspsychologie. Ein Lehrbuch* (2. Aufl.) (S. 237–260). Hogrefe.

- Hagen, M., Besser, M. & Blum, W. (2022). Individuelle Diagnose und lernprozessbegleitende Rückmeldung im Mathematikunterricht der Grundschule. In K. Eilerts, R. Möller & T. Huhmann (Hrsg.), *Auf dem Weg zum neuen Mathematiklehren und -lernen 2.0* (S. 2–748), Springer Spektrum. https://dx.doi.org/10.1007/978-3-658-33450-5_3
- Harris, L. R., Brown, G. T. & Harnett, J. A. (2014). Understanding Classroom Feedback Practices: A Study of New Zealand Student Experiences, Perceptions, and Emotional Responses. *Educational Assessment, Evaluation and Accountability*, 26(2), 107–133. <https://dx.doi.org/10.1007/s11092-013-9187-5>
- Hasselhorn, M., Schneider, W. & Trautwein, U. (Hrsg.) (2014). *Lernverlaufsdiagnostik* (= Tests und Trends, Jahrbuch der pädagogisch-psychologischen Diagnostik, Bd. 12). Hogrefe.
- Hattie, J. A. (2012). *Visible learning for teachers. Maximizing impact on learning*. Routledge. <https://dx.doi.org/10.4324/9780203181522>
- Hattie, J. A. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://dx.doi.org/10.3102/003465430298487>
- Hebbecke, K. & Souvignier, E. (2018). Formatives Assessment im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts. *Zeitschrift für Erziehungswissenschaft*, 21(4), 735–765. <https://dx.doi.org/10.1007/s11618-018-0834-y>
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K. & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50–62. <https://dx.doi.org/10.1016/j.edurev.2015.12.002>
- Helmke, A. & Klieme, E. (2008). Unterricht und Entwicklung sprachlicher Kompetenzen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 301–312). Beltz. <https://dx.doi.org/10.25656/01:3149>
- Hofmann, F., Rau, C. & Zeitler, S. (2019). *Formative Diagnosekompetenz entwickeln – eine Herausforderung für Lehramtsstudierende* (= SUN, Schulpädagogische Untersuchungen Nürnberg, Forschungsbericht Nr. 41). Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Schulpädagogik mit dem Schwerpunkt empirische Unterrichtsforschung.
- Hofmann, F. (2017). *Selbsteinschätzungen im Aufsatz- und Schreibunterricht. Eine empirische Untersuchung zu den Effekten von formativen Leistungsdiagnosen und Selbsteinschätzungen auf die Schreibkompetenz*. Klinkhardt.
- Holtappels, H. G. (2019). Transfer in der Schulentwicklung. Ansätze und Gelingensbedingungen aus der Perspektive von Schulentwicklungstheorie und -forschung. *Die Deutsche Schule*, 111(3), 274–293. <https://dx.doi.org/10.31244/ddS.2019.03.03>
- Hußmann, A. & Schröter, A. (2022). Förderplanung im Bereich Lesen. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 667–682). Universitätsbibliothek. <https://dx.doi.org/10.5283/pub.53149>
- IQB = Institut für Qualitätsentwicklung im Bildungswesen. (2021). *Wie entsteht ein VERA-Test?* Online unter: <https://www.iqb.hu-berlin.de/vera/ueberblick/testentwicklung> (letzter Abruf: 28.12.2022).
- IQB = Institut für Qualitätsentwicklung im Bildungswesen. (o. J.). *Beispielaufgaben und Materialien. Deutsch Sekundarstufe I: Sprachgebrauch*. Online unter: <https://www.iqb.hu-berlin.de/vera/-aufgaben/de1/> (letzter Abruf: 15.09.2022).
- Jacobs, B. (1998). *Aufgaben stellen und Feedback geben*. Saarbrücken: Medienzentrum der Philosophischen Fakultät der Universität Saarbrücken. Online unter: <https://psycharchives.de/jspui/bitstream/20.500.11780/1024/1/feedback.pdf> (letzter Abruf: 16.09.2022).

- Janssen, F., Westbroek, H., Doyle, W. & van Driel, J. (2013). How to make innovations practical. *Teachers College Record*, 115(7), 1–43. <https://doi.org/10.1177/01614681131150070>
- Jónsson, I. R., Smith, K & Geirsdóttir, G. (2018). Shared language of feedback and assessment. Perception of teachers and students in three Icelandic secondary schools. *Studies in Educational Evaluation*, 56, 52–58. <https://dx.doi.org/10.1016/j.stueduc.2017.11.003>
- Jürgens, E. (2022). Leistungsbeurteilung. In M. Harring, C. Rohlf's & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (2. akt. u. erw. Aufl.) (S. 552–562). Waxmann. <https://dx.doi.org/10.36198/9783838587967>
- Käter, C., Käter, T., Martenstein, R. & Hillenbrand, C. (2016). Leitfadengestützte Konstruktion eines Instruments der Lernverlaufdiagnostik (CBM) im Bereich Lesen. *Zeitschrift für Heilpädagogik*, 66(4), 168–179.
- Kingston, N. & Nash, B. (2011). Formative Assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Klauer, K. J. (2011). Lernverlaufdiagnostik – Konzept, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik*, 3(3), 207–224. <https://dx.doi.org/10.25656/01:9324>
- Klauer, K. J. (2014). Formative Leistungsdiagnostik. Historischer Hintergrund und Weiterentwicklung zur Lernverlaufdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufdiagnostik* (= Jahrbuch der pädagogisch-psychologischen Diagnostik, Bd. 12) (S. 1–18). Hogrefe.
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D. & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht. Projekt Co2CA. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (Zeitschrift für Pädagogik, Beiheft 56) (S. 64–74). Beltz. <https://dx.doi.org/10.25656/01:3324>
- Kluger, A. N. & DeNisi, A. S. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119(2), 254–284. <https://dx.doi.org/10.1037/0033-2909.119.2.254>
- KMK = Kultusministerkonferenz. (2004). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss. Beschluss der Kultusministerkonferenz vom 4.12.2003*. Wolters Kluwer.
- KMK = Kultusministerkonferenz. (2014). *Kompetenzstufenmodell zu den Bildungsstandards im Kompetenzbereich Sprache und Sprachgebrauch untersuchen für den Mittleren Bildungsabschluss*. Online unter: [file:///C:/Users/eq53ohed/Downloads/KSM_Sprachgebrauch_MSA_2014-12-10c%20\(2\).pdf](file:///C:/Users/eq53ohed/Downloads/KSM_Sprachgebrauch_MSA_2014-12-10c%20(2).pdf) (letzter Abruf: 28.12.2022).
- KMK = Kultusministerkonferenz. (2015). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Online unter: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf (letzter Abruf: 28.12.2022).
- Konrad, K. & Traub, S. (2018). *Selbstgesteuertes Lernen. Grundwissen und Tipps für die Praxis* (6. Aufl.). Schneider Verlag Hohengehren.
- Kopmann, H. (2021). Zwischen Anspruch und Alltäglichkeit: Konzeptuelle unterrichtspraktische Perspektiven auf inklusive Diagnostik. *Qfl – Qualifizierung für Inklusion*, 3(2). <https://dx.doi.org/10.21248/qfl.75>
- Kunter, M., Schümer, G., Artelt, A., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Max-Planck-Institut für Bildungsforschung.
- Lerche, T. (2019). Üben. In E. Kiel, B. Herzig, U. Maier & U. Sandfuchs (Hrsg.), *Handbuch Unterrichten an allgemeinbildenden Schulen* (= utb 5308) (S. 220–227). Klinkhardt.

- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (= Zeitschrift für Pädagogik, Beiheft Nr. 51) (S. 47–70). Beltz. <https://dx.doi.org/10.25656/01:7370>
- Ludin, F. (2021). *Diversität an Schulen – Etablierung von Diversitäts- und Antidiskriminierungskonzepten im Kontext der Handlungsstrategien und Handlungsmöglichkeiten von Schulleiter*innen. Eine qualitative Studie zur Umsetzung des Landesantidiskriminierungsgesetzes (LADG)* (= Masterarbeit, Studiengang Schul- und Bildungsmanagement). Universität Potsdam. Online unter: https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/56322/file/ludin_master.pdf (letzter Abruf: 12.12.2022).
- Maier, U. (2010). Formative Assessment – Ein erfolgsversprechendes Konzept zur Reform von Unterricht und Leistungsmessung. *Zeitschrift für Erziehungswissenschaft*, 13(2), 293–308. <https://dx.doi.org/10.1007/s11618-010-0124-9>
- Maier, U. (2015). *Leistungsdiagnostik in Schule und Unterricht* (= Studentexte Bildungswissenschaft, utb 4178). Klinkhardt. <https://dx.doi.org/10.36198/9783838541785>
- Maier, U., Hofmann, F. & Zeitler, S. (2012). *Formative Leistungsdiagnostik. Grundlagen und Praxisbeispiele* (= Schulmanagement Handbuch, Bd. 141). Oldenbourg.
- McDonald, B. & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, 10(2), 209–220. <https://dx.doi.org/10.1080/0969594032000121289>
- Meyer, H. (2014). *Leitfaden Unterrichtsvorbereitung* (10. Aufl.). Cornelsen Scriptor.
- Mory, E. (2004). Feedback Research Revisited. In D. Jonassen (Hrsg.), *Handbook of research for educational communications and technology* (S. 745–783). Lawrence Erlbaum.
- Moser Opitz, E. (2022). Diagnostisches und didaktisches Handeln verbinden: Entwicklung eines Prozessmodells auf der Grundlage von Erkenntnissen aus der pädagogischen Diagnostik und der Förderdiagnostik. *Journal für Mathematik-Didaktik*, 43(1), 205–230. <https://dx.doi.org/10.1007/s13138-022-00201-1>
- Narciss, S. (2006). *Informatives tutorielles Feedback*. Waxmann.
- Narciss, S. (2014). Modelle zu den Bedingungen und Wirkungen von Feedback in Lehr-Lernsituationen. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 43–82). Waxmann.
- Narciss, S. (2020). Feedbackstrategien für interaktive Lernaufgaben. In H. Niegemann & A. Weinberger (Hrsg.), *Handbuch Bildungstechnologie. Konzeption und Einsatz digitaler Lernumgebungen* (S. 369–392). Springer. https://dx.doi.org/10.1007/978-3-662-54368-9_35
- OECD. (2019). *PISA 2018. Results. Combined executive summaries. Volume I, II, III*. Paris: OECD. Online unter: https://www.oecd.org/pisa/publications/PISA2018_CN_DEU_German.pdf (letzter Abruf: 24.09.2022).
- Pawlik, K. (1976). Modell- und Praxisdimensionen psychologischer Diagnostik. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik* (S. 13–43). Klett.
- Pissarek, M. & Wild, J. (2018). Prä-/Post-/Follow-Up-Kontrollgruppendesign. In J. M. Boelmann (Hrsg.), *Empirische Forschung in der Deutschdidaktik. Band 1: Grundlagen* (S. 215–236). Schneider Verlag Hohengehren.
- Rakoczy, K., Klieme, E., Leiß, D. & Blum W. H.-J. (2017). Formative Assessment in Mathematics Instruction: Theoretical Considerations and Empirical Results of the Co2CA Project. In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Hrsg.), *Competence Assessment in Ed-*

- ucation. *Research, Models and Instruments* (= Methodology of Educational Measurement and Assessment) (S. 469–485). Springer. <https://dx.doi.org/10.1007/978-3-319-50030-0>
- Riemer, C. (2022). Intentional statt inzidentell. Zur Effektivität mündlichen korrekativen Feedbacks in der L2-Unterrichtsinteraktion. In E. Burwitz-Melzer, C. Riemer & L. Schmelter (Hrsg.), *Feedback beim Lehren und Lernen von Fremd- und Zweitsprachen* (= Giessener Beiträge zur Fremdsprachendidaktik) (S. 113–124). Narr.
- Rosidin, U., Haryanti, N., Lora, H. A. & Viyanti, V. (2020). Reconstruct the Class Assessment Strategy: Promoting the 21st-Century Learning. *Indonesian Journal of Science and Mathematics Education*, 3(1), 2–36. <https://dx.doi.org/10.24042/ijmsme.v3i1.6056>
- Ruiz-Primo, M. A. & Furtak, E. M. (2007). Exploring Teachers' Informal Formative Assessment Practices and Students' Understanding in the Context of Scientific Inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84. <https://dx.doi.org/10.1002/tea.20163>
- Scriven, M. (1967). *The methodology of evaluation*. American Educational Research Association.
- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B. & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103(1), 101602-101618. <https://dx.doi.org/10.1016/j.ijer.2020.101602>
- Schmidinger, E. (2013). Formative Leistungsbeurteilung. *Erziehung und Unterricht*, 163(9/10), 776–785. <https://dx.doi.org/10.17888/nbb2015-2>
- Schmidt, C. (2020). *Formatives Assessment in der Grundschule. Konzepte, Einschätzungen der Lehrkräfte und Zusammenhänge*. VS Springer. <https://dx.doi.org/10.1007/978-3-658-26921-0>
- Schrader, F.-W. & Helmke, A. (2014). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (3. Aufl.) (S. 45–58). Beltz.
- Schrader, S., Hasselhorn, M., Hetfleisch, P. & Goeze, A. (2020). Stichwortbeitrag Implementationsforschung: Wie Wissenschaft zu Verbesserungen im Bildungssystem beitragen kann. *Zeitschrift für Erziehungswissenschaft*, 23(1), 9–59. <https://dx.doi.org/10.1007/s11618-020-00927-z>
- Schroeder, J. (2016). Diversity Management im Schulsystem. In P. Genkova & T. Ringeisen (Hrsg.), *Handbuch Diversity Kompetenz. Band 1: Perspektiven und Anwendungsfelder* (= Springer Reference Psychologie) (S. 785–801). Springer. https://dx.doi.org/10.1007/978-3-658-08594-0_57
- Schütze, B., Souvignier, E. & Hasselhorn, M. (2018). Stichwort – Formatives Assessment. *Zeitschrift für Erziehungswissenschaft*, 21(4), 697–715. <https://dx.doi.org/10.1007/s11618-018-0838-7>
- Souvignier, E., Förster, N. & Kawohl, E. (2016). Implementation eines Förderkonzeptes zur diagnosebasierten individuellen Leseförderung in der Grundschule. In M. Philipp & E. Souvignier (Hrsg.), *Implementation von Lesefördermaßnahmen. Perspektiven auf Gelingensbedingungen und Hindernisse* (S. 77–98). Waxmann.
- Souvignier, E. (2022). Nutzbarmachung von Ergebnissen aus der Interventionsforschung für Wissenschaft und Praxis. In N. McElvany, M. Becker, F. Lauermaun, H. Gaspard & A. Ohle-Peters (Hrsg.), *Optimierung schulischer Bildungsprozesse – What works?* (= Dortmunder Symposium der Empirischen Bildungsforschung, Bd. 6) (S. 63–80). Waxmann.
- Souvignier, E. & Hasselhorn, M. (2018). Formatives Assessment. *Zeitschrift für Erziehungswissenschaft*, 21(4), 693–696. <https://dx.doi.org/10.1007/s11618-018-0839-6>
- Sprenger, J. (2019). Unser Pressure. Oder: Berufsvollzugsprobleme und Belastungen von Lehrpersonen. Eine empirische Studie. *R & amp;E-Source, Online Journal for Research*

- and Education, 12. Online unter: <https://journal.ph-noe.ac.at/index.php/resource/article/view/711/730> (letzter Abruf: 10.12.2022).
- Streber, D. (2016). Individuelle Förderung – theoretisch begründet. In K. Zierer (Hrsg.), *Jahrbuch für Allgemeine Didaktik 2016* (S. 184–203). Schneider Verlag Hohengehren.
- Stanat, P. (2022). *Ein halbes Schuljahr Rückstand. Blog Bildung und Politik v. 1.7.2022*. Online unter: <https://www.jmwiarda.de/2022/07/01/ein-halbes-schuljahr-r%C3%BCckstand> (letzter Abruf: 17.09.2022).
- Tröster, H. (2018). *Diagnostik in schulischen Handlungsfeldern. Methoden, Konzepte, praktische Ansätze*. Kohlhammer.
- Urton, K., Hertel, S. & Hennemann, T. (2021). Umsetzung und Evaluation eines multiprofessionellen und systemübergreifenden Inklusionskonzeptes für Kinder und Jugendliche im Schulalter. *ESE, Emotionale und Soziale Entwicklung in der Pädagogik der Erziehungshilfe und bei Verhaltensstörungen*, 3(3), 164–172. <https://dx.doi.org/10.35468/5903-13>
- Van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education*, 85, 175–189. <https://dx.doi.org/10.1016/j.tate.2019.06.010>
- Wagner, W., Helmke, A. & Rösner, E. (2009). *Deutsch Englisch Schülerleistungen international. Dokumentation der Erhebungsinstrumente für Schülerinnen und Schüler, Eltern und Lehrkräfte* (= Materialien zur Bildungsforschung, Bd. 25/1). GFPPF/DIPF.
- Wilde, M., Bätz, K., Kovaleva, A. & Urhahne, D. (2009). Überprüfung einer Kurzska-la intrinsischer Motivation (KIM). *Zeitschrift für Didaktik der Naturwissenschaften*, 15(1), 31–45.
- Winter, F. (2015). *Lerndialog statt Noten. Neue Formen der Leistungsbeurteilung*. Beltz.
- Wollenschläger, M., Möller, J. & Harms, U. (2012). Ist kompetenzielles Fremdfeedback überlegen, weil es als effektiver wahrgenommen wird? *Unterrichtswissenschaft*, 40(3), 197–212.
- Zierer, K., Busse, V., Wernke, S. & Otterspeer, L. (2015). Feedback in der Schule – Forschungsergebnisse. In C. G. Buhren (Hrsg.), *Handbuch Feedback in der Schule* (S. 31–50). Beltz.

Formatives Feedback ist nicht immer hilfreich: Feedbackrezeption und Lernzuwachs in einer Web-App zur Übung von Grammatik und Rechtschreibung

Uwe Maier | Christian Klotz

Pädagogische Hochschule Schwäbisch Gmünd

Abstract: In digitalen Lernsystemen wird formatives Feedback genutzt, um die Lernumgebung adaptiv an die Lernfortschritte anzupassen. Ein formatives Feedback ist jedoch nur dann lernwirksam, wenn die Lernenden diese Rückmeldung auch verstehen und für die Schließung von Wissenslücken nutzen. In dieser Studie wird auf Ebene einzelner Fehlerrückmeldungen untersucht, wie sich die Feedbacklesezeit zusammen mit der subjektiv eingeschätzten Fehlerverständlichkeit auf den Lernzuwachs auswirken. Forschungsgegenstand ist die Web-App MasteryX (www.masteryx.de), die verschiedene Minikurse zur Übung und Wiederholung von Grammatik- und Rechtschreibthemen anbietet. Die Lernprogression wird über formative Assessments gesteuert. Um herauszufinden, wie Feedbackrezeption und Lernzuwachs zusammenhängen, wurden die Log-Daten von 528 Personen (288 Schülerinnen und 240 Schüler) in 63 Klassen analysiert. Es zeigte sich ein positiver Zusammenhang zwischen der Anzahl der verstandenen Fehleritems und dem Lernzuwachs beim Thema Wortarten. Bei anderen Themen (Kommasetzung, Feldermodell) sind diese Zusammenhänge nicht vorhanden. Die Befunde legen nahe, dass die Lernwirksamkeit von formativem Feedback stark von der inhaltlichen Umsetzung abhängt.

1 Einleitung

Formatives Assessment (oder alternativ: *assessment for learning*) ist ein international diskutiertes, didaktisches Konzept, bei dem Leistungsmessungen in den laufenden Lehr-Lernprozess integriert werden, um daraus Schlussfolgerungen für die Optimierung des Unterrichts ziehen zu können (z. B. Maier, 2010; Souvignier et al., 2018). Die Beispiele hierfür reichen von Fragetechniken über Quizspiele bis hin zu systematisch angelegten Lernverlaufsdiagnosen. Mit der rasanten Entwicklung von digitalen Bildungstechnologien ergaben sich auch immer mehr Möglichkeiten, formative Assessments zeitsparend und adaptiv in den Unterrichtsprozess zu integrieren. In vielen digitalen Lernumgebungen (z. B. intelligente tutorielle Systeme, adaptive Übungssysteme, gamifizierte Lern-Apps, Lernmanagementsysteme) wird formatives Feedback in Form von kurzen Lernzielüberprüfungen eingebaut. Die Lernenden erhalten dabei ein sofortiges, schriftliches, computergeneriertes Feedback auf unterschiedlichen Ebenen (formatives Assessment bestanden, prozentualer Anteil korrekter Aufgaben, Rückmel-

dungen zu einzelnen Aufgaben), an dem sich die weiterführenden Lernwege in der digitalen Lernumgebung orientieren. Ein solches formatives Feedback ist zentraler Teil von formativem Assessment und relevant, weil die Lernenden die erhaltenen Informationen unmittelbar zur Optimierung ihres Lernens nutzen können. Die Forschungen zu formativem Assessment (Wiliam, 2011) und spezifisch zu formativem Feedback (Shute, 2008) legen die Schlussfolgerung nahe, dass die Lernwirksamkeit von formativen Assessments in digitalen Lernumgebungen unter anderem von der Feedbackrezeption abhängt. Die Rezeption von digitalem, schriftlichem Feedback wurde in den bisherigen Studien allerdings überwiegend über Lesezeiten und Clicks auf Feedbackseiten erfasst und dabei vernachlässigt, inwieweit die Lernenden das Feedback verstehen. Es gibt Hinweise, dass Lernende mit geringem Vorwissen und ungünstigen motivationalen Voraussetzungen längere Feedbacklesezeiten haben (Fox, 2013; Timmers et al., 2013). Unklar ist allerdings, ob längere Feedbacklesezeiten mit Verständnis einhergehen und diese Lernenden davon profitieren. In der hier vorgestellten Studie wurde die Rezeption des formativen Feedbacks deshalb wesentlich feinkörniger auf der Ebene einzelner Aufgaben operationalisiert. Am Beispiel einer Lern-App wurde untersucht, wie sich die Rezeption und dabei vor allem das Verständnis von Rückmeldungen zu Fehlern bei einzelnen Aufgaben auf den Lernzuwachs bei einem wiederholt durchgeführten, formativen Assessment auswirkt.

2 Theoretischer und empirischer Hintergrund der Studie

Formatives Assessment (alternativ: formative Leistungsmessung, formative Leistungsdiagnostik; engl.: *formative assessment*, *assessment for learning*) gilt in der internationalen Literatur als ein didaktisches Unterrichtsprinzip, das sich von einer an der Notengebung orientierten Kultur der summativen Leistungsdiagnostik abgrenzt (Köller, 2005; Black & Wiliam, 2009). Mündliche oder schriftliche Leistungsmessungen sollen nicht ausschließlich der Zensurengebung dienen, sondern sowohl für die Lehrkräfte als auch die Lernenden eine informative Rückmeldung für den weiterführenden Unterricht oder die individuelle Förderung bereitstellen. Hierzu eignen sich neue Formate der Leistungsmessung, die sich von typischen Klausuren oder Tests abgrenzen (z. B. spielerische Quiz-Formate). Einige Autoren betonen allerdings, dass formative und summative Leistungsmessungen nicht generell im Widerspruch stehen müssen (Black & Wiliam, 2009). Vielmehr können auch summative, für die Zensurengebung durchgeführte Leistungstests formativ genutzt werden, beispielsweise um Lernschwierigkeiten noch einmal aufzugreifen.

Was man genau unter einer Praxis des formativen Assessment versteht und wie diese zu einer Optimierung des nachfolgenden Lehr-Lernprozesses beitragen kann, wird in der Literatur sehr unterschiedlich verstanden. Viele Metaanalysen und Übersichtsartikel zur Thematik schreiben generell einer formativen Leistungsmessung hohe Lerneffekte zu (Fuchs & Fuchs, 1986; Black & Wiliam, 1998; Lee et al., 2020). Es gibt

jedoch auch Autorinnen und Autoren, die einen differenzierteren Blick auf die unterschiedlichsten Praktiken und Methoden formativer Leistungsmessung einfordern (z. B. Bennett, 2011; Kingston & Nash, 2011). Neben Fach, Inhaltsbereich und Altersstufe hängt die Lerneffektivität von formativen Leistungsmessung von vielen weiteren Faktoren ab, z. B. der Testgüte der Leistungsmessung, dem Feedback sowie der Fähigkeit bzw. Motivation von Lehrenden und Lernenden, aus dem Feedback Konsequenzen abzuleiten (Heitink et al., 2016).

Das didaktische Prinzip des formativen Assessment besteht aus verschiedenen Aspekten bzw. Teilschritten (Black & Wiliam, 2009), z. B. Definition und Absprache von Lernzielen sowie Bewertungskriterien, Einbindung von einzelnen, formativen Leistungstests in den Unterrichtsprozess, Rezeption und Nutzung des formativen Feedbacks. Je nach Fach, Inhalt und Implementationskonzept können auch Leistungsmessungen über einen längeren Zeitraum hinweg wiederholt werden, um eine Lernfortschrittsdiagnostik zu ermöglichen (z. B. Klauer & Strathmann, 2013). In vielen digitalen Lernumgebungen werden formative Leistungstests nach dem Prinzip des *mastery learning* eingesetzt (z. B. Bokhove & Drijvers, 2012; Paiva et al., 2017; Tempelaar, 2020). Dabei entsteht nicht eine Lernverlaufsdiagnostik, vielmehr wird mit kleinen, lernzielorientierten Tests geprüft, ob bestimmte Lerninhalte oder Kompetenzen beherrscht werden. Die Rückmeldungen werden in diesem Fall kriterial-dichotom interpretiert: Beherrschen die Lernenden eine Kompetenz (ja/nein)? Kann auf dem nächsten Level weitergelernt werden (ja/nein)?

Das allgemeine, didaktische Prinzip des formativen Assessment ist eng mit dem lernpsychologischen Konzept „Feedback“ assoziiert. Jedes einzelne formative Assessment generiert eine Rückmeldung an Lehrende und Lernende. Shute (2008, 154) definiert formatives Feedback als: „... information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning“. Die Lerneffektivität des formativen Assessment wird von vielen Autorinnen und Autoren vor allem mit den Effekten des formativen Feedbacks erklärt (z. B. Hattie & Timperley, 2007). Hinzu kommt, dass Metaanalysen zeigten, dass vor allem schriftliche, computergestützte Rückmeldungen lernwirksamer sind als mündliche, von den Lehrkräften erteilte Rückmeldungen, weil diese als weniger selbstwertbedrohlich empfunden werden (Kluger & DeNisi, 1996). Zudem fanden Van der Kleij et al. (2015) in einer Metaanalyse hohe Effektstärken von itembasiertem Feedback in digitalen Lernumgebungen. Vor allem elaboriertes Feedback, das eine Rückmeldung zur Richtigkeit einer Aufgabe mit Erklärungen und Lernhinweisen kombiniert, gilt als besonders vielversprechend.

Ob formatives, schriftliches Feedback das Lernen von Schülerinnen und Schülern positiv beeinflusst oder nicht, hängt nach Wiliam (2011) von zwei entscheidenden Bedingungen ab. Erstens müssen die durch ein formatives Assessment erzeugten Diagnoseinformationen nicht nur eine Lücke zwischen aktuellem Lernstand und Lernziel aufzeigen, sondern es müssen Hinweise auf unterrichtliche Maßnahmen zur Schließung dieser Lücke bereitgestellt werden. Zweitens ist notwendig, dass die Lernenden aktiviert werden, das formative Feedback zu lesen und sich mit den darin enthaltenen Informationen auseinanderzusetzen. Diese Problematik der Feedbackrezeption gilt in besonde-

rem Maße für digitale Lernumgebungen, in denen Schülerinnen und Schüler wesentlich selbständiger agieren müssen und die persönliche Ansprache und Unterstützung durch eine Lehrkraft häufig fehlt. Empirische Befunde zeigen, dass die Feedbackrezeption bei digitalen formativen Assessments stark variieren kann. In der Studie von Timmers und Veldkamp (2011) beispielsweise wurde den Studierenden in einem Online-Kurs elaboriertes, schriftliches Feedback zu schriftlich eingereichten Aufgabenstellungen angeboten. Ein Viertel der Studierenden öffnete keine einzige dieser Feedbackseiten und ein Viertel öffnete alle Feedbackseiten. Young et al. (2021) erstellten mithilfe einer Software ein automatisiertes, individualisiertes Feedback nach formativen Assessments in einem Online-Kurs für Erstsemester im Chemiestudium. Eine Befragung der Studierenden ergab, dass lediglich 52 % der Studierenden das Feedback anschauten.

Weiter zeigte sich, dass die Wahrnehmung von Feedback zu falsch gelösten Aufgaben (negatives Feedback) mit Lernerfolg korreliert (Cutumisu et al., 2017; Zimbardi et al., 2017; Maier, 2020). In der Studie von Cutumisu et al. (2017) erhielten Schülerinnen und Schüler ein automatisiertes, formatives Feedback zu einer Designaufgabe (Poster für eine Schulveranstaltung gestalten). Das Feedback bestand aus drei positiven und drei negativen (kritischen) Aspekten. Die Lernenden konnten selbst entscheiden, welche Aspekte des Feedbacks sie anklicken und lesen möchten. Ebenso konnten sie entscheiden, ob sie das Poster verbessern möchten oder nicht. Die Auswertungen zeigten, dass die Anzahl der angeklickten, negativen Aspekte (sog. „negative feedback seeking choices“) mit der Anzahl der Revisionen sowie mit der finalen Qualität des revidierten Posters korreliert.

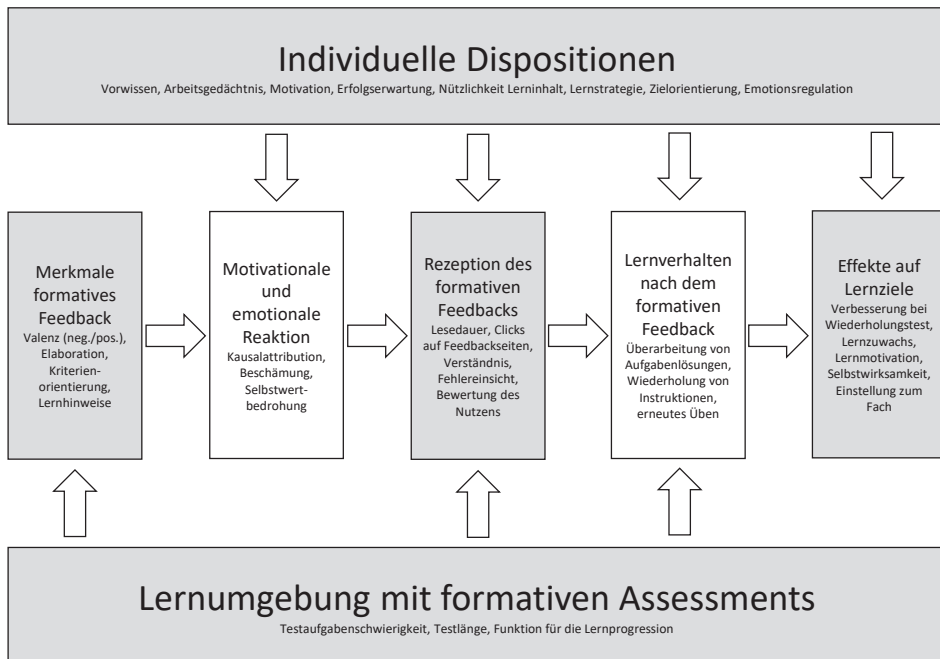


Abbildung 1: Rahmenmodell für die Feedbackrezeption (die in dieser Studie tangierten Elemente des Modells sind grau hinterlegt)

Es gibt mittlerweile eine Reihe von empirischen Studien und theoretischen Modellen, mit denen sich die Prozesse der Feedbackrezeption sowie moderierende Einflussfaktoren beschreiben lassen (Abbildung 1). Die Rezeption wird zunächst einmal durch Merkmale des Feedbacks beeinflusst (Kluger & DeNisi, 1996; Shute, 2008; Van der Kleij et al., 2015). Erstens sollte Feedback die Aufmerksamkeit auf die bearbeiteten Aufgaben und die Bewertungskriterien lenken. Weniger günstig ist selbstbezogenes Feedback, z. B. unspezifisches Lob, selbstwertbedrohliche Kritik oder sozial vergleichendes Feedback. Zweitens ist elaboriertes Feedback mit konkreten Hinweisen auf Fehler und Verbesserungsvorschlägen wirkungsvoller als einfaches Feedback (*knowledge of correct result*). Feedback darf allerdings nicht zu viel Information enthalten, um eine kognitive Überlastung zu vermeiden (Shute, 2008). Studien zeigen, dass sich bei Feedback mit einem größeren Informationsumfang (Länge der Feedbacktexte) die Lesedauer nicht gleichermaßen erhöht (Petrovic et al., 2017).

Merkmale der in die Lernumgebung integrierten formativen Assessments wirken sich ebenfalls auf die Feedbackrezeption aus. Es kann angenommen werden, dass formatives Feedback eher wahrgenommen wird, wenn beispielsweise die Lernprogression in einem Lernprogramm vom Bestehen des formativen Assessments abhängt (Maier, 2021). Timmers und Veldkamp (2011) zeigten zudem, dass die Anzahl angeklickter Feedbackseiten sinkt, wenn die Länge des formativen Assessments (Testdauer) oder die Testaufgabenschwierigkeiten ansteigen.

Aber auch dispositionale Merkmale von Lernenden wirken sich auf die Rezeption des formativen Feedbacks aus. Die Kapazität des Arbeitsgedächtnisses beispielsweise moderiert den Lerneffekt von Feedback. Bei Lernenden mit einer geringeren Kapazität ist einfaches Feedback lernwirksamer als elaboriertes Feedback (Fyfe et al., 2015). Die Feedbackrezeption hängt zudem mit dem Vorwissen und den strategischen Lernzielen der Lernenden zusammen. Personen mit geringen Testwerten zeigten in Experimenten die geringste Feedbackverweildauer, weil sie die gesamte Situation möglichst schnell beenden wollten (Hancock et al., 1995; Fox, 2013; Timmers & Veldkamp, 2011). Die Studie von Fox (2013) zeigt allerdings auch, dass es wichtig ist, das Feedbackverhalten genau zu erfassen. Die leistungsschwächeren Studierenden ignorierten zwar das Feedback eher, klickten allerdings im Falle des Nicht-Ignorierens mehr Feedbackseiten an. Dies spricht für eine ineffektive Feedbacklesestrategie. Das Feedback wird zwar aufgerufen, aber nicht gelesen oder nicht verstanden.

Nicht zuletzt spielen emotionale und motivationale Reaktionen auf formatives Feedback eine Rolle. Grundmann et al. (2020) nehmen in ihrem *model of motivated feedback disengagement* an, dass die emotionale Reaktion auf negatives Feedback im Zusammenspiel mit der motivationalen Zielorientierung das Feedbackverhalten beeinflussen. Wenn eine Person Feedback nach einer falsch gelösten Aufgabe als Bedrohung wahrnimmt und eine hedonistische Zielorientierung hat (sich gut fühlen vs. sich verbessern), wird Feedback bewusst ignoriert. Auch auf Basis von Erwartungs-Wert-Theorien lassen sich entsprechende Zusammenhänge zeigen. Studierende, die Feedback in einer digitalen Lernumgebung ignorieren, haben eher eine geringere Erfolgserwartung und

bewerten die Aufgabennützlichkeit kritischer (Timmers et al., 2013). Die motivationalen Variablen in dieser Studie korrelieren allerdings nicht mit der Feedbacklesedauer. Auch dies spricht dafür, dass die reine Feedbacklesedauer kein optimaler Indikator für die Feedbackrezeption ist.

3 Fragestellung

Die im vorangehenden Abschnitt skizzierten Befunde implizieren, dass man bei der Gestaltung und Optimierung von formativen Assessments in digitalen Lernumgebungen die Feedbackrezeption im Blick haben sollte. Bestimmte Merkmalskonstellationen bei den Lernenden und der Lernumgebung führen zu einer ungünstigen Feedbackwahrnehmung und dies wiederum beeinflusst das nachfolgende Lernverhalten negativ. In den bisherigen Studien wurde Feedbackrezeption in digitalen Lernumgebungen meistens über die Feedbacklesedauer und die Anzahl von Clicks auf Feedbackseiten operationalisiert. Dabei zeigten sich widersprüchliche Befunde, beispielsweise ist die reine Feedbacklesedauer kein Indikator für eine lernwirksame Feedbacknutzung. Dies kann unterschiedliche Ursachen haben. Gerade lernschwächere Schülerinnen und Schüler erhalten häufig ein längeres Feedback, was die Feedbackverweildauer erhöhen könnte, allerdings nicht zwangsläufig bedeutet, dass in dieser Zeit auch das Feedback gelesen oder verstanden wird. In dieser Studie soll deshalb zusätzlich die Verständlichkeit des formativen Feedbacks auf einer feinkörnigen Ebene von einzelnen, fehlerhaft bearbeiteten Aufgaben erfasst werden. Wir nehmen an, dass das Feedbackverständnis auf dieser Ebene einen zusätzlichen Beitrag zur Erklärung der Lerneffekte leistet.

Forschungsgegenstand ist die Web-App MasteryX (www.masteryx.de), die verschiedene Minikurse zur Übung und Wiederholung von Grammatik- und Rechtschreibthemen für Deutsch und Englisch anbietet (z. B. Wortarten, Zeitformen, Satzglieder usw.). Die einzelnen Themen sind in Levels gegliedert und die Lernenden starten mit einem formativen Assessment auf Level 1 (Abbildung 2). Die formativen Assessments prüfen, inwiefern grammatikalische Regeln in Satzbeispielen angewendet werden können. Im Kurs Kommasetzung beispielsweise müssen auf Level 1 in drei aufeinander folgenden, kurzen Texten die fehlenden Kommas ergänzt werden (Abbildung 3). Auf Level 1 werden dabei die Kommaeregeln „Aufzählung“ und „Satzreihe“ geprüft. Die Schwierigkeit der Kommaeregeln steigt mit den Levels an. Zeigen sich noch deutliche Verständnisschwierigkeiten bei der Bearbeitung eines formativen Assessments (Testwert unter 80 %) müssen Übungen zu den Regeln, die im Test fehlerhaft waren, absolviert werden. Bei einem Testwert von 80 % und mehr kann sofort das formative Assessment auf dem nächsten Level absolviert werden.

Im Fokus der Studie stehen die Schülerinnen und Schüler mit einem Testwert unter 80 %, die ein Assessment wiederholen müssen. Diese erhalten von der Lernumgebung ein formatives Feedback, das sie zur Wiederholung der Lerninhalte anregen soll. Das

Feedback ist mehrschrittig aufgebaut. Zunächst erfahren die Lernenden in einer kurzen Mitteilung ihren Punktwert (unter 80 %) und dass sie nach dem Absolvieren der Übungen den formativen Test noch einmal wiederholen sollen (Testfeedback). Danach werden die Lernenden gefragt, ob sie sich das Feedback zu den fehlerhaften Aufgaben ansehen möchten (Detailfeedback). Dabei wird für jede Aufgabe angegeben, welche Grammatikregel bzw. welche Rechtschreibregel man hätte beachten sollen (Abbildung 4). Die Lernenden erhalten somit nicht nur die korrekte Lösung, sondern auch einen Hinweis auf die Kommaregel, die in diesem Satz zur Anwendung kommt. Die Schülerinnen und Schüler können die fehlerhaften Aufgaben einzeln durchklicken und werden nach jedem einzelnen Fehler gefragt, ob sie verstanden haben, warum dies falsch war. Abschließend werden die Lernenden gefragt, ob das gesamte Feedback hilfreich war. Diese Frage konnten sie mit einem lachenden oder traurigen Smiley beantworten.

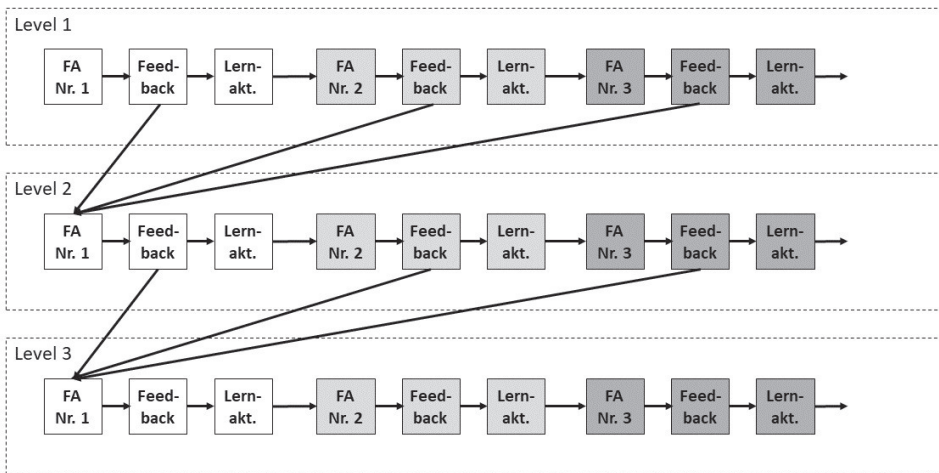


Abbildung 2: Mögliche Lernverläufe in einem Kurs

Setze Kommas! Klicke auf ein Wort, um ein Komma danach zu setzen.

Die Ölvorräte auf dieser Erde werden knapp, Millionen von Autos verschmutzen die Luft, Menschen leiden unter dem Lärm der Fahrzeuge. Das Auto war einmal ein riesiger Fortschritt und jetzt wird es zum Problem. Ingenieure Wissenschaftler Techniker und Designer arbeiten bereits an neuen Ideen. Elektroautos gelten als umweltfreundlich, sparsam und sauber. Doch man darf sich nicht täuschen lassen, die Produktion von Batterien kann auch große Umweltprobleme verursachen. Dann nehmen wir halt das Fahrrad, den Bus, den Zug oder die eigenen zwei Beine. Das ist auch gesünder.

Weiter

Abbildung 3: Beispiel für eine Testaufgabe (Kurs Kommasetzung, Level 1)

Im Test hattest Du 11 Fehler. Schau Dir genau an, wie es richtig gewesen wäre!

Richtige Lösung:

Ingenieure, Wissenschaftler, Techniker und Designer arbeiten bereits an neuen Ideen.

(Aufzählung)

(Fehler Nr. 1)

Ja klar, weiter

Verstehe ich nicht, weiter

Abbildung 4: Ausschnitt aus dem Detailfeedback (Kurs Kommasetzung, Level 1)

Die Studie soll klären, wie Feedbacklesedauer, Fehlerverständnis und Gesamtfeedbackbewertung mit dem Lernzuwachs beim wiederholt durchgeführten Assessment zusammenhängen. Hierzu werden beispielhaft die Daten aus den Kursen Kommasetzung, Feldermodell¹ und Wortarten untersucht. Hiermit soll geprüft werden, inwiefern die Effekte der Feedbackrezeption auf den Lernzuwachs auch vom Lerninhalt abhängen. Die Studie soll zudem einen Beitrag zur Erforschung der Feedbackrezeption in digitalen Lernumgebungen, die im Sekundarschulsystem zum Einsatz kommen, leisten. Viele Studien in diesem Forschungsfeld wurden an Universitäten durchgeführt (z. B. Timmers et al., 2013; Zimbardi et al., 2017; Young et al., 2021). Es ist anzunehmen, dass die Rezeption von digitalem Feedback im Sekundarschulsystem fragiler ist, weil die Lernenden in dieser Schulstufe weniger selbstreguliert lernen können als Studierende an Hochschulen. Van der Kleij et al. (2015) fanden beispielsweise in einer Metaanalyse höhere Effektstärken von Feedback in digitalen Lernumgebungen an Hochschulen im Vergleich zu Primar- und Sekundarschulen.

4. Methodische Vorgehensweise

4.1 Stichprobe

Die Web-App MasteryX wurde innerhalb eines Entwicklungs- und Forschungsprojektes an der Pädagogischen Hochschule Schwäbisch Gmünd einer Reihe von kooperierenden Sekundarschulen in Baden-Württemberg zur Verfügung gestellt. Die beteiligten Lehrkräfte wurden über Art, Umfang und Zweck der Datenerhebung ausführlich informiert und konnten das Lernangebot nur mit Zustimmung zu dieser Datenerhebung nutzen. Die in der Web-App angebotenen Online-Kurse sind curricular so organisiert, dass damit Rechtschreib- und Grammatikthemen

¹ Das sog. „Feldermodell“ ersetzt beispielsweise im Bildungsplan von Baden-Württemberg die Behandlung der Satzglieder und dient der Analyse von längeren Satzgefügen.

über mehrere Jahrgangsstufen der Sekundarstufe 1 abgedeckt werden. Die in diesem Beitrag untersuchte Stichprobe ist nicht mit der Gesamtzahl der Nutzerinnen und Nutzer identisch. Erstens wurden nur Daten von Schülerinnen und Schülern, die mindestens einen der drei Kurse Kommasetzung, Feldermodell und Wortarten bearbeitet haben, einbezogen. Zweitens wurden davon wiederum nur Lernende einbezogen, die mindestens eine Testwiederholungssequenz hatten, d. h. mindestens ein nicht bestandenenes Assessment einmal wiederholt haben. Damit ergab sich ein Stichprobenumfang von 528 Personen (288 Schülerinnen und 240 Schüler) in 63 Klassen. Tabelle 1 zeigt einen Stichprobenüberblick nach Jahrgangsstufen und Schulart. Die Web-App wurde schwerpunktmäßig an Realschulen und in den Jahrgangsstufen 8–10 erprobt. Da nur sehr wenige Datenpunkte von Hauptschulen, Gemeinschaftsschulen und Gymnasien stammen, wird die Variable Schulart in den nachfolgenden Analysen nicht mehr berücksichtigt.

Tabelle 1: Stichprobe nach Jahrgangsstufen und Schulart (Personen und Klassen)

	Gemeinschafts- schule	Gymnasium	Hauptschule	Realschule	Gesamt
Jgs. 6	7 (1)	3 (1)	4 (1)	7 (1)	21 (4)
Jgs. 7	5 (1)	6 (2)	3 (1)	28 (3)	42 (7)
Jgs. 8	19 (4)	4 (1)	-	129 (12)	152 (17)
Jgs. 9	37 (5)	-	4 (2)	126 (11)	167 (18)
Jgs. 10	-	-	3 (1)	143 (16)	146 (17)
gesamt	68 (11)	13 (4)	14 (5)	433 (43)	528 (63)

Anmerkung: Anzahl Schülerinnen und Schüler sowie Anzahl Klassen in Klammern

4.2 Datenaufbereitung, Analyseinheit und Variablen

Die in der Web-App generierten Log-Daten wurden auf verschiedenen Ebenen aufbereitet und analysiert, um die einzelnen Variablen zu extrahieren:

- Über die Meldebögen der Lehrkräfte wurden folgende Personenmerkmale erfasst: Jahrgangsstufe, Schulart, Geschlecht
- Bevor die Schülerinnen und Schüler einen Online-Kurs starten konnten, wurden sie gebeten, das Vorwissen und das Interesse am Thema selbst zu bewerten. Ebenso wurde gefragt, ob das Kursthema zeitgleich auch im Unterricht behandelt wird oder ob sie mit der Web-App lediglich ein Thema wiederholen.

Da in dieser Studie die Auswirkungen des Feedbackverhaltens auf die Wiederholung eines formativen Assessments untersucht werden, ist die Analyseinheit eine Test-Testwiederholungs-Sequenz. Leistungsstarke Lernende, die einen Kurs auf Anhieb durchlaufen, d. h. auf allen Levels das erste formative Assessment sofort bestehen, sind somit nicht in der Stichprobe enthalten. Leistungsschwächere Schülerinnen und

Schüler, die ein Kurslevel beginnen und am ersten formativen Assessment scheitern sowie einen Wiederholungsversuch unternehmen, steuern hingegen viele Daten bei. Diese Schülerinnen und Schüler könnten nach einem nicht bestandenen, zweiten Assessment und den obligatorischen Trainings einen dritten Versuch unternehmen (vgl. Abbildung 2). In dieser Studie wird allerdings nur die erste Testwiederholungssequenz ausgewertet. Damit stehen vor allem die Effekte der Feedbackrezeption von Lernenden im Fokus, die zwar über ein geringes Vorwissen verfügen, aber beim erstmaligen Scheitern nicht gleich aufgeben.

Für die erstmals nicht bestandenen, formativen Assessments liegen folgende Daten vor: Kurslevel (1, 2 oder 3), Testscore (wird als Vorwissen interpretiert), Lesedauer Testfeedback (Punktwert, Mitteilung ob Test bestanden und Hinweis auf Übung bzw. Wiederholung), Lesedauer Fehlerrückmeldungen (Feedback zu jeder einzelnen, fehlerhaften Aufgabe), Aufruf aller Fehlerrückmeldungen, Verständnis der einzelnen Fehlerrückmeldungen (Ja/Nein), Gesamtbewertung Feedback (Smileys). Weitere Kovariaten sind Aspekte des bisherigen Lernverhaltens in der Lernumgebung (bisherige Lernzeit in der App, Summe bisher bestandener, formativer Assessments als Indikator für das Leistungsniveau) sowie der zeitliche Abstand zwischen erstem formativem Assessment und Wiederholungstest als Indikator für die Arbeitsgeschwindigkeit. Die abhängige Variable ist der Testscore beim wiederholt durchgeführten, formativen Assessment (zweites formatives Assessment auf dem Kurslevel). Dieser Testwert muss nicht die letzte, durchgeführte Messung in einem Kurslevel sein (Abbildung 2). Schülerinnen und Schüler können auch nach einem nicht bestandenen, zweiten formativen Assessment weitere Wiederholungen durchführen, um das Kurslevel zu bestehen.

4.3 Testgüte der formativen Assessments

Die formativen Assessments strukturieren die Lernprogression in den Online-Kursen und sollen prüfen, ob Rechtschreib- oder Grammatikregeln durchweg korrekt in Satzbeispielen angewendet werden können. Im Kurs Kommasetzung wird beispielsweise geprüft, nach welchen Kommaeregeln die Schülerinnen und Schüler in einem kurzen Text fehlende Kommas ergänzen können (Abbildung 3). Die Prüfung der klassischen Testgütekriterien bei formativen Assessments gilt aus verschiedenen Gründen als Herausforderung (Way et al., 2010). Gerade wenn in Online-Kursen sehr viele, kleinschrittige Tests mit Wiederholungsmöglichkeiten zum Einsatz kommen sollen, sind aufwändige Pilotstudien schwierig. Die Items der formativen Assessments werden in diesem Projekt deshalb kontinuierlich geprüft und weiterentwickelt. Mittlerweile liegen formative Testitems aus einer mehrjährigen Projektentwicklung vor (Maier et al., 2017; Maier, 2020).

Da die Testversionen für die einzelnen Testwiederholungen aus einer Itemdatenbank zufällig zusammengestellt werden (vgl. Jia et al., 2012; Pelkola et al., 2018), können keine internen Konsistenzen (Cronbach's Alpha) berechnet werden. Um zu-

mindest einen Anhaltspunkt für die interne Konsistenz zu haben, wurde in einer Ergänzungsstudie ein Kommasetzungstest auf Niveau 3 mit 25 festen Items in der Web-App als Demotest freigeschaltet. Zum Zeitpunkt der Auswertung lagen 139 Testbearbeitungen vor ($M = 81.7\%$, $SD = 15.9\%$). Die interne Konsistenz (Cronbach's Alpha) für diesen Test lag mit 0.86 in einem guten Wertebereich. Die Itemtrennschärfen (item-total-correlation) der 25 Items liegen im Mittel bei 0.48 ($SD = 0.14$).

Um problematische Items auszusortieren, werden bei allen Items kontinuierlich Itemleichtigkeit und Trennschärfe analysiert. Vor allem die Trennschärfe steht in direktem Zusammenhang mit der internen Konsistenz und eignet sich deshalb, um die Reliabilität der Tests zu optimieren. Die Itemleichtigkeitswerte sind durchweg hoch und liegen im Mittel bei 0.86 ($SD = 0.15$). Sehr schwierige Aufgaben (Items mit geringer Itemleichtigkeit) wurden inhaltlich überprüft und überarbeitet bzw. gelöscht. In der Regel handelt es sich um fehlerhafte oder missverständlich formulierte Aufgaben. Sehr leichte Aufgaben wurden ebenfalls geprüft, bei Korrektheit jedoch im Aufgabenpool belassen, wenn ausreichend mittelschwere Items vorhanden sind. Die mittlere Trennschärfe liegt bei 0.45 ($SD = 0.26$) und damit ungefähr gleich wie beim Kommasetzungstest in der Ergänzungsstudie.

4.4 Analysestrategie

Für die 528 Schülerinnen und Schüler aus der Stichprobe liegen insgesamt 1006 Testwiederholungssequenzen vor (Kommasetzung: 311, Feldermodell: 365, Wortarten: 330). In einem ersten Schritt wurden alle Variablen auf der Ebene der Testwiederholungssequenzen in den drei Kursen deskriptivstatistisch ausgewertet. Anschließend wurde mit multiplen, linearen Regressionsanalysen in drei ausgewählten Kursen geprüft, wie sich die Feedbackvariablen auf den Punktescore beim wiederholt durchgeführten, formativen Assessment (AV) auswirken. Personenmerkmale, bisheriges Lernverhalten in der Web-App und Merkmale des ersten formativen Assessments (Testscore, Bearbeitungsdauer, subjektiv eingeschätzte Schwierigkeit) werden dabei als Kovariaten eingesetzt. Es werden standardisierte Regressionskoeffizienten, Varianzaufklärungsmaße sowie die Effektstärke Cohens f -Quadrat berichtet. Die Prüfung auf Multikollinearität mittels einer Korrelationsmatrix ergab, dass lediglich die Korrelation zwischen „Lernzeit in der App“ und „bisher bestandene, formative Assessments“ über $r = 0.7$ liegt (0.73). Alle anderen Korrelationen liegen unter 0.3. Damit sind lediglich zwei Variablen stark voneinander abhängig. Wenn man eine dieser Variablen aus den Regressionsmodellen herausnimmt, ändert sich allerdings nichts an den Ergebnissen. Die beiden Variablen sind zudem keine signifikanten Prädiktoren. Wir haben beide Variablen in den Modellen belassen, um zu zeigen, dass diese Informationen berücksichtigt wurden.

5 Ergebnisse

Die deskriptivstatistische Auswertung (Tabelle 2) verdeutlicht eine Reihe von Eigenheiten des gesamten Datensatzes sowie der ausgewählten Kurse. Bezüglich des Mittelwerts der Jahrgangsstufe gibt es keine substanziellen Unterschiede. Ebenso verteilen sich die Testwiederholungssequenzen bei allen drei Kursen über sehr unterschiedliche Zeitpunkte innerhalb der Gesamtnutzung der Web-App. Die Verteilung der bisherigen Lernzeit ist dabei linkssteil verteilt. In den meisten Fällen wurde das Kursthema bereits im Unterricht behandelt und die Lehrkräfte nutzen MasteryX als Übungs- und Wiederholungsmöglichkeit. Besonders hoch ist dieser Wert beim Kurs Wortarten. Dort schätzen die Schülerinnen und Schüler auch ihr Vorwissen am höchsten ein. Dies korrespondiert mit dem mittleren Punktescore des ersten formativen Assessments, der beim Kurs Wortarten mit 70 % deutlich höher liegt als beim Kurs Feldermodell mit 61 %. Der mittlere Punktescore sowie der Median der Testwiederholung liegen bei den Kursen Feldermodell und Wortarten bei über 80 %. Damit ist in mehr als der Hälfte der Fälle in diesen Kursen die Testwiederholung erfolgreich. Der Kurs Kommasetzung ist dagegen schwieriger und verzeichnet geringere Lernzuwächse.

Die Testbearbeitungsdauer schwankt sowohl zwischen den Kursen als auch innerhalb der Kurse sehr stark. Dies liegt einmal am Schwierigkeitsgrad und der variierenden Testlänge. Andererseits finden sich bei den Werten auch Ausreißer, die auf Bedienungsfehler zurückzuführen sind, beispielsweise wenn das formative Assessment nach kurzer Zeit aufgrund technischer Probleme abgebrochen wird. Diese Fälle wurden für die nachfolgende Auswertung ausgeschlossen. Für die Auswertung wichtig ist der Umstand, dass nur in 8 % bis 12 % der Testwiederholungssequenzen das Feedback zu den Fehlern übersprungen wird (Klick auf „Fehlerdetails interessieren mich nicht“). Der Umfang des Detailfeedbacks (Anzahl Worte) schwankt sowohl in Abhängigkeit der Kurse, der Levels als auch der Fehler pro Assessment. Ein sehr umfangreiches Detailfeedback gibt es beispielsweise beim formativen Assessment für Kommasetzung auf Level 4 mit sehr vielen Fehlern. Wichtig ist auch der Umstand, dass nicht in allen Testwiederholungssequenzen alle Fehler angeschaut wurden. Beim Kurs Kommasetzung wurden in 43 % der Fälle alle Fehlerdetails durchgeklickt. Beim Kurs Feldermodell ist dieser Wert deutlich geringer. Die Gesamtbewertung des Feedbacks liegt in allen drei ausgewählten Kursen mit über 60 % eher im positiven Bereich.

Tabelle 2: Mittelwert, Standardabweichung und Median der Variablen für die Regressionsanalyse in den drei ausgewählten Kursen

	Kommasetzung			Feldermodell			Wortarten		
	<i>M</i>	<i>SD</i>	<i>Med.</i>	<i>M</i>	<i>SD</i>	<i>Med.</i>	<i>M</i>	<i>SD</i>	<i>Med.</i>
Jahrgangsstufe	8.59	1.09	9.00	8.80	1.03	9.00	9.04	0.98	9.00
Geschlecht (0: m, 1: w)	0.56	0.50	1.00	0.59	0.49	1.00	0.58	0.49	1.00
Bisherige Lernzeit in der App (in Stunden)	3.40	4.89	2.10	2.91	4.99	1.37	3.07	4.10	1.77
Bisher bestandene for- mative Assessments	5.82	10.92	3.00	4.27	6.07	2.00	5.26	5.97	3.00
Kursthema im Unter- richt (Nein=0, Ja=1)	0.75	0.43	1.00	0.72	0.45	1.00	0.90	0.30	1.00
Vorwissen Kursthema (Nein=0, Ja=1)	0.55	0.50	1.00	0.44	0.50	0.00	0.60	0.49	1.00
Interesse Kursthema (Nein=0, Ja=1)	0.40	0.49	0.00	0.43	0.50	0.00	0.44	0.50	0.00
Kurslevel (1-3 in allen Kursen außer Komma mit 4 Levels)	1.73	0.89	1.00	1.76	0.76	2.00	1.65	0.73	2.00
Punktwert des 1. form. Assessments	0.67	0.16	0.70	0.61	0.24	0.67	0.70	0.21	0.74
Einschätzung, ob Testaufgaben schwer waren (Nein=0, Ja=1)	0.50	0.50	0.00	0.49	0.50	0.00	0.55	0.50	1.00
Testbearbeitungs- dauer (sec.)	320.66	229.06	265.92	247.53	227.08	186.82	399.06	272.28	326.11
Lesedauer Testfeedback (in Sekunden)	8.16	10.70	5.21	7.75	11.32	4.52	7.82	10.84	4.80
Fehlerdetails interessieren nicht	0.12	0.32	0.00	0.11	0.32	0.00	0.08	0.28	0.00
Umfang Fehlerdetails (Anzahl Worte)	1700.68	1254.86	1372.00	1176.29	808.48	1035.00	1574.76	1103.78	1381.00
Lesedauer Fehlerdetails	45.53	344.98	11.66	26.03	158.50	0.00	52.59	384.96	6.87
Anzahl Klicks „Fehler verstanden“	1.28	1.86	1.00	0.78	1.39	0.00	1.25	1.95	0.00
Anzahl Klicks „Fehler unklar“	0.37	0.88	0.00	0.30	0.94	0.00	0.49	1.24	0.00
Alle Fehlerdetails an- geklickt (Nein=0, Ja=1)	0.43	0.50	0.00	0.28	0.45	0.00	0.32	0.47	0.00
Gesamtbewertung Feedback (negativ=0, positiv=1)	0.64	0.48	1.00	0.67	0.47	1.00	0.62	0.49	1.00
Punktwert des 2. form. Assessments (AV)	0.76	0.16	0.78	0.84	0.17	0.88	0.81	0.17	0.85

Die Ergebnisse der Regressionsanalysen für die Lernzuwächse in den Testwiederholungssequenzen der drei Kurse sind in Tabelle 3 dargestellt. Zunächst fällt auf, dass die drei Kurse kein gemeinsames Muster signifikanter Prädiktoren zeigen. Beim Kurs Wortarten finden sich insgesamt die meisten signifikanten beta-Werte, die höchste Gesamtvarianzaufklärung und die Effektstärke der Prädiktoren liegt in einem hohen Bereich (Cohen's f-Quadrat). Deutlich geringer sind die Varianzaufklärungen bei den beiden anderen Kursen. Wie zu erwarten, hat bei allen drei Kursen der Punktwert des ersten formativen Assessments den höchsten Erklärungswert für den Punktwert des zweiten, wiederholt durchgeführten Assessments. Bei einem höheren Level ist bei den Kursen Feldermodell und Wortarten mit einem geringeren Lernzuwachs zu rechnen. Die subjektiv eingeschätzte Schwierigkeit der Testaufgaben sowie eine hohe Testbearbeitungsdauer haben zumindest beim Kurs Wortarten einen geringen, negativen Effekt auf den Testscore des wiederholten Assessments.

Weiter fällt auf, dass sich sowohl Jahrgangsstufe als auch Geschlecht im Kurs Kommasetzung auf den Lernzuwachs auswirken. Vor allem älteren Schülern scheint es deutlich besser zu gelingen, die Kommaregeln anhand der Web-App aufzufrischen. Die bisherige Lernzeit in der Web-App sowie die bisher positiv absolvierten Assessments hängen nicht mit dem Lernfortschritt innerhalb einer Testwiederholungssequenz zusammen. Dies lässt sich vor allem darauf zurückführen, dass die Lehrkräfte die einzelnen Kurse in sehr unterschiedlicher Reihenfolge bearbeiten lassen und somit der Gesamtlernerfolg in der Web-App keine Information zum spezifischen Vorwissensstand auf einem Kurslevel liefert.

Bei den Feedbackvariablen ergibt sich ein sehr unterschiedliches Bild für die drei Kurse. Für die Kurse Kommasetzung und Feldermodell zeigen die Regressionsanalysen keine signifikanten Zusammenhänge an. Das Feedbackverhalten nach dem ersten formativen Assessment wirkt sich nur beim Kurs Wortarten auf den Punktezuwachs beim zweiten Assessment aus. Interessant ist hierbei zunächst, dass das absichtliche Überspringen der gesamten Detailrückmeldung (Fehlerdetails interessieren nicht) mit dem Lernzuwachs zusammenhängt. Dies ist nur in 8 % der Testwiederholungssequenzen im Kurs Wortarten der Fall. Hier scheinen sich die Lernenden aber eher sicher zu sein, welche Fehler sie gemacht haben. Werden die Detailrückmeldungen aufgerufen, dann zeigt sich, dass sowohl die als „verstanden“ als auch die als „nicht verstanden“ angeklickten Fehlerdetails positiv mit dem Lernzuwachs innerhalb der Testwiederholungssequenz zusammenhängen. Wenn allerdings alle Fehlerdetails angeklickt werden, ist der Lernzuwachs geringer. Eine mögliche Erklärung hierfür ergibt sich aus Beobachtungen zum Umgang mit der Web-App. Es gibt eine Reihe von Lernenden, die alle Fehlerdetails lediglich schnell durchklicken, ohne diese zu lesen.

Tabelle 3: Ergebnisse der linearen Regressionsanalyse mit dem Wert des zweiten formativen Assessments als abhängiger Variable (standardisierte Regressionskoeffizienten)

	Kommasetzung	Feldermodell	Wortarten
Prädiktoren	Beta (S.E.)	Beta (S.E.)	Beta (S.E.)
<i>Personenmerkmale</i>			
(Intercept)	-0.21 (0.29)	0.00 (0.05)	-0.00 (0.04)
Jahrgangsstufe	0.23 (0.07) ***	0.01 (0.05)	0.01 (0.05)
Geschlecht: weiblich	-0.13 (0.06) *	0.05 (0.05)	0.00 (0.05)
<i>Bisheriges Lernverhalten</i>			
Lernzeit in der App	-0.17 (0.10)	0.04 (0.09)	-0.06 (0.06)
Bestandene formative Assessments	0.11 (0.09)	0.09 (0.08)	0.08 (0.06)
<i>Angaben Lernende zum Kursthema</i>			
Kursthema im Unterricht	0.06 (0.06)	-0.01 (0.06)	-0.05 (0.05)
Vorwissen Kursthema	0.10 (0.06)	0.08 (0.06)	0.01 (0.05)
Interesse Kursthema	0.03 (0.06)	-0.02 (0.05)	0.04 (0.05)
<i>Merkmale des ersten formativen Assessments</i>			
Kurslevel	-0.04 (0.07)	-0.21 (0.06) ***	-0.13 (0.06) *
Punktwert	0.26 (0.06) ***	0.40 (0.09) ***	0.37 (0.10) ***
Testaufgaben waren schwer	-0.01 (0.06)	-0.04 (0.05)	-0.11 (0.05) *
Testbearbeitungsdauer	-0.07 (0.06)	0.05 (0.06)	-0.11 (0.05) *
<i>Merkmale Feedback nach dem ersten formativen Assessment</i>			
Lesedauer Testfeedback	0.03 (0.06)	-0.09 (0.05)	-0.03 (0.05)
Fehlerdetails interessieren nicht (Ja)	0.06 (0.06)	0.05 (0.05)	0.10 (0.05) *
Umfang Fehlerdetails (Anzahl Worte)	-0.05 (0.07)	0.14 (0.05)	-0.15 (0.10)
Lesedauer Fehlerdetails	0.03 (0.06)	-0.01 (0.10)	0.05 (0.05)
Anzahl Klicks „Fehler verstanden“	0.08 (0.06)	-0.02 (0.06)	0.26 (0.06) ***
Anzahl Klicks „Fehler unklar“	0.07 (0.06)	0.06 (0.06)	0.15 (0.05) **
Alle Fehlerdetails angeklickt (Ja)	0.06 (0.07)	0.05 (0.07)	-0.14 (0.06) *
Gesamtbewertung Feedback	-0.07 (0.06)	0.05 (0.06)	0.03 (0.05)
<i>Modellangaben</i>			
Multiples R-Quadrat	0.19	0.16	0.40
Adjustiertes multiples R-Quadrat	0.14	0.12	0.36
Effektstärke Cohens f-Quadrat	0.24	0.20	0.66
N = Anzahl Testwiederholungssequenzen	311	365	330

Anmerkungen: Lineare Regression mit simultaner Einführung aller Prädiktoren; alle metrischen Prädiktoren sind z-standardisiert; Signifikanz: *** $p < .001$; ** $p < .01$; * $p < .05$; Angaben in Klammern sind die Standardfehler der beta-Werte

6 Diskussion der Befunde und Schlussfolgerungen

Die Ergebnisse der Studie ermöglichen einen Einblick in die konkrete Wirkweise des formativen Feedbacks der Web-App MasteryX auf einer feinkörnigen Ebene einzelner, fehlerhafter Aufgaben. In allen drei Kursen ist die reine Lesedauer der detaillierten Fehlerrückmeldungen kein Prädiktor für den Lernzuwachs, wenn ein formatives Assessment wiederholt werden muss. Dies korrespondiert mit den Ergebnissen der Studie zur Feedbacknutzung von Timmers et al. (2013). Dafür zeigte sich zumindest beim Kurs Wortarten, dass die Feedbackrezeption mit Lernzuwächsen korreliert. Obwohl alle drei Kurse strukturell aufgebaut sind, scheint es den Lernenden im Kurs Wortarten wesentlich besser zu gelingen, aus den Fehlerrückmeldungen zu lernen. Unsere Vermutung ist, dass dies von der inhaltlichen Ausgestaltung der Kurse abhängt. Beim Kurs Wortarten fragen die einzelnen Testitems jeweils präzise die Kenntnis einer Wortart ab. Schülerinnen und Schüler mit geringem Vorwissen im Bereich Wortarten erhalten nach dem formativen Assessment eine hohe Anzahl an Fehlerrückmeldungen. Wenn diese Lernenden dann möglichst präzise zwischen den einzelnen Fehlern differenzieren, ob sie diese verstanden haben oder nicht, ist dies ein Hinweis auf eine produktive, lernwirksame Auseinandersetzung mit dem Feedback. Dieses ausgeprägte Fehlerverständnis erhöht vermutlich die Bereitschaft, die nachfolgenden Trainings als Lernchance zu sehen. Bei den beiden anderen Kursen Kommasetzung und Feldermodell ist der innere Zusammenhang zwischen Fehlerrückmeldung und zu übender Grammatikregel weniger offensichtlich. Sowohl die Aufgabenstellung als auch die Fehlerrückmeldungen umfassen mehr Text und die Lernenden müssen für das Verständnis der fehlerhaften Antworten die komplexe Grammatikregel noch einmal auf das jeweilige Textbeispiel anwenden. Das Feedback ist in diesen beiden Kursen somit weniger eindeutig als beim Kurs Wortarten. Damit sind auch die Angaben der Schülerinnen und Schüler, ob sie die jeweiligen Fehler verstehen oder nicht, weniger mit der tatsächlichen Einsicht in die eigenen Lerndefizite assoziiert bzw. die Fehlerrückmeldung und die durch die App initiierte Reflexion erreichen eine wesentlich geringere Lernwirksamkeit. Für diese Interpretation sprechen auch die deskriptiven Daten. Beim Kurs Wortarten ist die Lesedauer des Detailfeedbacks im Mittel am höchsten und es werden im Mittel am meisten Fehler als unklar bezeichnet. Bei den Kursen Kommasetzung und Feldermodell ist hingegen die Lesedauer im Vergleich zum Textumfang des Feedbacks deutlich geringer, was auf eine oberflächlichere Auseinandersetzung mit dem Feedback bzw. auf ein „Durchklicken“ der Fehlerrückmeldungen hindeutet.

Die Befunde dieser Studie knüpfen damit auch an die Metaanalyse zu itembasiertem Feedback in digitalen Lernumgebungen von Van der Kleij et al. (2015) an. In dieser Metaanalyse zeigte sich, dass elaboriertes Feedback auf Itemebene vor allem dann mit hohen Lerneffekten assoziiert ist, wenn die Rückmeldung zur Korrektheit von Aufgaben („knowledge of result“) mit gezielten Erklärungen und Lernhinweisen kombiniert wird. Diese Kombination kann sowohl gut als auch weniger gut umgesetzt werden. Van der Kleij et al. (2015) vermuten beispielsweise, dass die Kombination aus Feedback mit

richtiger Lösung („knowledge of correct result“) und elaboriertem Feedback weniger lernwirksam ist als die Kombination aus Korrektheit der Aufgabe („knowledge of result“) und elaboriertem Feedback. Im ersten Fall wird den Lernenden bereits die korrekte Lösung angezeigt. Damit sinkt die Motivation, sich mit den Erklärungen und Lernhinweisen zu beschäftigen. Wird allerdings nur angezeigt, dass die Aufgabe falsch war, könnte dies eher dazu beitragen, dass sich die Schülerinnen und Schüler über die Erklärungen zur Aufgabenstellung Gedanken machen, um die richtige Lösung noch eigenständig herauszufinden. Im Hinblick auf die Weiterentwicklung der Lernapp wird bei den Kursen Feldermodell und Kommasetzung deshalb zu prüfen sein, wie man das Detailfeedback neu arrangiert, sodass in Zukunft eine Feedbackrezeption auch zu Lernzuwächsen beitragen kann. Eine Möglichkeit wäre, eine Aufgabe lediglich als falsch anzuzeigen („knowledge of result“) und die Lernenden aufzufordern, anhand der eingblendeten Regel die richtige Lösung selbst zu finden.

Die Befunde der Studie müssen vor dem Hintergrund forschungsmethodischer Limitationen kritisch eingeordnet werden. Bei der Analyse der Zusammenhänge wurde die hierarchische Datenstruktur nicht berücksichtigt. Die Daten sind mehrfach geschichtet (Klassen, Lernende, formative Assessments zu den Kursthemen). Aufgrund der sehr unterschiedlichen Nutzung der Web-App in den einzelnen Klassen wäre ein hierarchisch strukturierter Datensatz jedoch nicht ausbalanciert. Die Anzahl einzelner Kursbearbeitungen pro Klasse schwankt stark. Damit bleibt die auf Unterschiede zwischen Klassen zurückzuführende Varianz in den Analysen unberücksichtigt. Eine weitere Limitation der hier vorgestellten Studie ist die Reduktion auf einzelne Facetten der Feedbackrezeption (Abbildung 1). Es fehlen beispielsweise Daten zur unmittelbaren, emotionalen Reaktion auf das negative Feedback. Zusammen mit Selbstwirksamkeitserwartungen und der motivationalen Zielorientierung ist die unmittelbare emotionale Reaktion eine wichtige Determinante für das Feedbackengagement (Grundmann et al., 2020). Eine Erfassung von emotionalen Facetten der Feedbackrezeption wäre technisch zwar prinzipiell möglich, würde aber vermutlich den Arbeitsfluss in der Lern-App noch mehr stören. Die Operationalisierung des Detailfeedbackverständnisses ist zudem fehleranfällig. Vermutlich klicken manche Schülerinnen und Schüler bei der Frage, ob sie den Fehler verstanden haben, immer mit „Ja“, um schnell alle Fehler einsehen zu können, auch wenn sie nicht jeden Fehler nachvollziehen können. Letztendlich ist in dieser Studie das Lernverhalten nach dem formativen Feedback nicht abgebildet. Es wäre beispielsweise von Interesse, ob und wie die Lehrkräfte einzelne Schülerinnen und Schüler nach einem nicht bestandenen, formativen Assessment unterstützen. Dieser zusätzliche Blick auf die instruktionale Rahmung der Feedbackrezeption wäre in einer weiterführenden Beobachtungsstudie denkbar.

Insgesamt zeigt die Studie, dass es sich lohnt, die einzelnen Phasen und Aspekte der Rezeption und Nutzung von formativem Feedback genau aufzuschlüsseln. Einfach zu messende Variablen wie die Feedbacklesedauer oder die Anzahl von Feedback-Clicks haben nur eine begrenzte Aussagekraft (z. B. Fox, 2013). Vielmehr muss man die Rezeption und Nutzung von formativem Feedback als einen vielschichtigen Prozess

verstehen, bei dem sowohl kognitive Verarbeitungsprozesse als auch emotionale und motivationale Bewertungen eine Rolle spielen (vgl. Rahmenmodell in Abbildung 1). Gerade weil formative Assessments in vielen digitalen Lernmaterialien eine Steuerungsfunktion übernehmen, sollte die Vielschichtigkeit der Feedbackrezeption beim Design in den Blick genommen werden. Zudem sollte man in weiteren Untersuchungen bzw. Modellentwicklungen dem detaillierten Verständnis von einzelnen Rückmeldungen mehr Aufmerksamkeit widmen.

Literatur

- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Bokhove, C. & Drijvers, P. (2012). Effects of a digital intervention on the development of algebraic expertise. *Computers & Education*, 58, 197–208. <https://doi.org/10.1016/j.compedu.2011.08.010>
- Cutumisu, M., Blair, K. P., Chin, D. B. & Schwartz, D. L. (2017). Assessing whether students seek constructive criticism: The design of an automated feedback system for a graphic design task. *International Journal of Artificial Intelligence in Education*, 27, 419–447. <https://doi.org/10.1007/s40593-016-0137-5>
- Fox, J.-P. (2013). Multivariate zero-inflated modeling with latent predictors: Modeling feedback behavior. *Computational Statistics & Data Analysis*, 68, 361–374. <https://doi.org/10.1016/j.csda.2013.07.003>
- Fuchs, L.S. & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53, 199–208. <https://doi.org/10.1177/001440298605300301>
- Fyfe, E. R., DeCaro, M. S. & Rittle-Johnson, B. (2015). When feedback is cognitively-demanding: The importance of working memory capacity. *Instructional Science*, 43(1), 73–91. <https://doi.org/10.1007/s11251-014-9323-8>
- Grundmann, F., Scheibe, S. & Epstude, K. (2020). When ignoring negative feedback is functional: Presenting a model of motivated feedback disengagement. *Current Directions in Psychological Science*. <https://doi.org/10.1177/0963721420969386>
- Hancock, T. E., Thurman, R. A. & Hubbard, D. C. (1995). An expanded control model for the use of instructional feedback. *Contemporary Educational Psychology*, 20(4), 410–425. <https://doi.org/10.1006/ceps.1995.1028>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review* 17, 50–62. <https://doi.org/10.1016/j.edurev.2015.12.002>
- Jia, J., Chen, Y., Ding, Z. & Ruan, M (2012). Effects of a vocabulary acquisition and assessment system on students' performance in a blended learning class for English subject. *Computers & Education*, 58(1), 63–76. <https://doi.org/10.1016/j.compedu.2011.08.002>

- Kingston, N. & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Klauer, K. J. & Strathmann, A. M. (2013). Lernverlaufsdiagnostik Mathematik: Test auf Änderungssensibilität bei rechenschwachen Grundschulern. *Psychologie in Erziehung und Unterricht*, 60(4), 241–252. <https://doi.org/10.2378/peu2013.art18d>
- Kluger, A. N. & DeNisi, A. (1996). The effects of Feedback Interventions on performance: A historical review, a meta-analysis, and a preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Kölller, O. (2005). Formative assessment in classrooms: A review of the empirical German literature. In OECD (Hrsg.), *Formative Assessment – Improving Learning in Secondary Classrooms* (S. 265–279). OECD.
- Lee, H., Chung, H. Q., Zhang, Y., Abedi, J. & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, 33(2), 124–140. <https://doi.org/10.1080/08957347.2020.1732383>
- Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, 13, 293–308. <https://doi.org/10.1007/s11618-010-0124-9>
- Maier, U. (2020). Modellierung der Effekte lernzielorientierter Tests nach dem Prinzip des mastery measurement innerhalb einer digitalen, formativen Leistungsmessung. *Zeitschrift für Erziehungswissenschaft*, 23, 769–791. <https://doi.org/10.1007/s11618-020-00958-6>
- Maier, U. (2021). Self-referenced vs. reward-based feedback messages in online courses with formative mastery assessments: A randomized controlled trial in secondary classrooms. *Computers and Education*, 174, 104–306. <https://doi.org/10.1016/j.compedu.2021.104306>
- Maier, U., Ramsteck, C. & Hoffmann, K. (2017). Formative Leistungsdiagnostik und Learning Analytics: Entwicklung, Nutzung und Optimierung eines onlinebasierten Kurses für die Diagnostik und Förderung von Grundwissen im Kompetenzbereich Sprachbetrachtung. *Zeitschrift für Erziehungswissenschaft*, 20, 728–747. <https://doi.org/10.1007/s11618-017-0776-9>
- Paiva, R. C. Ferreira, M. S. & Frade, M. M. (2017). Intelligent tutorial system based on personalized system of instruction to teach or remind mathematical concepts. *Journal of Computer Assisted Learning*, 33(4), 370–381. <https://doi.org/10.1111/jcal.12186>
- Pelkola, T., Rasila, A. & Sangwin, C. (2018). Investigating Bloom's learning for mastery in mathematics with online assessment. *Informatics in Education*, 17(2), 363–380. <https://doi.org/10.15388/infedu.2018.19>
- Petrovic, J., Pale, P. & Jeren, B. (2017). Online formative assessments in a digital signal processing course: Effects of feedback type and content difficulty on students learning achievements. *Education and Information Technologies*, 22, 3047–3061. <https://doi.org/10.1007/s10639-016-9571-0>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Souvignier, E., Schütze, B. & Hasselhorn, M. (2018). Stichwort – Formatives Assessment. *Zeitschrift für Erziehungswissenschaft*, 21, 697–715. <https://doi.org/10.1007/s11618-018-0838-7>
- Tempelaar, D. (2020). Supporting the less-adaptive student: The role of learning analytics, formative assessment and blended learning. *Assessment & Evaluation in Higher Education*, 45(4), 579–593. <https://doi.org/10.1080/02602938.2019.1677855>

- Timmers, C. F., Braber-van den Broek, J. & van den Berg, S. M. (2013). Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment. *Computers & Education*, 60, 25–31. <https://doi.org/10.1016/j.compedu.2012.07.007>
- Timmers, C. & Veldkamp, B. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education*, 56, 923–930. <https://doi.org/10.1016/j.compedu.2010.11.007>
- Van der Kleij, F. M., Feskens, R. C. W. & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475–511. <https://doi.org/10.3102/0034654314564881>
- Way, W.D., Dolan, R.P. & Nichols, P. (2010). Psychometric challenges and opportunities in implementing formative assessment. In H.L. Andrade & G.J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 297–315). Routledge.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Young, K. R., Schaffer, H. E., James, J. B. & Gallardo-Williams, M. T. (2021). Tired of failing students? Improving student learning using detailed and automated individualized feedback in a large introductory science course. *Innovative Higher Education*, 46, 133–151. <https://doi.org/10.1007/s10755-020-09527-5>
- Zimbardi, K., Colthorpe, K., Dekker, A., Engstrom, C., Bugarcic, A., Worthy, P., Victor, R., Chunduri, P., Lluka, L. & Long, P. (2017). Are they using my feedback? The extent of students' feedback use has a large impact on subsequent academic performance. *Assessment & Evaluation in Higher Education*, 42(4), 625–644. <https://doi.org/10.1080/02602938.2016.1174187>

Der Prozess formativen Assessments beim Experimentieren

Welche Rolle spielt die kognitive Belastung?

Mathias Ropohl¹ | Hilda Weimar² | Julia Arnold³ | Svenja Boegel¹ | Nele Kampa⁴ | Maria Opfermann⁵

¹Universität Duisburg-Essen | ²Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik | ³Pädagogische Hochschule FHNW | ⁴Universität Wien |

⁵Bergische Universität Wuppertal

Abstract: Der vorliegende Beitrag betrachtet die Rolle formativen Assessments im Zusammenhang mit naturwissenschaftlichen Denk- und Arbeitsweisen, insbesondere dem Planen von Experimenten. Im Fokus steht die Frage, ob der Erwerb von Wissen über die Denk- und Arbeitsweisen davon profitiert, dass mithilfe fortlaufenden formativen Assessments Erkenntnisse zum individuellen Kompetenzgewinn gewonnen werden können, die wiederum beispielsweise durch individuelle Rückmeldungen den weiteren Lernprozess positiv beeinflussen. Des Weiteren wird die Frage nach der Rolle der kognitiven Belastung in einer komplexen Lerngelegenheit aufgegriffen.

Im Beitrag wird zunächst auf die theoretischen Grundlagen des Experimentierens als wissenschaftlicher Denk- und Arbeitsweise sowie auf Grundlagen und Wirkung formativen Assessments eingegangen. Im Anschluss wird eine Studie vorgestellt, welche die Rolle von Rückmeldungen beim Experimentieren mit unterschiedlichen Öffnungsgraden beleuchtet. Dabei zeigen sich insbesondere beim Experimentplanen mit vorgegebenen Fragestellungen und Hypothesen Rückmeldungen als lernwirksam, während die Befundlage zur kognitiven Belastung weiterer Forschung bedarf.

1 Einführung

Der Aufbau von Kompetenzen im Bereich der naturwissenschaftlichen Denk- und Arbeitsweisen wie dem Experimentieren ist in Deutschland Teil der naturwissenschaftlichen Grundbildung (KMK, 2005). Dabei ist ihr Erwerb für Schülerinnen und Schüler herausfordernd (z. B. Weirich et al., 2019), weil Wissen über den zu untersuchenden fachlichen Inhalt mit Wissen über naturwissenschaftliche Denk- und Arbeitsweisen kombiniert werden muss (Arnold et al., 2014; Meier & Mayer, 2014). Es wird vermutet, dass entsprechende Lerngelegenheiten für Lernende kognitiv belastend sind und daher Lernprozesse regelmäßig nicht zu Lernzuwachs führen.

Infolgedessen ist die Untersuchung von Formaten der Lernunterstützung zur Reduzierung der kognitiven Belastungen ein Schwerpunkt naturwissenschaftsdidaktischer Forschung (z. B. Arnold et al., 2017). Häufig gehen diese Formate mit formativem

Assessment einher, welches als Schrittabfolge beschrieben werden kann, deren konzentrierte Aktivitäten einen Beitrag zur Individualisierung von Lehr-Lern-Prozessen leisten (Harlen, 2013; Ropohl & von Aufschnaiter, 2021).

Bisher wurde formatives Assessment insbesondere im Zusammenhang mit Wissen über den fachlichen Inhalt operationalisiert und weniger mit Wissen über Denk- und Arbeitsweisen wie dem Experimentieren (z. B. Hild et al., 2020; Ropohl & Scheuermann, 2018). Vor diesem Hintergrund wird im Beitrag zunächst der Prozess formativen Assessments am Beispiel des Experimentierens erläutert und dann erste empirische Befunde zur Wirksamkeit formativen Assessments auf das Planen von Experimenten berichtet (z. B. Scheuermann, 2017). Trotz der Befunde sind Fragen ungeklärt: Zum einen welche Wirkungen Lerngelegenheiten unterschiedlicher Öffnungsgrade beim Experimentieren mit zusätzlichem formativem Assessment haben und zum anderen welche Rolle dabei die wahrgenommene kognitive Belastung der Lernenden spielt. In dem nachfolgenden Beitrag wird eine Untersuchung zur Beantwortung dieser Fragestellungen präsentiert.

2 Experimentieren als naturwissenschaftliche Arbeitsweise

Im naturwissenschaftlichen Unterricht werden fachspezifische Arbeitsweisen wie das Experimentieren eingeführt (z. B. MBWK SH, 2019), u. a. um eigenständiges Untersuchen von Fragestellungen zu ermöglichen (Emden et al., 2019; Rönnebeck et al., 2016). Die Spezifik des Experimentierens wird in der gezielten und kontrollierten Untersuchung von kausalen Zusammenhängen gesehen (Schulz et al., 2012). Schülerinnen und Schüler müssen dabei die Variablenkontrollstrategie anwenden, indem sie die potenzielle Ursache für einen Effekt aktiv manipulieren und im Anschluss daran etwaige Veränderungen erfassen. Alle weiteren potenziellen Einflussfaktoren müssen konstant gehalten oder gezielt kontrolliert werden (Chen & Klahr, 1999).

Mit Blick auf die Kompetenzen der Lernenden wurden in verschiedenen Untersuchungen deren Schwierigkeiten beim Anwenden naturwissenschaftlicher Arbeitsweisen, gerade auch des Experimentierens, diagnostiziert (z. B. Arnold et al., 2014; Künsting et al., 2008; Walpuski & Sumfleth, 2007; Zenner et al., 2017). Ein Grund kann in dem fehlenden Wissen über die Arbeitsweisen gesehen werden (Weirich et al., 2019), das häufig die Auswahl eines weniger zielführenden Vorgehens beim Experimentieren – beispielsweise das Formulieren einer nicht passenden Fragestellung, das Aufstellen einer nicht zielführenden Hypothese oder das Planen eines unsystematischen Experiments – begründet (Wahser & Sumfleth, 2008).

Aus diesen und weiteren Befunden wurden ein erhöhter Unterstützungs- und Strukturierungsbedarf für die Arbeitsweise des Experimentierens abgeleitet (Blanchard et al., 2010; Minner et al., 2010). Daher wurden Lerngelegenheiten entwickelt und evaluiert, die unterschiedliche Öffnungsgrade hinsichtlich des Fachinhalts, der Strategie, der Methode, der Anzahl der Lösungen und Lösungswege repräsentieren (Priemer, 2011).

Hinsichtlich der hier relevanten Arbeitsweise des Experimentierens kann die Offenheit bestimmt werden, indem der Grad der Beteiligung der Lehrkraft an der Umsetzung abgewogen wird (Abrams et al., 2008; Blanchard et al., 2010; Furtak et al., 2012). Eine geringe Offenheit ist gekennzeichnet durch stark lehrendengelenktes forschendes Lernen, eine hohe Offenheit durch stark lernendengelenktes forschendes Lernen. Als Übergangsform wird das geführte forschende Lernen definiert, bei dem sowohl der Lehrkraft als auch den Lernenden eine hohe Beteiligung am Experimentieren zugesprochen wird.

Empirische Befunde zeigen, dass prozessbezogene Kompetenzen insbesondere beim geführten forschenden Lernen gefördert werden und dass dieses lernwirksamer als stark lehrendengelenktes forschendes Lernen ist (Blanchard et al., 2010; Furtak et al., 2012) sowie als regulärer Unterricht, der nicht der Konzeption des forschenden Lernens folgt (Dean & Kuhn, 2007). Hinsichtlich der Wirksamkeit von geführtem forschendem Lernen im Kontrast zu stark lernendengelenktem forschendem Lernen wurde kein signifikanter Unterschied identifiziert (Furtak et al., 2012). Insgesamt ist die Befundlage zur Wirkung der Öffnungsgrade auf Lernprozesse in Kombination mit formativem Assessment unzureichend. Es ist davon auszugehen, dass durch die zunehmende Komplexität von Lerngelegenheiten die kognitive Belastung deren Wirkung mediiert.

3 Formatives Assessment beim Experimentieren

Zweck formativen Assessments ist die Optimierung von Lernprozessen. Dabei werden verschiedene aufeinander aufbauende Komponenten formativen Assessments unterschieden (Ropohl & von Aufschnaiter, 2021; vgl. Abbildung 1). Ausgehend von der Festlegung eines konkreten Lernziels für die Schülerinnen und Schüler muss ein darauf aufbauender Diagnoseprozess initiiert werden, in dessen Zuge Entscheidungen zum Diagnoseverfahren, den relevanten Daten und zu Konsequenzen aus den Beobachtungen getroffen werden müssen, und an dessen Ende eine Rückmeldung an die Lernenden formuliert wird.

Eine beispielhafte Umsetzung des Prozesses formativen Assessments hat Scheuermann (2017) ausgearbeitet (s. Abbildung 1). Der Prozess wird auf das Planen eines Experiments bezogen: (1) Das Lernziel ist die Anwendung der Variablenkontrollstrategie bei der Planung eines Experiments. (2) Die schriftliche und individuelle Planung wird als geeignete Erhebungsmethode für das Lernziel ausgewählt. (3) Die Lernenden verschriftlichen ihre individuelle Planung zu einer vorgegebenen Fragestellung entsprechend ihrer Vermutungen. (4) Für die Bewertung der Planungen wird ein Schema erstellt, mit dem (5) die zentralen Aspekte der Variablenkontrollstrategie analysiert werden können. (6) Basierend auf dieser Bewertung wird die Qualität der Anwendung der Variablenkontrollstrategie bestimmt und (7) eine standardisierte Rückmeldung für zukünftige Planungen gegeben. In Ergänzung zum Prozess des formativen Assessments (Ropohl & von Aufschnaiter, 2021) erfolgt am Ende des Prozesses zusätzlich eine Reflexion der erhaltenen Rückmeldungen durch die Lernenden, um die Verarbeitungstiefe zu erhöhen.

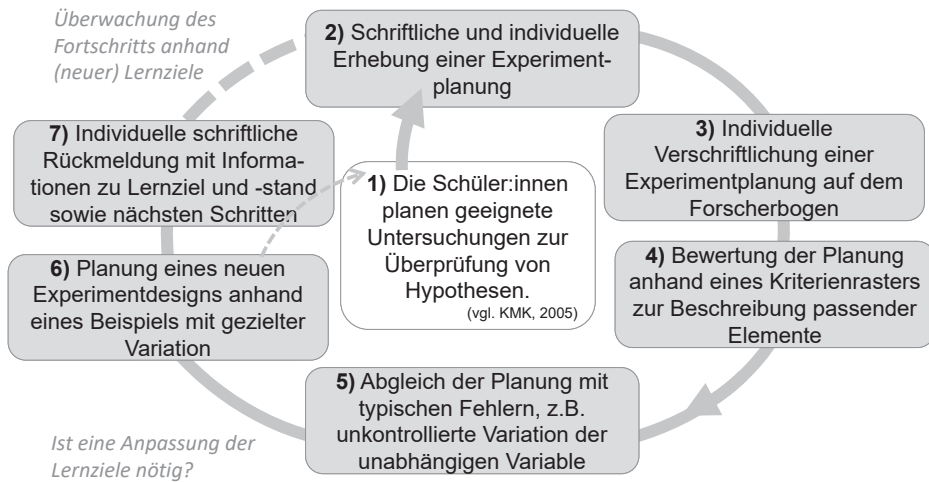


Abbildung 1: Prozess formativen Assessments am Beispiel des Planens eines Experiments nach Scheuermann (2017).

4 Wirkung formativen Assessments beim Experimentieren

Inwiefern formatives Assessment den Aufbau von Kompetenzen beim Experimentieren unterstützt, wurde empirisch untersucht. Im Fokus standen insbesondere Rückmeldungen an die Lernenden als im Unterricht sichtbarstes Element des Prozesses formativen Assessments (Hild et al., 2020; Ropohl & Scheuermann, 2018; Wollenschläger et al., 2011, 2012). Gemeinsam ist den folgenden drei Untersuchungen, dass jeweils unterschiedliche Rückmeldearten kontrastiert wurden.

Wollenschläger et al. (2011) kommen zu dem Schluss, dass Rückmeldungen, die v.a. die Kompetenzen der Lernenden adressieren, wirksamer sind als solche ohne den Bezug zu erreichten Stufen eines Kompetenzmodells. Der Befund kann als Beleg für die Wirkung von Rückmeldungen gesehen werden, die gezielt Informationen über die weitere Bearbeitung von kompetenzorientierten Aufgaben enthalten. In gewissem Widerspruch zu diesem Ergebnis stehen die Befunde von Hild et al. (2020), die belegen, dass Lernende vor allem von einer Rückmeldung des aktuellen Leistungsstandes profitieren und weniger von einer Kombination von Rückmeldungen zum aktuellen Leistungsstand und nächsten Schritten. Insgesamt waren die Unterschiede zwischen den Wirkungen der Interventionsbedingungen jedoch gering. Ropohl und Scheuermann (2018) widerlegen diesen Befund wiederum, da in ihrer Studie Lernende deutlich besser abschnitten, die in ihren Rückmeldungen als Komponente des Prozesses formativen Assessments (s. Abbildung 1) Informationen zum Lernziel, zum Lernstand und zu nächsten Schritten im Lernprozess erhielten (sogenannte lösungsprozessbezogene Rückmeldungen).

Die Befunde legen nahe, dass die unterschiedlichen Wirkungen aus den jeweiligen Untersuchungsbedingungen resultieren und diese bei Analysen als Mediator- oder Moderatorvariablen zu berücksichtigen wären. Rückmeldungen wirken demnach stark differenziell, d. h. nicht eine spezifische Rückmeldeform ist anderen Formen grundsätzlich überlegen. Denkbar ist, dass die Menge der rückgemeldeten Informationen die kognitive Belastung während des Lernens beeinflusst.

5 Kognitive Belastung bei Öffnungsgraden und Rückmeldungen

Hinter jeder Form von Lernunterstützung steht die positive Intention Schülerinnen und Schülern das Lernen zu erleichtern. Dies gilt für Öffnungsgrade wie auch Rückmeldungen. Gleichzeitig können solche Unterstützungsformate die kognitive Belastung (Cognitive Load; Sweller, 1988) für die Lernenden ungewollt erhöhen (Leutner et al., 2014), wenn sie als Gestaltungselement einer Lerngelegenheit hinzugefügt werden. Die kognitive Belastung ist ein Maß, wie stark die Kapazitäten des Arbeitsgedächtnisses während des Lernens beansprucht werden. Unterschieden werden drei Formen der Belastung (Sweller et al., 1998). Der Intrinsic Cognitive Load (ICL) betrifft die Komplexität des Lerninhaltes an sich, welche vom Vorwissen der Lernenden abhängt. Der Extraneous Cognitive Load (ECL) bezieht sich auf die kognitive Belastung, welche primär durch Merkmale des Lernmaterials und der Lernumgebung verursacht wird (Öffnungsgrade und Rückmeldungen). Die nach Berücksichtigung von ICL und ECL noch verbleibenden Kapazitäten des Arbeitsgedächtnisses können für kognitive Prozesse im Sinne des Germane Cognitive Load (GCL) aufgewendet werden, d. h. Schema-konstruktion, Elaboration und Verständnis.

In Bezug auf das stark lernendengelenkte forschende Lernen wird kritisiert (Kirschner et al., 2006), dass das Lernen in einer solchen Lerngelegenheit mit einer hohen kognitiven Belastung verbunden sei. Diese führe dazu, dass Schülerinnen und Schüler einen geringeren Lernzuwachs erfahren würden, insbesondere wenn sie wenig Wissen über Arbeitsweisen haben. Wird in solch einem Fall von hohem ICL eine instruktionale Unterstützung hinzugefügt, z. B. in Form von Rückmeldungen, kann dies den ECL erhöhen, da zusätzliche Kapazitäten aufgewendet werden müssen, um das Unterstützungsformat zu verstehen. Dies kann dazu führen, dass das Arbeitsgedächtnis überlastet wird und der Lernprozess abbricht (Hmelo-Silver et al., 2007). Mit Blick auf die berichteten Untersuchungen zur Wirksamkeit von Rückmeldungen, muss daher angenommen werden, dass der ECL für die Lernenden je nach Voraussetzungen besonders hoch ist. Die Intention der Rückmeldungen ist allerdings dem ECL entgegenzuwirken und diesen sogar zu senken. In einer Untersuchung, die auf der Untersuchung von Ropohl und Scheuermann (2018) aufbaut und nachfolgend dargestellt wird, wurde die Rolle des ECL beim Geben von Rückmeldungen zum Planen von Experimenten untersucht.

6 Ziel und Fragestellungen der Untersuchung

Das Ziel ist die Untersuchung der Lernwirksamkeit unterschiedlicher Öffnungsgrade beim Experimentieren mit zusätzlicher Unterstützung durch lösungsprozessbezogene Rückmeldungen sowie die Untersuchung der vermittelnden Funktion des ECL. Kontrastiert werden das lernendengelenkte mit dem lehrendengelenkten und dem geführten forschenden Lernen. Die folgenden Forschungsfragen werden dazu aufgestellt:

- (1) Unterscheidet sich die Lernwirksamkeit der Öffnungsgrade beim Experimentieren inkl. Rückmeldungen hinsichtlich der Planung von Experimenten?
- (2) Mediiert die kognitive Belastung den Einfluss der verschiedenen Öffnungsgrade des forschenden Lernens auf den Lernzuwachs?

7 Methodisches Vorgehen

In einer Interventionsstudie wurden die Öffnungsgrade als unabhängige Variable und die Kompetenz der Probandinnen und Probanden, ein Experiment zu planen, als abhängige Variable konzeptualisiert. Die Öffnungsgrade unterschieden sich hinsichtlich der Vorgabe der Fragestellung und der Hypothese. Die lösungsprozessbezogenen Rückmeldungen wurden lernbegleitend im Sinne formativen Assessments in allen Interventionsgruppen gleich realisiert (vgl. Ropohl & Scheuermann, 2018; Abbildung 1). Die kognitive Belastung wird als Mediatorvariable in die Analysen einbezogen.

7.1 Stichprobe und Datenerhebung

An der dreistündigen Interventionsstudie nahmen nach Ausschluss von Probandinnen und Probanden, die nicht an allen Terminen anwesend waren, $N = 266$ Schülerinnen und Schüler ($M_{\text{Alter}} = 13.61$; $SD = 0.56$; 55,6 % weiblich) aus 13 Klassen der 8. Jahrgangsstufe an sieben Gymnasien in Schleswig-Holstein teil. Die Klassen unterschieden sich zum Prä-Messzeitpunkt hinsichtlich ihres Fachwissens zum Thema Metalle, $F(12, 294) = 5.26$, $p < .001$, $\eta^2 = .18$, und ihres Wissens zu naturwissenschaftlichen Arbeitsweisen, $F(12, 289) = 1.84$, $p = .041$, $\eta^2 = .07$. Um gleiche Ausgangsbedingungen in den Interventionsgruppen zu garantieren, wurden die Probandinnen und Probanden nach dem Prä-Messzeitpunkt klassenweise den drei Interventionsbedingungen balanciert zugeordnet. Nach der Zuordnung unterschieden sich die drei Gruppen weder hinsichtlich des Fachwissens zum Thema Metalle, $F(2, 304) = 0.95$, $p = .387$, $\eta^2 = .01$, noch hinsichtlich des Wissens zu naturwissenschaftlichen Arbeitsweisen, $F(2, 299) = 0.77$, $p = .047$, $\eta^2 = .01$.

Die drei Interventionsgruppen repräsentieren jeweils einen anderen Öffnungsgrad (Tabelle 1). In der Experimentalgruppe wurden den Lernenden die Fragestellung und die Hypothese vorgegeben (lehrendengelenkt). In der Vergleichsgruppe 1 wurde den Lernenden lediglich die Fragestellung vorgegeben (geführt). In der Vergleichsgruppe 2 haben die Lernenden alle drei Aktivitäten selbst durchgeführt (lernendengelenkt). Die

drei Gruppen haben zuvor ein Demonstrationsexperiment präsentiert bekommen, welches einen kognitiven Konflikt und damit die das Experiment leitende Frage auslösen sollte. Die Schülerinnen und Schüler aller Gruppen haben ein Experiment geplant, dieses aber nicht durchgeführt.

Tabelle 1: Bedingungen und Anzahl der Schüler:innen in den Interventionsgruppen

Schüler:innen-aktivitäten	Interventionsgruppen		
	Experimentalgruppe [n = 77]	Vergleichsgruppe 1 [n = 86]	Vergleichsgruppe 2 [n = 103]
Formulieren von Fragestellungen	vorgegeben	vorgegeben	Formulieren von Fragestellungen
Aufstellen von Hypothesen	vorgegeben	Aufstellen von Hypothesen	Aufstellen von Hypothesen
Planen eines Experiments	Planen eines Experiments	Planen eines Experiments	Planen eines Experiments

In einer ersten Unterrichtsstunde wurden die Merkmale der Experimentplanung inklusive der Umsetzung der Variablenkontrollstrategie als Lernziel erarbeitet (Aktivität 1 in Abbildung 1). In den beiden folgenden Interventionsstunden planten die Lernenden jeweils ein Experiment unter Anwendung der Variablenkontrollstrategie und verschriftlichten ihre Planungen auf Forscherbögen (Aktivitäten 2 und 3 in Abbildung 1). Es wurde jeweils ein anderer Aspekt der Reaktion von Metallen mit Sauerstoff thematisiert. Die Experimentplanungen wurden anhand eines Kriterienrasters im Hinblick auf die Vollständigkeit der Merkmale der Experimentplanung bewertet (Aktivität 4 in Abbildung 1; dabei Berücksichtigung der Aktivitäten 5 und 6; s. Unterkapitel 7.2). Basierend auf der Bewertung wurde für alle Lernenden eine individuelle Rückmeldung mit den Informationen zum Lernziel, Lernstand und nächsten Lernschritten standardisiert formuliert und in der jeweils nachfolgenden Stunde auf einem Rückmeldebogen ausgehändigt (Aktivität 7 in Abbildung 1). Auf die Reflexion der Rückmeldungen mit Hilfe eines Reflexionsbogens folgte jeweils die Planung eines weiteren Experiments, bei dem die Rückmeldung durch die Lernenden berücksichtigt werden sollte (die Reflexion zur zweiten Experimentplanung erfolgte zum Post-Messzeitpunkt). Auf diese Weise wurde der Prozess formativen Assessments zweimal durchlaufen (Abbildung 1).

Um sicherzustellen, dass die Durchführung der Intervention dem beschriebenen Vorgehen genügt, wurden die beiden Interventionsstunden von den Forschenden zur Kontrolle beobachtet und protokolliert. Die Reflexion der Rückmeldungen durch die Lernenden diente ebenfalls der Sicherstellung der Interventionsbedingungen. Durch die Reflexion wird garantiert, dass die Probandinnen und Probanden die Rückmeldeinformation zu jedem Merkmal der Experimentplanung gelesen und verarbeitet haben. Zudem wurde die Intervention von den Forschenden basierend auf einem Manual standardisiert durchgeführt. Für jeden Fehler, den die Lernenden machen konnten, gab es

einen vorformulierten Textbaustein. Die Bausteine wurden entsprechend der Fehler, die die Lernenden gemacht haben, zu einer Gesamtrückmeldung zusammengefügt und auf das jeweilige inhaltliche Beispiel angepasst. Wenn beispielsweise Lernende das Konstanthalten von Störvariablen richtig geplant haben, erhielten sie die Rückmeldung: „Mit deinem geplanten Experiment berücksichtigst du mehr als zwei Störvariablen, indem du a, b und c konstant hältst.“ Geling dieser Planungsschritt nicht, wurde die folgende Rückmeldung gegeben: „Berücksichtige mehr als zwei Störvariablen, z. B. indem a, b oder c konstant gehalten werden.“

7.2 Instrumente

Als abhängige Variable wurde die Kompetenz der Lernenden, ein Experiment zu planen, zum Prä-, Post- und Follow-up-Messzeitpunkt erhoben, die unmittelbar vor und nach bzw. 12 Wochen nach den drei Unterrichtsstunden der Intervention lagen. Die resultierenden drei Experimentplanungen je Lernender oder Lernendem wurden anhand des gleichen Kriterienrasters von den Forschenden kodiert, das auch für die beiden Experimentplanungen aus den beiden Interventionsstunden verwendet wurde. Zehn Merkmale einer Experimentplanung unter Anwendung der Variablenkontrollstrategie wurden jeweils auf einer dreistufigen Skala (0 = nicht erfüllt; 1 = teilweise erfüllt; 2 = voll erfüllt) bewertet. Die Güte der Kodierung wurde durch eine Doppelkodierung von 10 % der schriftlichen Experimentplanungen mit einem zweiten hypothesenblinden Rater sichergestellt. Die Interrater-Reliabilität Kendalls Tau τ liegt zu allen Messzeitpunkten in einem guten bis sehr guten Bereich zwischen $\tau = .68$ und $\tau = 1.00$ (Wirtz & Caspar, 2002).

Zur Berechnung der Reliabilität wurden die Merkmale der Experimentplanung als Items einer offenen Aufgabe betrachtet (Tabelle 2). Die Reliabilitäten aus Post- und Follow-Up-Testung sind für Gruppenvergleiche ausreichend (Lienert & Raatz, 1998). Die geringe Reliabilität zum Prä-Messzeitpunkt kann dahingehend erklärt werden, dass die Lernenden die Merkmale der Experimentplanung zu diesem Zeitpunkt nicht kannten und daher nicht systematisch anwenden konnten.

Tabelle 2: Reliabilität des Instruments zur Kompetenz der Planung eines Experiments

Variable	Reliabilität [α]		
	Prä	Post	Follow-up
Planen von Experimenten	.30	.66	.66

Zur Messung des ECL als Mediatorvariable wurden die Lernenden in Abhängigkeit von der Gruppenbedingung schriftlich gefragt, wie anstrengend und schwierig sie das Experimentieren wahrgenommen haben. Für die Erhebung des ECL wurden zwei etablierte geschlossene Items mit 7-stufiger Rating-Skala verwendet, die die wahrgenommene Aufgabenschwierigkeit sowie die investierte mentale Anstrengung von *sehr leicht* bis *sehr schwer* erfragen (Kalyuga et al., 1999; Paas, 1992). Die Items wurden sowohl di-

rekt nach dem Formulieren von Fragestellungen und dem Aufstellen von Hypothesen als auch nach dem Planen von Experimenten eingesetzt. Es liegen gute bis sehr gute Reliabilitäten für die Skala des ECL vor (Tabelle 3).

Tabelle 3: Reliabilität der Skala zur Erhebung des ECL

Variable	Reliabilität [α]	
	1. Interventionsstunde	2. Interventionsstunde
Formulieren von Fragestellungen	.79	.92
Aufstellen von Hypothesen	.91	.92
Planen von Experimenten	.87	.85

7.3 Statistische Analysen

Die Wirksamkeit der unterschiedlichen Interventionsbedingungen auf das Planen von Experimenten unter Anwendung der Variablenkontrollstrategie wurde mit einfaktoriellen Varianzanalysen mit Messwiederholung überprüft und hierzu Experimentplanungen von allen drei Messzeitpunkten herangezogen. Mithilfe der einfaktoriellen Varianzanalyse mit Messwiederholung wurde der Haupteffekt – die Veränderung über die Messzeitpunkte – und der Interaktionseffekt – die Veränderung zwischen den Interventionsbedingungen über die Messzeitpunkte – geprüft. Mittels Bonferroni-korrigierter Post-hoc-Tests wurde die Effektstärke für die Veränderung innerhalb der Interventionsbedingungen über die Messzeitpunkte bestimmt. Als Maß für die Effektstärke der Varianzanalysen wird η^2 herangezogen (kleiner Effekt: $\eta^2 = .01$, mittlerer Effekt: $\eta^2 = .06$, großer Effekt: $\eta^2 = .14$; Cohen, 1988).

Zur Analyse des potentiellen Mediatoreffekts der wahrgenommenen kognitiven Belastung wurde eine Pfadanalyse durchgeführt und dabei die Bootstrapping-Methode angewandt. Auf diese Weise können Kausalbeziehungen zwischen den im Modell enthaltenen Variablen untersucht werden. Der Lernzuwachs vom Prä- zum Post-Messzeitpunkt ist die abhängige Variable (Y), der Öffnungsgrad des forschenden Lernens die multikategoriale unabhängige Variable (Dummy-Kodierung der drei unterschiedlichen Öffnungsgrade; D_1 : Experimentalgruppe = 1, Vergleichsgruppe 1 = 0, Vergleichsgruppe 2 = 0; D_2 : Experimentalgruppe = 0, Vergleichsgruppe 1 = 1, Vergleichsgruppe 2 = 0) und die wahrgenommene kognitive Belastung die Mediatorvariable (M). Aufgrund empirisch ermittelter Korrelationen zwischen den Modellvariablen können über Strukturgleichungen die Pfadkoeffizienten (a, b oder c) geschätzt werden. Die Größe eines Pfadkoeffizienten entspricht jenem Teil der Standardabweichung einer Variablen, der durch eine kausal vorgeordnete Variable determiniert ist, wenn alle anderen kausal vorgeordneten Modellvariablen konstant gehalten werden. Das Bootstrapping wird zur Analyse eines Mediatoreffekts empfohlen, da es keine Normalverteilung des indirekten Effekts voraussetzt. Diese wird durch Replikation des Datensatzes generiert (10.000-fache Replikation des Datensatzes; 98,75 % Konfidenzintervall).

8 Ergebnisse zur Wirkung der Öffnungsgrade und zur Rolle des ECL

Zum Prä-Messzeitpunkt weisen die drei Interventionsgruppen die gleichen Ausgangsbedingungen im Planen von Experimenten auf, $F(2, 314) = 2.50, p = .084, \eta^2 = .02$. Vom Prä- zum Post-Messzeitpunkt liegt ein Lernzuwachs im Planen von Experimenten vor, $F_{\text{Prä-Post}}(1, 263) = 497.21, p < .001, \eta^2 = .65$. Dieser Lernzuwachs ist in Abhängigkeit von der Gruppenbedingung unterschiedlich stark, $F_{\text{Prä-Post}}(2, 263) = 4.50, p = .011, \eta^2 = .03$. Die Bonferroni-korrigierten Post-hoc-Tests zeigen, dass sich alle drei Interventionsbedingungen hinsichtlich der mittleren Differenz zwischen Prä- und Post-Messzeitpunkt jeweils signifikant voneinander unterscheiden: Experimentalgruppe ($M = 0.28, SD = 0.02$), Vergleichsgruppe I ($M = 0.24, SD = 0.02$) und Vergleichsgruppe II ($M = 0.23, SD = 0.02$; Abbildung 2). Die Experimentalgruppe weist im Vergleich zu den beiden anderen Gruppen den stärksten Zuwachs auf.

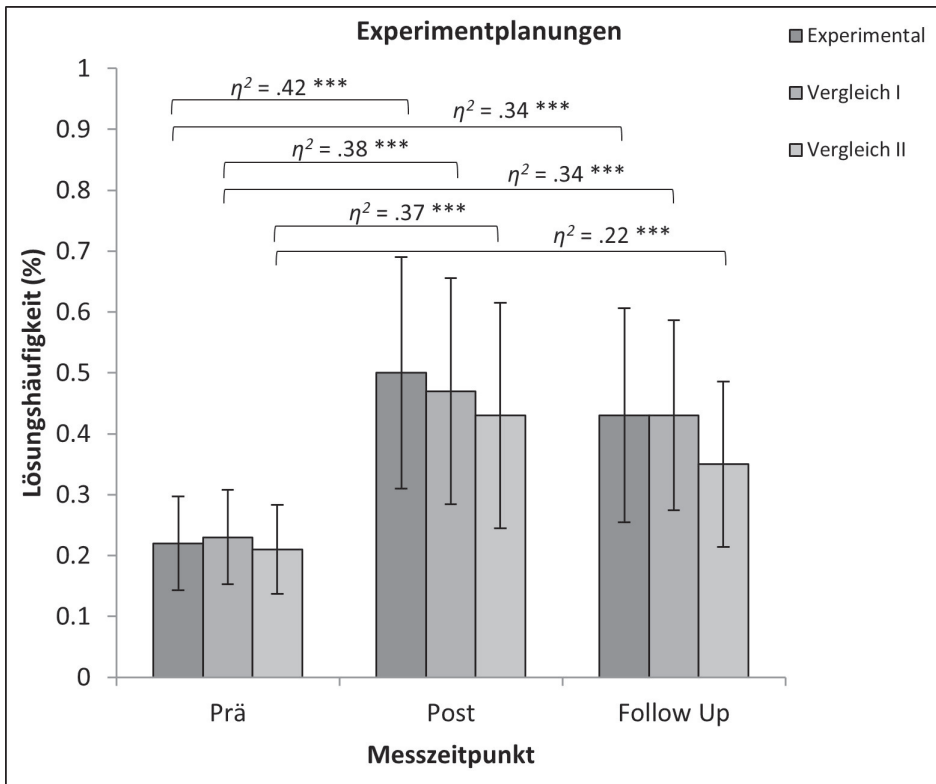


Abbildung 2: Ergebnisse der Experimentplanungen für die Interventionsgruppen zu den drei Messzeitpunkten.

Anmerkung: Die Effektstärken η^2 geben die Effekte aus Bonferroni Post-hoc-Tests für den Prä- und Post-Messzeitpunkt wieder. *** $p < .001$.

Auch für den Gesamtzeitraum vom Prä- zum Follow-up-Messzeitpunkt liegt eine signifikante Veränderung hinsichtlich des Planens von Experimenten vor, $F_{\text{Prä-Follow Up}}(1, 290) = 378.76, p < .001, \eta^2 = .57$. Zwischen den Gruppen liegen dabei Unterschiede vor, $F_{\text{Prä-Follow Up}}(2, 290) = 6.93, p = .001, \eta^2 = .05$. Die Bonferronikorrigierten Post-hoc-Tests zeigen, dass sich alle drei Interventionsbedingungen hinsichtlich der mittleren Differenz zwischen Prä- und Follow up-Messzeitpunkt jeweils signifikant voneinander unterscheiden: Experimentalgruppe ($M = 0.21, SD = 0.02, \eta^2 = .34$), Vergleichsgruppe I ($M = 0.20, SD = 0.02, \eta^2 = .34$) und Vergleichsgruppe II ($M = 0.13, SD = 0.02, \eta^2 = .22$).

Die Pfadanalyse zeigt, dass sich die drei Interventionsgruppen in der Wahrnehmung des ECL nicht signifikant unterscheiden ($a_1 = -0.12, p = .498; a_2 = 0.14, p = .461$; Abbildung 3). Bei Betrachtung der Mittelwerte zeigt sich lediglich die Tendenz, dass die Vergleichsgruppe I einen höheren ECL wahrnimmt als die Experimental- und Vergleichsgruppe II. Ferner liegt kein signifikanter direkter Effekt von ECL auf den Lernzuwachs in den Experimentplanungen vor ($b = 0.26, p = .234$). So ergibt das Bootstrapping keine hinreichende Evidenz für eine Vermittlung des ECL bezüglich des Effekts des jeweiligen Öffnungsgrades auf den Lernzuwachs in der Experimentplanung ($D_1: a_1 b = -0.03, p = .694, \text{Bootstrap-KI (98,75 \%)} = [-0.234, 0.100]$; $D_2: a_2 b = 0.04, p = .300, \text{Bootstrap-KI (98,75 \%)} = [-0.112, 0.254]$).

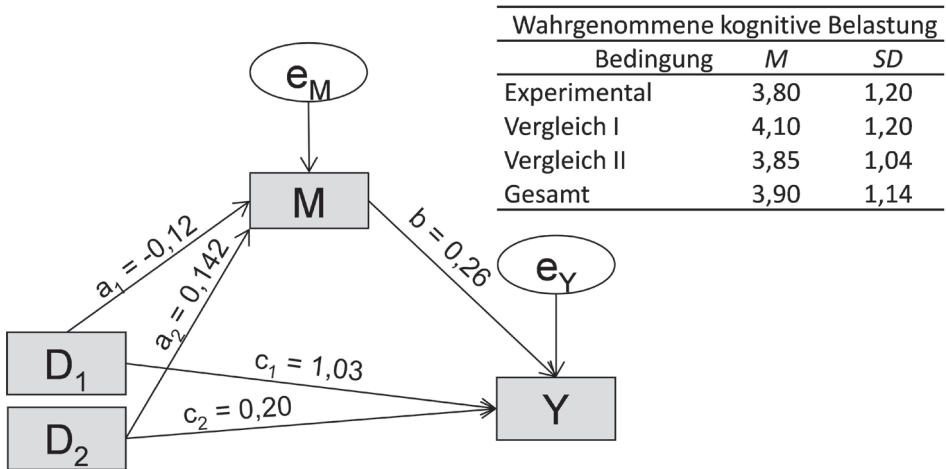


Abbildung 3: Mediatormodell für den ECL

Anmerkung. D_1 & D_2 = Dummy-kodierte Öffnungsgrade des forschenden Lernens; M = wahrgenommene kognitive Belastung; Y = Lernzuwachs vom Prä- zum Post-Messzeitpunkt; a, b & c = Pfadkoeffizienten.

9 Diskussion der Ergebnisse

Für die erste Forschungsfrage kann geschlussfolgert werden, dass die Lernenden der Experimentalgruppe, denen Fragestellungen und Hypothesen vorgegeben wurden, den größten Lernzuwachs beim Planen von Experimenten erfahren haben. Das Experimentieren bei einem stark lehrendengelenkten forschenden Lernen ist in der Untersuchung den beiden anderen Öffnungsgraden überlegen. Die Lernenden aus den Vergleichsgruppen I und II, die in geführten bzw. stärker geöffneten Lernsituationen Experimente geplant haben, zeigen einen geringeren Lernzuwachs. Dieser Effekt besteht auch nach drei Monaten. Insgesamt kann festgehalten werden, dass dieselbe Rückmeldeform in Abhängigkeit vom Öffnungsgrad unterschiedlich gut wirkt.

Die Ergebnisse stützen die Kritik von Kirschner et al. (2006) am forschenden Lernen, da die höhere Verantwortung der Lernenden für die Planung des Erkenntnisgewinnungsprozesses mit einem niedrigeren Lernzuwachs einhergeht. Je mehr Planungsentscheidungen Lernende treffen müssen, desto mehr Fehler machen sie. Dies gilt offenbar auch in einer Lernumgebung, in der die Schülerinnen und Schüler eine Unterstützung in Form von Rückmeldungen erhalten. Hartinger et al. (2013) empfehlen daher den Bearbeitungsanteil von Lernenden am Planungsprozess sukzessive und entsprechend ihrer Voraussetzungen zu erhöhen. Einen anderen Ansatz haben Wahser und Sumfleth (2008) vorgeschlagen und untersucht: sogenannte Strukturierungshilfen. Diese Hilfen leiten Schülerinnen und Schüler bereits während der Planung und Durchführung von Erkenntnisgewinnungsprozessen an, was mit Blick auf die Wirkung der Hilfen ein Vorteil ist (Künsting et al., 2008). Schriftliche Rückmeldungen haben demgegenüber einen Nachteil: die Rückmeldeinformation kann nicht direkt bei der Planung berücksichtigt werden, sondern erst danach.

Mit Blick auf die zweite Forschungsfrage zeigt sich, dass die Schülerinnen und Schüler beim Lernen in den unterschiedlichen Öffnungsgraden einen ähnlich hohen ECL wahrnehmen. Die Gruppe, die eine eigene Hypothese hinsichtlich einer vorgegebenen Fragestellung aufgestellt hat, zeigt die Tendenz zu einem höheren ECL als die übrigen Lernenden. Es liegt aber weder ein direkter noch ein indirekter Effekt des ECL auf den Lernzuwachs im Planen von Experimenten vor. Der Effekt des jeweiligen Öffnungsgrades des forschenden Lernens auf den Lernzuwachs wird folglich nicht über den wahrgenommenen ECL mediiert.

Dass die unterschiedliche Wirkung der Rückmeldungen auf den Lernzuwachs der Schülerinnen und Schüler in der vorliegenden Untersuchung nicht auf die kognitive Belastung der Lernenden zurückgeführt werden konnte, ist nicht erwartungskonform, da eine offenere Lerngelegenheit zum Experimentieren eigentlich den ECL verringern müsste. Wirth et al. (2009) konnten zeigen, dass ein höherer Öffnungsgrad, der mit einem unspezifischen Lernziel einhergeht, den ECL verringert. In diesem Fall können die Lernenden ihre eigenen Lernstrategien abrufen und ausreichend kognitive Ressourcen für den Wissenserwerb nutzen. Kirschner et al. (2006) haben wiederum herausgestellt, dass gerade ein hoher Öffnungsgrad bei vielen Lernenden zu einem

hohen ECL führt. Zu beachten ist, dass in der vorliegenden Untersuchung neben dem Öffnungsgrad Rückmeldungen als Unterstützungsformat implementiert wurden und dadurch die Komplexität der Lerngelegenheit für die Schülerinnen und Schüler erhöht wurde; und somit potentiell auch der ECL (Leutner et al., 2014). Erhalten die Lernenden beispielsweise während des Experimentierens Rückmeldungen, müssen sowohl die Informationen der Lehr-Lern-Materialien als auch die Rückmeldeinformationen von ihnen verarbeitet und aufeinander bezogen werden. Dies kann zum einen zu kognitiver Belastung – also zu einem erhöhten ECL – und darüber vermittelt zu einer verringerten Motivation führen (Leutner et al., 2014).

10 Limitationen der Untersuchung

Die dargestellte Untersuchung ist eine quasiexperimentelle Laboruntersuchung: Die Umgebung ist laborartig gestaltet und die Durchführung erfolgt in natürlichen Gruppen. Aufgrund dieser Umstände ist die generelle Gültigkeit der Untersuchungsbefunde im Vergleich zu anderen Untersuchungsvarianten eingeschränkt (z. B. Bortz & Döring, 2006). An zwei Aspekten kann verdeutlicht werden, inwiefern die saubere Untersuchungskonzeption und die Übertragbarkeit sich gegenseitig bedingen und wie beide Thematiken für die vorliegende Untersuchung einzuschätzen sind: (1) In der Untersuchung wurde die Kompetenz des Planens von Experimenten und der Einfluss von Rückmeldungen auf diese Kompetenz in den Blick genommen. Aufgrund dieser Fokussierung wurden angrenzende Kompetenzen, wie die Durchführung eines selbstgeplanten Experiments, nicht in die Untersuchung einbezogen. Im regulären Chemieunterricht ist jedoch wahrscheinlich, dass Schülerinnen und Schüler ein selbstgeplantes Experiment auch durchführen, dabei gleichzeitig Optimierungen an der ursprünglichen Planung vornehmen und damit der Weg der Erkenntnisgewinnung in Teilen iterativ umgesetzt wird. Damit lassen sich die beobachteten Effekte nicht uneingeschränkt und ohne weitere Untersuchungen auf Lerngelegenheiten im Unterricht übertragen. (2) Die laborähnlichen Bedingungen der Untersuchungskonzeption garantieren die weitgehende Ausschaltung bzw. Kontrolle möglicher Störvariablen, die die abhängige Variable beeinflussen könnten. Dazu zählen das Experimentieren der Lernenden in Einzelarbeit, wodurch Effekte der sozialen Interaktion in Kleingruppen auf den Lernprozess ausgeschlossen werden, und das Umsetzen der Intervention mithilfe eines Testleiters bzw. einer Testleiterin, wodurch Effekte abweichender Rückmeldungen durch die Lehrkraft verhindert werden. Damit können Veränderungen der abhängigen Variable mit sehr hoher Wahrscheinlichkeit ursächlich auf die unabhängige Variable, die unterschiedlichen Rückmeldeformen, zurückgeführt werden. Dies ist jedoch zu einem gewissen Grad mit einer geringeren Übertragbarkeit der Ergebnisse verbunden.

11 Desiderata und Ausblick

Bei Betrachtung der bisherigen Forschungsergebnisse bleibt die Aufklärung der Rolle der kognitiven Belastung beim Experimentieren mit formativem Assessment offen. Es kann angenommen werden, dass Lernende mit unterschiedlichen Voraussetzungen in Abhängigkeit vom Öffnungsgrad eine hohe kognitive Belastung erfahren. So können sie z. B. beim lernendengelenkten forschenden Lernen überfordert werden, wenn sie wenige Vorerfahrungen haben oder im lehrendengelenkten forschenden Lernen überfordert sein, wenn sie viele Vorerfahrungen haben und die Anweisungen mit den entwickelten Wissensschemata kollidieren. Folglich gilt es zu prüfen, ob Rückmeldungen, die sich hinsichtlich der enthaltenen Informationen unterscheiden und damit inhaltlich den Voraussetzungen der Schülerinnen und Schüler angepasst werden, die kognitive Belastung reduzieren können. Ein Desiderat ist damit die Untersuchung des Zusammenwirkens zwischen dem Vorwissen der Lernenden (aptitude) und verschiedenen Arten der Rückmeldung (treatment).

In der berichteten Interventionsstudie ist die Motivation der Lernenden nicht gezielt untersucht worden. Wollenschläger et al. (2011) legen jedoch einen Einfluss von Rückmeldungen auf motivationale Aspekte nahe. Im Hochschulbereich zeigten Pitt und Norton (2016), dass positive Rückmeldungen die intrinsische Motivation verstärken, negative diese jedoch verringern. Ob sich ähnliche Effekte bei Schülerinnen und Schülern bei der spezifischen Arbeitsweise des Experimentierens zeigen, ist fraglich. Ein zweites Desiderat ist daher die Untersuchung des Einflusses unterschiedlicher Ausgestaltungen einer Lerngelegenheit zum Experimentieren mit formativem Assessment auf die Motivation der Lernenden.

Literatur

- Abrams, E., Southerland, S. A. & Evans, C. (2008). Introduction: Inquiry in the Classroom: Identifying necessary components of a useful definition. In E. Abrams, S. A. Southerland & P. Silva (Hrsg.), *Integrating inquiry in the classroom: Realities and opportunities* (S. xi–xlii). Informations Age Inc.
- Arnold, J., Kremer, K. & Mayer, J. (2014). Understanding students' experiments—What kind of support do they need in inquiry tasks? *International Journal of Science Education*, 36(16), 2719–2749. <http://dx.doi.org/10.1080/09500693.2014.930209>
- Arnold, J., Kremer, K. & Mayer, J. (2017). Scaffolding beim Forschenden Lernen – Eine empirische Untersuchung zur Wirkung von Lernunterstützungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 23, 21–37. <https://doi.org/10.1007/s40573-016-0053-0>
- Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A. & Granger, E. (2010). Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education*, 94(4), 577–616. <https://doi.org/10.1002/sce.20390>
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Springer. <https://doi.org/10.1007/978-3-540-33306-7>

- Chen, Z. & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70, 1098–1120. <https://doi.org/10.1111/1467-8624.00081>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Dean, D. & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, 91(3), 384–397. <https://doi.org/10.1002/sce.20194>
- Emden, M., Bewersdorff, A. & Baur, A. (2019). Kann Experimentieren in der Schule bilden? Ein Beitrag zur bildungstheoretischen Legitimation eines selbstverständlichen Gegenstandes des Naturwissenschaftsunterrichts. *Zeitschrift für Pädagogik*, 65(5), 710–729. <https://doi.org/10.25656/01:24115>
- Furtak, E. M., Seidel, T., Iverson, H. & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82(3), 300–329. <https://doi.org/10.3102/0034654312457206>
- Harlen, W. (2013). *Assessment & inquiry-based science education: issues in policy and practice*. Trieste, Italy: Global Network of Science Academies (IAP) Science Education Programme (SEP).
- Hartinger, A., Grygier, P., Tretter, T. & Ziegler, F. (2013). *Lernumgebungen zum naturwissenschaftlichen Experimentieren* (Modul für SINUS an Grundschulen). IPN.
- Hild, P., Buff, A., Gut, C. & Parchmann, I. (2020). Adaptives kompetenzbezogenes Feedback beim selbstständigen praktisch-naturwissenschaftlichen Arbeiten: Eine empirische Untersuchung zur Wirksamkeit unterschiedlicher Feedbackformen. *Zeitschrift für Didaktik der Naturwissenschaften*, 26, 19–35. <https://doi.org/10.1007/s40573-020-00109-8>
- Hmelo-Silver, C. E., Duncan, R. G. & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107. <https://doi.org/10.1080/00461520701263368>
- Kalyuga, S., Chandler, P. & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351–371. [https://doi.org/10.1002/\(SICI\)1099-0720\(199908\)13:4<351::AID-ACP589>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6)
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- KMK (Hrsg.) (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Luchterhand.
- Künsting, J., Thillmann, H., Wirth, J., Fischer, H. E. & Leutner, D. (2008). Strategisches Experimentieren im naturwissenschaftlichen Unterricht. *Psychologie in Erziehung und Unterricht*, 55(1), 1–15.
- Leutner, D., Opfermann, M. & Schmeck, A. (2014). Lernen mit Medien. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (S. 297–322). Beltz. <https://doi.org/10.1007/978-3-662-54368-9>
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Beltz.
- Meier, M. & Mayer, J. (2014). Selbständiges Experimentieren: Entwicklung und Einsatz eines anwendungsbezogenen Aufgabendesigns. *Mathematisch und naturwissenschaftlicher Unterricht (MNU)*, 67(1), 4–10.
- MBWK SH (Ministerium für Bildung, Wissenschaft und Kultur des Landes Schleswig-Holstein). (Hrsg.). (2019). *Fachanforderungen Chemie, Allgemeinbildende Schulen, Sekundarstufe I, Sekundarstufe II*. Verfügbar unter: <https://fachportal.lernnetz.de/files/>

- Fachanforderungen%20und%20Leitf%C3%A4den/Sek.%20I_II/Fachanforderungen/
Fachanforderungen_Chemie_f%C3%BCr_die_Sekundarstufe_I_II.pdf
- Minner, D. D., Levy, A. J. & Century, J. (2010). Inquiry-based science instruction – what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496. <https://doi.org/10.1002/tea.20347>
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>
- Pitt, E. & Norton, L. (2017). ‘Now that’s the feedback I want!’ Students’ reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*, 42(4), 499–516. <https://doi.org/10.1080/02602938.2016.1142500>
- Priemer, B. (2011). Was ist das Offene beim offenen Experimentieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 315–337.
- Rönnebeck, S., Bernholt, S. & Ropohl, M. (2016). Searching for a common ground – a literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.25656/01:12695>
- Ropohl, M. & Scheuermann, H. (2018). Welche Rückmeldungen wirken am besten? Ergebnisse einer empirischen Untersuchung von Rückmeldeformen beim Planen von Experimenten. *Zeitschrift für Didaktik der Naturwissenschaften*, 24, 151–165. <https://doi.org/10.1007/s40573-018-0080-0>
- Ropohl, M. & von Aufschnaiter, C. (2021). Formative assessment. In H.E. Fischer & R. Girwidz (Hrsg.), *Physics Education* (S. 413–437). Springer International. https://doi.org/10.1007/978-3-030-87391-2_15
- Scheuermann, H. (2017). *Entwicklung und Evaluation von Unterstützungsmaßnahmen zur Förderung der Variablenkontrollstrategie beim Planen von Experimenten*. Logos.
- Scheuermann, H. & Ropohl, M. (2017). Abhängige Variable, unabhängige Variable, Störvariable!? Die Einführung der Variablenkontrollstrategie. *Naturwissenschaften im Unterricht – Chemie*, 28(158), 19–23.
- Schulz, A., Wirtz, M. & Starauscek, E. (2012). Das Experiment in den Naturwissenschaften. In W. Rieß, M. Wirtz, B. Barzel & A. Schulz (Hrsg.), *Experimentieren im mathematisch-naturwissenschaftlichen Unterricht. Schüler lernen wissenschaftlich denken und arbeiten* (S. 15–38). Waxmann.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Sweller, J., van Merriënboer, J. & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Wahser, I. & Sumfleth, E. (2008). Training experimenteller Arbeitsweisen zur Unterstützung kooperativer Kleingruppenarbeit im Fach Chemie. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 219–241.
- Walpuski, M. & Sumfleth, E. (2007). Strukturierungshilfen und Feedback zur Unterstützung experimenteller Kleingruppenarbeit im Chemieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 181–198.
- Weirich, S., Becker, B. & Holtmann, M. (2019). Kompetenzstufenbesetzungen in den naturwissenschaftlichen Fächern. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 169–200). Waxmann.

- Wirth, J., Künsting, J. & Leutner, D. (2009). The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior*, 25(2), 299–305. <https://doi.org/10.1016/j.chb.2008.12.004>
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Hogrefe-Verlag.
- Wollenschläger, M., Möller, J. & Harms, U. (2011). Effekte kompetenzieller Rückmeldung beim wissenschaftlichen Denken. *Zeitschrift für Pädagogische Psychologie*, 25(3), 197–202. <https://doi.org/10.1024/1010-0652/a000040>
- Wollenschläger, M., Möller, J. & Harms, U. (2012). Ist kompetenzielles Fremdfeedback überlegen, weil es als effektiver wahrgenommen wird? *Unterrichtswissenschaft*, 40(3), 197–212.
- Zenner, C., Scheuermann, H. & Ropohl, M. (2017). Das Kupfer-Problem 2.0. Untersuchung der Schülervorstellungen zur Oxidation von Kupfer. *Chemie Konkret*, 24(5), 370–380. <https://doi.org/10.1002/ckon.201710311>

Formatives Assessment in Arbeitsphasen

Eine explorative Analyse der Lernunterstützung in Interaktionen zwischen der Lehrperson und Schüler:innen mit verschiedenen kognitiv-motivationalen Profilen

Merle Ruelmann^{1,2} | Urs Grob² | Alois Buholzer¹ | Sandra Zulliger¹

¹Pädagogische Hochschule Luzern | ²Universität Zürich

Abstract: Interaktionen zwischen Lehrenden und Lernenden in Arbeitsphasen sind zentral für die Lernunterstützung. Bisherige Studien untersuchten überwiegend, wie Lehrpersonen Strategien des formativen Assessments (z. B. Fragen, Feedback) einsetzen, um Lernende zu unterstützen. In der vorliegenden Untersuchung wurden ergänzend Schüler:innenfaktoren (verbale Äußerungen; kognitive und motivationale Voraussetzungen) bei der Analyse von Interaktionen einbezogen. Dazu wurden mathematische Arbeitsphasen von 52 vierten Primarschulklassen der Zentralschweiz mit Videocodierungen ausgewertet. Es konnten sechs Interaktionsmuster identifiziert werden, die sich hinsichtlich der Schüler:innenäußerungen und der Unterstützung durch die Lehrperson unterscheiden. Die Ergebnisse zeigen, dass Interaktionen der Lehrperson mit mehreren Lernenden häufiger dialogische Interaktionsmuster enthalten, Einzelinteraktionen jedoch mehrheitlich eng geführt werden. In Einzelinteraktionen unterschieden sich die Interaktionsmuster zudem in Abhängigkeit von den kognitiven und motivationalen Voraussetzungen der Schüler:innen. Die Befunde verdeutlichen, dass Schüler:innenfaktoren bei der Lernunterstützung konsequent einbezogen werden sollten.

1 Einleitung

Angebots-Nutzungs-Modelle des Unterrichts gehen davon aus, dass schulische Lernprozesse in einem Spannungsfeld von bereitgestellten Lerngelegenheiten und der aktiven Auseinandersetzung von Schüler:innen mit diesen Angeboten stattfinden (Helmke, 2014; Vieluf et al., 2020). Zugrunde liegt die (sozial-)konstruktivistische Annahme, dass neues Wissen (ko-)konstruktiv erworben wird und in diesem Prozess vielfältige Faktoren seitens der Lehrperson als auch der Lernenden Einfluss nehmen (Reusser & Pauli, 2010; Vieluf et al., 2020). Eine zentrale Bedeutung für die (ko-)konstruktive Unterstützung von Lernprozessen haben *Interaktionen* zwischen Lehrenden und Lernenden (bzw. *Lehrer-Schüler-Interaktionen*¹) im Unterricht (Lipowsky

1 In anderen Arbeiten wird teilweise der Begriff *Lehrer-Schüler-Interaktionen* verwendet (z. B. Jurik et al., 2015). Aus Gründen der gendergerechten Sprache verwenden wir in der vorliegenden Arbeit den Begriff *Interaktionen* oder synonym *Unterrichtsinteraktionen*.

et al., 2008; Pauli & Reusser, 2015). Darunter werden im Rahmen der vorliegenden Arbeit Dialoge der Lehrperson mit einzelnen, mehreren oder allen Lernenden während des Unterrichts verstanden. In solchen Interaktionen können Lehrpersonen die Lernprozesse von Schüler:innen unterstützen, indem sie mikroadaptiv zwischen dem Lerninhalt und dessen Nutzung durch Lernende vermitteln (Krammer et al., 2010; Krammer, 2016; Schnebel, 2017). Für die gezielte Unterstützung des Lernens im Unterricht eignen sich insbesondere Arbeitsphasen², in denen Schüler:innen weitestgehend selbstständig an Aufgaben arbeiten und die Lehrperson somit individuell auf sie eingehen kann (Krammer, 2009; Pohlmann-Rother et al., 2018). Studien zur Lernunterstützung in Arbeitsphasen lenkten den Blick bislang überwiegend auf das Handeln von Lehrpersonen (z. B. den Einsatz von Fragen und Feedback) und bezogen dabei Schüler:innenäußerungen kaum ein (van de Pol et al., 2010). Gemäß den Annahmen des Angebots-Nutzungs-Modell werden letztere jedoch ebenfalls als wesentlich für das Gelingen von Lernunterstützung erachtet (Vieluf et al., 2020). Die Bedeutung wechselseitig aufeinander bezogener Gesprächsbeiträge verdeutlichen Videostudien zum Klassenunterricht (Interaktionen der Lehrperson mit der gesamten Klasse). Sie legen nahe, dass die Wirkung von Unterstützungshandlungen der Lehrperson sich insbesondere dann entfaltet, wenn diese elaborierte Schüler:innenäußerungen anregen und ko-konstruktiv unterstützen (Reusser & Pauli, 2013; Ruiz-Primo & Furtak, 2006, 2007; Walshaw & Anthony, 2008; Webb et al., 2014).

Neben den Gesprächsbeiträgen von Lernenden sind zudem ihre kognitiven und motivationalen Voraussetzungen wichtige Faktoren in Unterrichtsinteraktionen. Befunde zum differenziellen Handeln von Lehrpersonen in Arbeitsphasen zeigen, dass Lehrpersonen sich leistungsschwachen Schüler:innen zeitlich intensiver zuwenden (Lipowsky et al., 2008; Pohlmann-Rother et al., 2018, 2020; Pohlmann-Rother & Kürzinger, 2015). Zudem existieren erste Hinweise, dass auch die motivationalen Voraussetzungen von Lernenden (z. B. niedrige Selbstwirksamkeit) differenzielle Reaktionen von Lehrpersonen in Interaktionen hervorrufen können (Boyer et al., 2007; Pielmeier et al. 2018). Personenzentrierte Untersuchungen heben hervor, dass kognitive und motivationale Voraussetzungen eng zusammenhängen. Sie liegen bei Lernenden in verschiedenen Ausprägungen (kognitiv-motivationalen Profilen) vor und können auch divergent sein (z. B. hohes Vorwissen bei gleichzeitig niedriger Selbstwirksamkeit; Lazarides et al., 2019; Linnenbrink-Garcia et al., 2012; Ruelmann et al., 2021; Seidel, 2006). Die Rolle von kognitiv-motivationalen Profilen der Lernenden in Bezug auf das Lehrpersonenhandeln in Unterrichtsinteraktionen wurde bislang kaum erforscht. Diese zu untersuchen, könnte zu einem tieferen Verständnis relevanter Faktoren für die Umsetzung von Lernunterstützung in Interaktionen beitragen. In der vorliegenden Studie werden die genannten Desiderata adressiert und Unterrichtsinteraktionen zwischen Lehrpersonen und Lernenden in Arbeitsphasen

2 In anderen Arbeiten wird der Begriff *Schülerarbeitsphasen* verwendet (z. B. Krammer, 2009). Aus Gründen der gendergerechten Sprache verwenden wir den Begriff Arbeitsphasen.

unter Berücksichtigung der Beiträge von Lernenden und ihrer kognitiv-motivationalen Profile exploriert.

Im Folgenden werden theoretische Grundlagen und empirische Erkenntnisse zur Lernunterstützung in Interaktionen zwischen Lehrenden und Lernenden dargelegt. Dazu wird zunächst die Relevanz von Arbeitsphasen für die Lernunterstützung thematisiert (2.1). Des Weiteren wird der Forschungsstand zu wechselseitigen Unterrichtsinteraktionen dargelegt (2.2) und vertiefend auf Befunde zum differenziellen Interaktionshandeln von Lehrpersonen eingegangen (2.3), woraus sich die Fragestellungen der vorliegenden explorativen Untersuchung ergeben (2.4).

2 Lernunterstützung in Unterrichtsinteraktionen

2.1 Relevanz von Arbeitsphasen für die Lernunterstützung

In der Unterrichtsforschung wird zwischen verschiedenen Sozialformen unterschieden, welche eine strukturelle Grundlage für stattfindende Interaktionen bilden. Oftmals wird zwischen *Arbeitsphasen* und *Klassenunterricht* differenziert (z. B. Kobarg & Seidel, 2007). In Arbeitsphasen beschäftigen sich Lernende weitgehend selbstständig entweder einzeln (Sozialform: Einzelarbeit) oder gemeinsam (Sozialformen: Partner:innen- oder Gruppenarbeit) mit Arbeitsaufträgen. Demnach können in Arbeitsphasen sowohl Einzelinteraktionen (eins-zu-eins mit der Lehrperson) als auch Interaktionen der Lehrperson mit mehreren Lernenden stattfinden. In der Sozialform Klassenunterricht adressiert die Lehrperson die gesamte Klasse in einem Klassengespräch und interagiert folglich mit allen Lernenden.

In der vorliegenden Untersuchung liegt der Fokus auf Arbeitsphasen. Arbeitsphasen machen mit durchschnittlich 50 bis 70 % einen bedeutenden Anteil von Unterrichtsstunden aus (Krammer et al., 2010) und sind somit wichtige Zeitfenster für lernunterstützende Interaktionen. Während Lernende selbstständig an Aufträgen arbeiten, kann die Lehrperson mit Schüler:innen über aktuell stattfindende Lernprozesse in den Dialog treten, sie situativ unterstützen und so zwischen dem Lernangebot und dessen optimaler Nutzung durch die Lernenden vermitteln (Schnebel, 2017). Eine solche unterrichtsbegleitende Lernunterstützung wird als entscheidender Faktor für einen adaptiven Unterricht und somit eine hohe Unterrichtsqualität gesehen (Krammer, 2016; Pauli & Reusser, 2015).

Vielfältige Ansätze beschäftigen sich mit der Unterstützung des Lernens in Interaktionen, unter anderem das *formative Assessment* (Bürgermeister & Saalbach, 2018) und das *Scaffolding* (van de Pol et al., 2010). Die genannten Konzepte haben das Anliegen einer individuell-adaptiven Förderung von Lernenden auf Basis diagnostischer Informationen gemein, also die Idee einer (ko-)konstruktiven Verknüpfung von Diagnose und Förderung im Unterricht. Im deutschsprachigen Raum wird dies allgemeiner auch als *Lernunterstützung* bezeichnet (Krammer, 2009; Krammer et al., 2010; Reusser &

Pauli, 2010). Lernunterstützung wird daher im Folgenden als Oberbegriff für verschiedene Ansätze der Diagnose und Förderung von Schüler:innen in Interaktionen verwendet. Der Lehrperson kommt bei der Lernunterstützung aus theoretischer Sicht nach Wygotskis Konzept der Zone der proximalen Entwicklung eine große Bedeutung zu. Die Zone der proximalen Entwicklung kann im Kontext von Unterrichtsinteraktionen zwischen Lehrpersonen und Lernenden als potenzieller nächster Lernschritt von Schüler:innen verstanden werden, welchen sie mit der Unterstützung einer Lehrperson erreichen können, jedoch nicht ohne diese (Bakker et al., 2015). Eine Vielzahl von Studien zur Lernunterstützung in Arbeitsphasen konzentriert sich daher auf die Angebotsseite, das Handeln der Lehrperson, mit dem Ziel, lernerfolgssteigernde Unterstützungsstrategien zu identifizieren (für eine Übersicht z. B. van de Pol et al., 2010). Zwei zentrale Unterstützungshandlungen, die als Schlüsselstrategien von erfolgreichem formativem Assessment gelten und wiederkehrend in verschiedenen Arbeiten zur Lernunterstützung genannt werden, sind *Eliciting Evidence of Learning* (lernbegleitende Diagnose) und *Feedback* (lernförderliche Rückmeldungen; Buholzer et al., 2020; Bürgermeister & Saalbach, 2018; Krammer, 2009; Krammer et al., 2010; Pohlmann-Rother et al., 2018). *Eliciting Evidence of Learning* kann mit „Lernnachweise entlocken“ übersetzt werden und umschreibt Handlungen der Lehrperson, die darauf abzielen, Einblicke in die Lernprozesse von Schüler:innen zu gewinnen (z. B. indem offene, aktivierende Fragen gestellt werden; Buholzer et al., 2020; Heritage, 2010). Lernförderliches Feedback knüpft an den Äußerungen von Lernenden gezielt an und unterstützt sie dabei, die Differenz zwischen dem aktuellen Lernstand und den Lernzielen zu überwinden (Hattie & Timperley, 2007; Sadler, 1989).

Zusammenfassend wird Arbeitsphasen eine hohe Relevanz für die gezielte Unterstützung der Lernprozesse von Schüler:innen zugesprochen. In der empirischen Forschung zur Umsetzung dieser Unterstützung wurde bislang jedoch weitgehend auf die Lehrperson fokussiert. Die gängige Praxis, nur Lehrpersonenäußerungen zu kodieren, bemängeln beispielsweise van de Pol und Kolleg:innen (2010) in ihrer Metaanalyse und heben hervor, dass auch die Beiträge von Lernenden einbezogen werden sollten.

2.2 Forschungsbefunde zu wechselseitigen Interaktionen

Eine breite Forschung zu wechselseitigen Interaktionen, in der auch Schüler:innenäußerungen berücksichtigt wurden, existiert zum Klassengespräch (z. B. Denn et al., 2016; Howe & Abedin, 2013; Park et al., 2020; Pfister et al., 2015). Studien zum Klassengespräch konnten als gängigstes Interaktionsmuster das fragend-entwickelnde Unterrichtsgespräch identifizieren (z. B. Ufer et al., 2015), das sich durch das häufige Auftreten sogenannter IRE- bzw. IRF-Sequenzen (Teacher Initiation, Student Response, Teacher Evaluation/ Follow up) auszeichnet (Howe & Abedin, 2013; Mehan, 1979). Lehrpersonen lenken das Klassengespräch dabei mittels geschlossener Fragen, während Schüler:innen als sogenannte „Stichwortgeber:innen“ agieren, indem sie die Gedankenführung der Lehrperson durch knappe Antworten ergänzen. Diese werden durch

korrigierendes Feedback der Lehrperson entweder verifiziert oder falsifiziert (Mehan, 1979). Fragend-entwickelnde Unterrichtsgespräche nach dem beschriebenen IRE/F-Muster wirken sich deaktivierend auf die emotionale und kognitive Partizipation von Lernenden aus (Reusser & Pauli, 2013; Kobarg & Seidel, 2007; Mehan, 1979; Seidel & Prenzel, 2006).

Im Kontrast dazu konnten im Klassengespräch auch dialogische, lernförderliche Interaktionsmuster unter anderem in Studien zum formativen Assessment identifiziert werden (Ruiz-Primo & Furtak, 2006, 2007). Ein lernförderliches Interaktionsmuster ist der sogenannte *Feedbackloop* (Heritage, 2010) oder auch *ESRU-Cycle* (Elicit Response, Student Response, Recognize, Use; Ruiz-Primo & Furtak, 2006, 2007). Dieses Muster ist durch offene, aktivierende Fragen (Eliciting Evidence) der Lehrperson charakterisiert, welche die Schüler:innen zur Verbalisierung ihrer Denk- und Lernwege (elaborierte Beiträge) anregen. An diese Äußerungen knüpfen Lehrpersonen gezielt mit Feedback an, das nicht auf die Korrektheit von Antworten fokussiert. Vielmehr greift es Überlegungen von Lernenden beispielsweise durch Paraphrase und Revoicing auf, führt sie weiter und fordert von Lernenden Erklärungen und Begründungen ein (Chin, 2006; Furtak et al., 2017; Ruiz-Primo & Furtak, 2006, 2007). Feedbackloops sind durch eine Abfolge reziproker Gesprächsbeiträge charakterisiert, die sich als ko-konstruktiver Dialog zwischen Lehrperson und den Lernenden entfalten. Befunde zur Wirksamkeit von Feedbackloops bzw. dialogischen Klassengesprächen deuten darauf hin, dass vollständige Feedbackloops, bei denen an Äußerungen von Lernenden angeknüpft wird, zu besseren Lernerfolgen führen als unvollständige Feedbackloops (Ruiz-Primo & Furtak, 2006, 2007; Walshaw & Anthony, 2008; Webb et al., 2014).

Studien zu wechselseitigen Interaktionen beziehen sich vor allem auf das Klassengespräch. Wie Interaktionen in Arbeitsphasen gestaltet sind, ist bislang wenig systematisch in größeren Stichproben erforscht worden. Jedoch dürfte die dialogische Einbindung der lehrpersonenseitigen Unterstützung auch hier relevant sein (Bakker et al., 2015). Erste Befunde zu Einzelinteraktionen existieren aus der Forschung zum Tutoring. Dabei wurden Gesprächssituationen zwischen einem Tutor oder einer Tutorin (Expert:innen) und einem Schüler oder einer Schülerin (Noviz:innen) analysiert. In einer Linie mit den Studien zum Klassenunterricht legen sie nahe, dass beim Tutoring ebenfalls eine enge Gesprächsführung dominiert (z. B. Chi et al., 2001; Graesser & Person, 1994). Wie im Klassenunterricht wurden überwiegend IRE/F-Muster festgestellt. Diese können als eine Art direktiv geführtes Scaffolding³ beschrieben werden: Die Tutor:innen zergliederten komplexe Arbeitsaufträge in kleinschrittige Teilaufgaben und die Lernenden brachten nur kurze Beiträge ein, in denen sie die kleinen Teilschritte lösten (Graesser & Person, 1994). Inwieweit auch Feedbackloops in diesen Einzelinteraktionen umgesetzt werden, ist bislang kaum erforscht. Die Ergebnisse von Chi et al. (2001) deuten jedoch analog zu den Befunden aus dem Klassenunterricht darauf

3 Scaffolding wurde in diesen Studien als eine dialogische Episode mit mehreren Sprechendenwechseln definiert, die konkrete Unterstützungen zu demselben Konzept oder Thema beinhaltet.

hin, dass auch beim Tutoring eine stärker dialogische Gesprächsführung, ähnlich dem Muster von Feedbackloops, lernförderlich ist: So führte insbesondere eine Kombination anregender Fragen des Tutors oder der Tutorin in Kombination mit elaborierten Beiträgen der Lernenden zu einem Lernzuwachs (Chi et al., 2001; Chi, 2009).

Zusammenfassend zeigen Studien zu wechselseitigen Interaktionen, dass Unterstützungshandlungen der Lehrperson, wie z. B. anregende Fragen, ihr Potential entfalten, wenn sie zu elaborierten Schüler:innenbeiträgen und somit einer lernförderlichen (ko-) konstruktiven Interaktion führen (Walshaw & Anthony, 2008). Erklärungen und Begründungen mit anderen zu teilen, fordert Schüler:innen heraus, eigene Überlegungen zu strukturieren, zu hinterfragen und Missverständnisse zu erkennen, sodass Beziehungen zwischen neuen Lerninhalten und bestehenden Wissensstrukturen hergestellt werden können (Lipowsky et al., 2008; Pauli et al., 2008; Webb et al., 2014).

2.3 Kognitiv-motivationale Voraussetzungen von Schüler:innen und differenzielles Lehrpersonenhandeln

Sowohl theoretisch als auch empirisch wird davon ausgegangen, dass kognitive Voraussetzungen (z. B. Vorwissen) und motivationale Voraussetzungen (z. B. Selbstwirksamkeit und intrinsische Motivation) für Lernprozesse von Belang sind (Linnenbrink-Garcia et al., 2012; Ruelmann et al., 2021; Vieluf et al., 2020). Ein Fokus der Forschung liegt darauf, wie kognitiv-motivationale Voraussetzungen von Lernenden die Nutzung des Unterrichts beeinflussen. Personenzentrierte Untersuchungen, welche Kombinationen von kognitiven und motivationalen Voraussetzungen (kognitiv-motivationale Profile) berücksichtigen, konnten Zusammenhänge zwischen den kognitiv-motivationalen Profilen von Lernenden und ihrer Wahrnehmung und Nutzung des Unterrichtsangebots sowie ihrer aktiven Beteiligung feststellen (Jurik et al., 2013, 2014, 2015; Lazarides et al., 2019, 2020; Linnenbrink-Garcia et al., 2012; Ruelmann et al., 2021; Seidel, 2006). Was die Wahrnehmung der Lernunterstützung betrifft, deuten die Ergebnisse von Ruelmann et al. (2021) daraufhin, dass diese je nach kognitiv-motivationalem Profil von Schüler:innen unterschiedlich ausfällt. Beispielsweise nahmen Lernende mit einem *überforderten Profil* (niedriges Vorwissen, niedrige Selbstwirksamkeit und niedrige intrinsische Motivation) und *unmotiviertem Profil* (niedrige intrinsische Motivation und Selbstwirksamkeit, aber durchschnittliches Vorwissen) die Lernunterstützung ihrer Lehrperson im Vergleich zu anderen Profilen negativer wahr.

Ergebnisse zur aktiven Beteiligung von Lernenden mit unterschiedlichen kognitiven und motivationalen Voraussetzungen liegen ebenfalls überwiegend aus Videoanalysen des Klassengesprächs vor. Demnach melden sich Lernende mit großem Vorwissen bzw. ausgeprägter Leistungsfähigkeit sowie hoher Selbstwirksamkeit häufiger. Sie beteiligen sich öfter und mit längeren Beiträgen am Klassengespräch, verglichen mit ihren Mitschüler:innen. Diese aktive Beteiligung ist zudem mit einem größeren Lernzuwachs verbunden (Boyer et al., 2007; Decristan, Fauth et al., 2020; Denn, 2021; Jurik et al., 2013; Pauli & Lipowsky, 2007; Pielmeier et al., 2018; Schnitzler et al., 2021). Einige Ergebnisse

deuten auch für Arbeitsphasen auf eine stärkere Eigeninitiative leistungsstärkerer und motivierterer Schüler:innen (Nemeth et al., 2019; Pauli & Lipowsky, 2007; Pohlmann-Rother et al., 2020).

Zum differenziellen Handeln von Lehrpersonen liegen ebenfalls hauptsächlich Ergebnisse zum Klassengespräch vor. Diese konnten einen stärkeren Einbezug von Leistungsstarken und Schüler:innen mit einem positiveren Selbstkonzept feststellen. Im Vergleich zu Mitschüler:innen wurden sie öfter aufgerufen und Lehrpersonen richteten häufiger aktivierende Fragen an sie (Decristan, Fauth et al., 2020; Lipowsky et al., 2007; Pauli & Lipowsky, 2007; Pielmeier et al., 2018). Für Arbeitsphasen hingegen existieren nur wenige Beobachtungsstudien zum differenziellen Handeln von Lehrpersonen. Konträr zum Klassenunterricht legen diese nahe, dass Lehrpersonen in Arbeitsphasen häufiger Interaktionen mit leistungsschwächeren Schüler:innen initiieren und diese zeitlich intensiver unterstützen (Lipowsky et al., 2008; Lotz & Lipowsky, 2014; Pohlmann-Rother et al., 2018, 2020). Fast keine Befunde existieren zur Art der Unterstützung bzw. zu Interaktionsmustern. Im Hinblick auf motivationale Faktoren ergaben Analysen von Boyer et al. (2007) mit Studierenden, dass Tutor:innen Studierende mit niedrigerer Selbstwirksamkeit und geringem Vorwissen häufiger lobten und ihnen häufiger einfaches Feedback als ihren Kommiliton:innen gaben (Boyer et al., 2007).

2.4 Die vorliegende explorative Untersuchung

Die vorangegangenen Ausführungen verdeutlichen, dass Arbeitsphasen ein relevantes Zeitfenster im Unterricht darstellen, in dem Lehrpersonen Lernende gezielt unterstützen können. Obwohl Studien zum Klassenunterricht und Tutoring verdeutlichen, dass Interaktionen zwischen Lehrperson und Schüler:innen sehr unterschiedlich gestaltet sein können, existieren bislang kaum Untersuchungen, die Lernunterstützung in Arbeitsphasen unter Einbezug wechselseitiger Gesprächsbeiträge der Lehrperson und Schüler:innen untersuchen. Arbeitsphasen unterschieden sich strukturell von dem gut erforschten Klassenunterricht und Tutoringsituationen, da sie in verschiedenen Sozialformen (Einzelarbeit vs. Partner:innen oder Gruppenarbeit) stattfinden können. Je nach Sozialform ist der Interaktionskontext somit eher dem Klassenunterricht oder dem Tutoring ähnlich. Bei der Sozialform Einzelarbeit finden mehrheitlich Einzelinteraktionen der Lehrperson mit einzelnen Lernenden statt (ähnlich wie beim Tutoring). Bei Partner:innen- und Gruppenarbeiten sind hingegen mehrere Lernende in die Interaktion mit der Lehrperson einbezogen (eher vergleichbar mit dem Klassenunterricht). Es gilt somit zu überprüfen, inwieweit Interaktionsmuster, die im Klassenunterricht und Tutoring identifiziert wurden (z. B. Feedbackloops oder IRE/F-Muster), in Arbeitsphasen realisiert werden und ob diese zwischen den Interaktionskontexten, die durch die Sozialformen determiniert sind, variieren.

Zudem ist es relevant, in diesem Kontext auch zu untersuchen, inwieweit sich Interaktionsmuster für Lernende mit verschiedenen kognitiv-motivationalen Profilen un-

terscheiden. Ob Schüler:innen dieselbe oder eine unterschiedliche Lernunterstützung erhalten, könnte Konsequenzen für die Interpretation des Zusammenhangs zwischen Lernunterstützung und der Leistungsentwicklung von Schüler:innen haben. In der vorliegenden Untersuchung werden die dargelegten Desiderata adressiert und die folgenden Forschungsfragen explorativ untersucht:

Forschungsfrage 1: Welche Interaktionsmuster zwischen Lehrperson und Schüler:innen lassen sich in Arbeitsphasen im Mathematikunterricht identifizieren und wie sind diese verteilt? Bestehen diesbezüglich Unterschiede zwischen Einzelinteraktionen und Interaktionen der Lehrperson mit mehreren Lernenden?

Forschungsfrage 2: Unterscheiden sich die Interaktionsmuster in Abhängigkeit von den kognitiv-motivationalen Profilen der Schüler:innen im Fach Mathematik?

3 Methode

3.1 Stichprobe und Datenerhebung

Die nachfolgenden Analysen basieren auf Video- und Fragebogendaten der TUFA-Studie (Buholzer et al., 2020), die in der ersten Jahreshälfte 2017 erhoben wurden. An der Studie nahmen 52 Lehrpersonen und ihre 633 Schüler:innen der vierten Primarschulklassen aus der Zentralschweiz teil. Die Lehrpersonen waren durchschnittlich 36 Jahre alt ($SD = 10.6$; Range: 25–60 Jahre), zu 76,9 % weiblich und hatten eine mittlere Berufserfahrung von 10,6 Jahren ($SD = 10.3$; Range: 1,5–39,0 Jahre). Die Lernenden waren durchschnittlich 10,5 Jahre alt ($SD = 0.49$; Range: 9–12,6) und zu 50,6 % Mädchen. Etwa 38 % der Schüler:innen waren mehrsprachig. Sie gaben an, zu Hause vorwiegend eine weitere oder andere Sprache als (Schweizer-)Deutsch zu sprechen.

Von jeder der 52 teilnehmenden Schulklassen wurde eine 90-minütige Mathematik- lektion zum Thema „Einführung in die halbschriftliche Division“ videografiert. Diese Thematik ist ein Lerninhalt im vierten Schuljahr in der (Deutsch-)Schweiz und wurde allen teilnehmenden Klassen vorgegeben, um vergleichbare Rahmenbedingungen zu gewährleisten.

Die Videoaufzeichnung des Unterrichts orientierte sich an einem standardisierten Kameraskript und wurde von Mitgliedern der Forschungsgruppe sowie geschulten studentischen Hilfskräften durchgeführt. Die Kamera folgte stets der Lehrperson, welche ein Ansteckmikrofon trug, um eine hohe Video- und Audioqualität zu erzielen. Lehrpersonen, Eltern und Kinder stimmten der Studienteilnahme gemäß den forschungsethischen Richtlinien im Vorfeld schriftlich zu. Schüler:innen, die nicht an der Studie teilnahmen, saßen außerhalb der Kamerareichweite.

Vor Beginn der Unterrichtseinheit zur halbschriftlichen Division wurde das Vorwissen der Schüler:innen mit einem standardisierten Leistungstest erfasst. Die Selbst-

wirksamkeit und intrinsische Motivation der Lernenden wurden anschließend mittels einer standardisierten Befragung durch Mitglieder der Forschungsgruppe und geschulte studentische Hilfskräfte in Abwesenheit der Lehrperson erhoben. Dabei wurden die Fragebogenitems laut vorgelesen, um eine Konfundierung der Ergebnisse mit der Lesekompetenz der Schüler:innen auszuschließen.

3.2 Instrumente

3.2.1 Beobachtungsinstrument zur Erfassung der Lernunterstützung in Interaktionen

Bestehende Beobachtungsinstrumente zur Erfassung von Lernunterstützung in Unterrichtinteraktionen untersuchten diese mit dem Fokus auf Lehrpersonenhandlungen, indem z. B. Feedback (Hinweise und Erklärungen) oder Fragetechniken erhoben wurden (z. B. Denn, 2021; Krammer, 2009; Lotz, 2016; Pohlmann-Rother et al., 2018, 2020). Forschungsergebnisse verweisen jedoch auf die Lernförderlichkeit einer interaktiv eingebetteten Unterstützung (vgl. 2.3). Um diese zu analysieren, ist es notwendig, unterstützende Handlungen von Lehrpersonen mit dem Auftreten elaborierter Schülerbeiträge simultan zu erfassen und Interaktionen somit ganzheitlich in den Blick zu nehmen. Zu diesem Zweck wurde für die vorliegenden Auswertungen mittels eines induktiv-deduktiven Verfahrens ein Kategoriensystem zur Erfassung von Interaktionsmustern entwickelt. Dabei wurden zentrale Unterstützungsstrategien von Lehrpersonen (Eliciting Evidence und Feedback) sowie das Zusammenspiel von Lehrpersonen- und Schüler:innenäußerungen einbezogen (vgl. Abb. 1).

Das Kategoriensystem knüpft an existierende Klassifizierungssysteme an und führt diese weiter (Chi et al., 2001; Helmke et al., 2007; Krammer, 2009; Ruiz-Primo & Furtak, 2007). Eine Schichtung verschiedener Verfahren ermöglicht die Strukturierung und Quantifizierung des Videomaterials im Hinblick auf das untersuchte Merkmal (Pauli, 2012). Daher wurden niedrig inferente Codierungen, die kaum interpretative Schlussfolgerungen erfordern (Basiscodierung), als Grundlage für mittel inferente Codierungen (vertiefende Codierung) verwendet, die sich zur Beurteilung komplexerer Merkmale (d. h. der Lernunterstützung in Interaktionen) eignen.

Mit einem Eventsampling wurden alle Interaktionen zwischen Lehrperson und Schüler:innen in Arbeitsphasen mit Anfangs- und Endzeitpunkt erfasst. Anhand der Basiscodierung (vgl. Abb. 1) wurden grundlegende Charakteristika jeder Interaktion erhoben: die Initiierung durch die Lehrperson oder Lernende (Initiierung), wie viele Lernende (Anzahl) und welche Lernende (ID-Zuordnung) an der jeweiligen Interaktion beteiligt waren. Die Schüler:innen wurden in den Videos auf der Basis von Sitzplänen anhand differenzierter Codierregeln identifiziert. Dabei wurde die Richtigkeit der Sitzpläne u. a. anhand von Namensnennungen in den Interaktionen sichergestellt und geprüft.

Im Zuge der vertiefenden Codierung (vgl. Abb. 1) wurden die Interaktionen inhaltlich analysiert. Es wurde zum einen festgehalten, inwieweit sie sich auf die Arbeit an

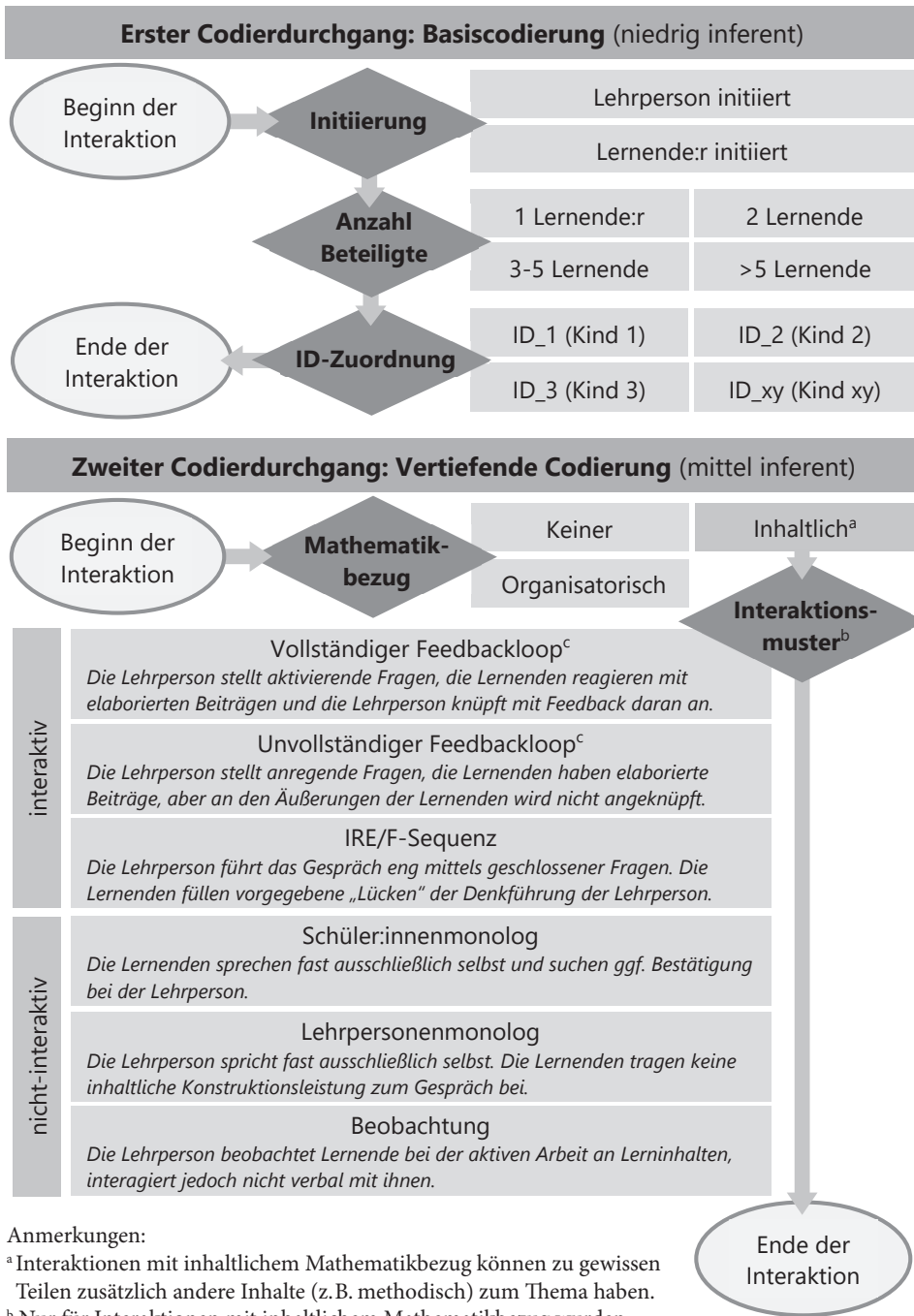
mathematischen Themen beziehen (Mathematikbezug). Für alle Interaktionen mit inhaltlichem Mathematikbezug wurden die zugrunde liegenden Interaktionsmuster bestimmt (Interaktionsmuster). Dabei konnten analog zur existierenden Forschung zum Klassenunterricht und Tutoring verschiedene Interaktionsmuster identifiziert werden, die sich hinsichtlich der Interaktivität und des Grades der Schüler:innenbeteiligung unterscheiden (vgl. Abb. 1). Es wurde zwischen sechs Interaktionsmustern differenziert: Vollständige und unvollständige Feedbackloops sind interaktiv und können als dialogisch charakterisiert werden, da sie wechselseitige Gesprächsbeiträge der Lehrperson und Schüler:innen enthalten und sich durch eine hohe Schüler:innenbeteiligung auszeichnen. Enggeführte Interaktionen (IRE/F-Sequenzen) sind zwar interaktiv, enthalten also wechselseitige Gesprächsbeiträge, jedoch ist die Schüler:innenbeteiligung gering. Schüler:innenmonologe wiederum zeichnen sich zwar durch eine hohe Schüler:innenbeteiligung aus, sind jedoch nicht interaktiv. Ebenfalls nicht interaktiv sind Lehrpersonenmonologe und „stille“ Beobachtungen. Des Weiteren unterscheidet sich der Grad der Lehrpersonendominanz zwischen den Interaktionsmustern: Lehrpersonenmonologe und enggeführte Interaktionen (IRE/F-Sequenzen) lassen sich im Gegensatz zu den anderen Mustern als lehrpersonendominiert beschreiben, da die Lehrperson das Gespräch stark lenkt.

3.2.2 Kognitiv-motivationale Schüler:innenprofile

Für die vorliegende Arbeit wurden die kognitiv-motivationalen Schüler:innenprofile von Ruelmann et al. (2021) verwendet.⁴ Diese wurden mittels latenter Profilanalysen anhand von drei Variablen geschätzt: Das Vorwissen wurde mit einem Leistungstest zum Thema „halbschriftliches Dividieren“ bestimmt, indem ein Summenscore gebildet wurde. Beispielsweise wurden basale Fähigkeiten in der Multiplikation und Division abgefragt (Bsp.: $8 \times 7 = _$). Die intrinsische Motivation und Selbstwirksamkeit wurden mathematikspezifisch mit zwei Skalen erfasst (Beispielitems: intrinsische Motivation, „Mathematik macht mir Spaß.“; Selbstwirksamkeit, „Ich kann fast alle Aufgaben in Mathematik schaffen, wenn ich mich anstreng.“; Rimm-Kaufman et al., 2015).

Ruelmann et al. (2021) identifizierten vier kognitiv-motivationale Profile mit unterschiedlichen Ausprägungen der drei Variablen: (1) *starkes Profil* mit hoch ausgeprägter Selbstwirksamkeit, hoch ausgeprägter intrinsischer Motivation und hoch ausgeprägtem Vorwissen; (2) *motiviertes Profil* mit hoch ausgeprägter intrinsischer Motivation, durchschnittlicher Selbstwirksamkeit und geringem Vorwissen; (3) *unmotiviertes Profil* mit geringer Selbstwirksamkeit, geringer intrinsischer Motivation und mittlerem Vorwissen; (4) *überfordertes Profil* mit niedrigen Werten aller Variablen.

4 Genauere Informationen zur Identifikation der Profile können bei Ruelmann et al. (2021) nachgelesen werden.



Anmerkungen:

^a Interaktionen mit inhaltlichem Mathematikbezug können zu gewissen Teilen zusätzlich andere Inhalte (z.B. methodisch) zum Thema haben.

^b Nur für Interaktionen mit inhaltlichem Mathematikbezug wurden Interaktionsmuster unterschieden.

^c Diese Interaktionsmuster wurden codiert, sobald sie auftraten, und umfassen nicht immer die gesamte Dauer einer Interaktion.

Abbildung 1: Codingverfahren zur Erfassung von Lernunterstützung in Unterrichtsinteraktionen

3.3 Statistische Analysen

3.3.1 Reliabilität und Validität des verwendeten Beobachtungsinstrumentes

Die Videos wurden anhand eines detaillierten Codebooks (Zusammenfassung des Verfahrens in Abb. 1) von zwei geschulten Codierern (der Erstautorin und einer Praktikantin) mit dem Programm MAXQDA 2020 codiert. Um die Reliabilität des Beobachtungssystems sicherzustellen, wurde die Interraterreliabilität (Cohens Kappa) zwischen den beiden Codierern für 5 bis 10 % des Videomaterials berechnet. Die Interraterreliabilität zwischen den Codierern lag für die Facetten der Basiscodierung zwischen 0.91 und 0.94 und für die Facetten der vertieften Codierung zwischen 0.87 und 0.88, was einer hohen Übereinstimmung entspricht (vgl. Landis & Koch, 1977; Wirtz & Caspar, 2002).

Um die Validität der Interaktionsmuster zu überprüfen, wurden Quervergleiche mit Codierungen der TUFA-Gesamtstudie durchgeführt. Diese zeigten erwartungsgemäße Unterschiede zwischen den Interaktionsmustern und deuten darauf hin, dass die Lernunterstützung und Schüler:innenbeteiligung in den Interaktionsmustern valide abgebildet wurde.

3.3.2 Kontingenzanalysen: Unterschiede in der Verteilung der Interaktionsmuster

Zur Analyse von Unterschieden in der Verteilung der Interaktionsmuster wurden sowohl univariate deskriptive Auswertungen vorgenommen als auch Kontingenzanalysen (Kreuztabellen, Chi²-Tests) durchgeführt. Da Arbeitsphasen sich durch verschiedene Sozialformen auszeichnen, wurde zur Analyse der Verteilung der Interaktionsmuster zwischen Einzelinteraktionen und Interaktionen der Lehrperson mit mehreren Schüler:innen (eins-zu-eins versus eins-zu-mehrere) unterschieden. Bei den Analysen von Unterschieden in der Verteilung der Interaktionsmuster zwischen den kognitiv-motivationalen Profilen wurden nur die Einzelinteraktionen ausgewertet. In diesem Setting kann eher von einer Orientierung des Handelns der Lehrperson an den kognitiv-motivationalen Voraussetzungen ausgegangen werden als in Interaktionen mit mehreren.

Die vorliegende Stichprobe unterscheidet sich durch die Verschachtelung von Schüler:innen in Klassen von einer einfachen Zufallsstichprobe. Deshalb wurde bei der Berechnung der Kontingenzanalysen für die hierarchische Datenstruktur korrigiert. Es wurde eine Korrektur sowohl des Chi²-Tests auf Unabhängigkeit wie der adjustierten Residuen nach dem von Rao und Scott (1981) entwickelten Verfahren durchgeführt. Dieses ist in SPSS im Rahmen der Prozeduren für komplexe Stichproben (complex samples) implementiert (IBM Corporation, 2021). Dabei werden die Stichprobenfehler unter Berücksichtigung des Grades der überzufälligen Ähnlichkeit der Merkmalstragenden innerhalb der Klassen geschätzt. Die adjustierten standardisierten Residuen folgen einer t- bzw. z-Verteilung, sodass Werte, die im Betrag größer 2 sind, als Hinweis auf eine signifikante zellbezogene Abweichung (mit $p < .05$) vom Erwartungswert unter Annahme der Unabhängigkeit der Variablen gelten (Haberman, 1973). Ein signifikan-

ter positiver standardisierter Residualwert bedeutet folglich, dass die in der betreffenden Zelle abgebildete Merkmalskombination häufiger vorkommt, als statistisch zu erwarten wäre, wenn keine Beziehung zwischen den Variablen bestünde. Konträr bedeutet ein signifikanter negativer standardisierter Residualwert, dass die entsprechende Merkmalskombination überzufällig seltener vorkommt als erwartet.

4 Ergebnisse

4.1 Interaktionsmuster in Arbeitsphasen

Die Arbeitsphasen hatten durchschnittlich eine Länge von 33,8 Minuten und einen Anteil an der Gesamtlektion von 42 %. Eine Interaktion dauerte im Mittel 44 Sekunden ($N = 2041$; $SD = 47$ Sek.; Range = 3 Sek. – 6,6 Min.). Einzelinteraktionen kamen mit 77,6 % ($N = 1584$; $M = 40$ Sek.; $SD = 43$ Sek.; Range = 3 Sek. – 6,2 Min.) am häufigsten in Arbeitsphasen vor. Demgegenüber fanden nur 21,9 % der Interaktionen mit mehreren Lernenden statt ($N = 447$; $M = 56$ Sek.; $SD = 55$ Sek.; Range = 00:03 – 6:35).⁵ Die Interaktionsmuster unterschieden sich in ihrer durchschnittlichen Dauer: Eine längere Dauer konnte bei IRE/F-Sequenzen ($N = 627$; $M = 74$ Sek.; $SD = 56$ Sek.) festgestellt werden. Unvollständige Feedbackloops ($N = 123$; $M = 79$ Sek.; $SD = 61$ Sek.) sowie vollständige Feedbackloops ($N = 45$; $M = 90$ Sek.; $SD = 95$ Sek.) wurden ebenfalls eher in längeren Interaktionen realisiert. Lehrpersonenmonologe ($N = 920$; $M = 25$ Sek.; $SD = 24$ Sek.) und Schüler:innenmonologe ($N = 254$; $M = 27$ Sek.; $SD = 23$ Sek.) lassen sich als kürzere Inputs beschreiben. Die kürzeste Dauer hatten Beobachtungen ($N = 72$; $M = 11$ Sek.; $SD = 5$ Sek.).

Im Hinblick auf die erste Forschungsfrage, welche Interaktionsmuster sich in Arbeitsphasen identifizieren lassen und wie diese verteilt sind, zeigen die Ergebnisse signifikante Unterschiede in der Verteilung der Interaktionsmuster zwischen Einzelinteraktionen und Interaktionen der Lehrperson mit mehreren Lernenden ($\chi^2 = 324.20$; $F_{adj} (4.27, 217.67) = 28.42$; $p < .001$). Wie Abbildung 2 und Tabelle 1 zu entnehmen ist, lassen sich Einzelinteraktionen zu 84,0 % durch eine lehrpersonendominierte Gesprächsführung beschreiben, in denen sich Schüler:innen kaum äußern: In diesen Interaktionen sind 50,6 % reine Monologe der Lehrperson, bei denen die Lernenden ausschließlich zuhören, und 33,4 % sind enggeführte Gesprächsmuster (IRE/F-Sequenzen), bei denen Schüler:innen als Stichwortgeber:innen fungieren. Interaktionsmuster, die eine stärkere Schüler:innenbeteiligung zulassen (Schüler:innenmonologe, unvollständige und vollständige Feedbackloops), waren in Einzelinteraktionen selten. Von diesen umfassen Schüler:innenmonologe mit 8,0 % den größten Anteil. Vollständige Feedbackloops, in denen Lehrpersonen durch offene, aktivierende Fragen längere Beiträge anregen und mit Feedback an diesen anknüpfen, machen lediglich 0,8 % aller Einzelinteraktionen

5 Bei den restlichen 0,5 % der Interaktionen konnte z. B. wegen mangelnder Sicht- oder Hörbarkeit der Aufzeichnungen die Anzahl der beteiligten Lernenden nicht bestimmt werden.

aus. Weitere 3,5 % der Einzelinteraktionen sind unvollständige Feedbackloops, in denen Lehrpersonen anregende Fragen stellen und die Schüler:innen teilweise ihre Denk- und Rechenwege teilen, jedoch anschließend keine Anknüpfung seitens der Lehrperson stattfindet.

Demgegenüber sind Interaktionen der Lehrperson mit mehreren Lernenden durch eine deutlich offenere Gesprächsführung charakterisiert. Lehrpersonendominierte Interaktionsmuster und solche mit stärkerer Schüler:innenbeteiligung kommen zu gleichen Teilen vor. Schüler:innenmonologe umfassen mit 28,6 % sogar den größten Anteil an Interaktionen der Lehrperson mit mehreren Lernenden. Vollständige Feedbackloops werden in 7,2 % der Interaktionen mit mehreren Lernenden realisiert und unvollständige Feedbackloops in 14,8 %.

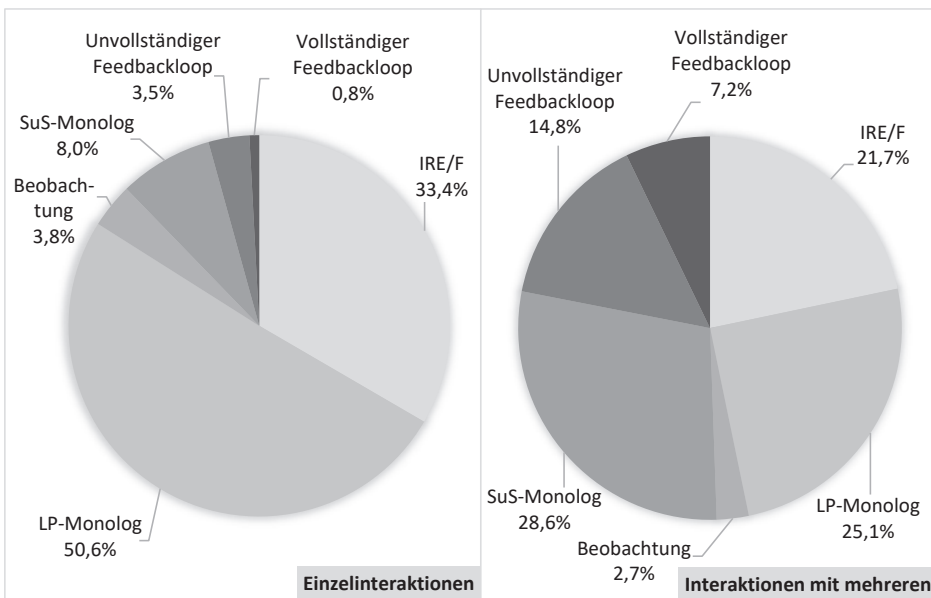


Abbildung 2: Interaktionsmuster in Einzelinteraktionen und Interaktionen mit mehreren

Unvollständige und vollständige Feedbackloops kommen zusammengenommen somit in diesen Interaktionen genauso häufig vor wie IRE/F-Sequenzen, die 21,7 % der Interaktionen mit mehreren Lernenden ausmachen. Lehrpersonenmonologe kommen mit 25,1 % nur halb so häufig vor wie in Einzelinteraktionen. Die beschriebenen Unterschiede in der Verteilung der Interaktionsmuster zwischen Einzelinteraktionen und Interaktionen der Lehrperson mit mehreren Lernenden sind für alle Interaktionsmuster mit Ausnahme von Beobachtungen der Lehrperson mit $p < .05$ signifikant (vgl. Tabelle 1).

Tabelle 1: Interaktionsmuster in Einzelinteraktionen und Interaktionen mit mehreren Lernenden

	Einzelinteraktionen			Interaktionen mit mehreren			Insgesamt	
	<i>N</i>	%	Std. Res.	<i>N</i>	%	Std. Res.	<i>N</i>	%
IRE/F	529	33,4	2.90*	97	21,7	-2.90*	626	30,8
LP-Monolog	801	50,6	6.29*	112	25,1	-6.29*	913	45,0
Beobachtung	60	3,8	0.89	12	2,7	-0.89	72	3,5
SuS-Monolog	126	8,0	-8.24*	128	28,6	8.24*	254	12,5
Unvollst. FB-Loop	55	3,5	-4.74*	66	14,8	4.74*	121	6,0
Vollst. FB-Loop	13	0,8	-5.74*	32	7,2	5.74*	45	2,2
Insgesamt	1584	78,0		447	22,0		2031	100

Anmerkungen. *N*: Anzahl der Interaktionen; %: Spaltenprozentwerte und unterste Zeile (Insgesamt) Reihenzusammenfassung; Std. Res.: Standardisierte adjustierte Residuen; *Signifikante Unterschiede ($p < .05$; |standardisierte adjustierte Residuen| > 2). Minimaler Erwartungswert in einer Zelle = 9.9.

4.2 Unterschiede zwischen den kognitiv-motivationalen Profilen

Zwischen den kognitiv-motivationalen Profilen zeigen sich signifikante Unterschiede im Hinblick auf die Verteilung der Interaktionsmuster in den Einzelinteraktionen ($\chi^2 = 54.01$; $F_{adj} (10.14, 516.99) = 2.55$; $p = .005$). Die Unterschiede sind in Abbildung 3 visualisiert und die genauen Werte sowie signifikanten Unterschiede pro Interaktionsmuster für jedes Profil sind in Tabelle 2 dargestellt.

Was die lehrpersonendominierten Interaktionsmuster (Lehrpersonenmonologe, IRE/F-Sequenzen) angeht, sind Schüler:innen sowohl mit einem überforderten Profil wie auch mit einem motivierten Profil signifikant häufiger an IRE/F-Interaktionen beteiligt. Diese machen rund die Hälfte der Einzelinteraktionen bei diesen Profilen aus. Bei Schüler:innen mit einem starken Profil umfassen IRE/F-Sequenzen hingegen mit 27,0 % nur ca. ein Drittel der Einzelinteraktionen. Hinsichtlich der Lehrpersonenmonologe zeigt sich hingegen ein leicht gegenteiliger Trend. Obwohl Lehrpersonenmonologe bei allen Profilen mit über 40 % stark dominieren, machen diese bei dem starken Profil mit 53,9 % den größten Anteil aus, der zudem einer signifikanten Übervertretung entspricht. Diese Schüler:innen werden auch am häufigsten, und signifikant übervertreten, ohne Eingreifen der Lehrpersonen beobachtet, während Schüler:innen mit einem überforderten Profil im Vergleich am seltensten nur beobachtet werden, d. h., ohne dass sich eine verbale Interaktion mit der Lehrperson anschließt.

Was die Interaktionsmuster mit einer stärkeren Schüler:innenbeteiligung (vollständige und unvollständige Feedbackloops, Schüler:innenmonologe) betrifft, wird deutlich, dass Schüler:innenmonologe ebenfalls beim starken Profil mit 9,0 % vergleichsweise am häufigsten vorkommen, während diese beim motivierten und überforderten Profil nur

ca. 6 % ausmachen. Diese Unterschiede sind jedoch nicht signifikant. Unvollständige und vollständige Feedbackloops konnten bei allen Profilen selten beobachtet werden. Während diese Interaktionsmuster zusammengenommen beim starken und unmotivierten Profil ca. 4 % der Einzelinteraktionen ausmachen, sind Schüler:innen mit einem überforderten Profil signifikant seltener an diesen Interaktionsmustern beteiligt. Vollständige Feedbackloops konnten bei diesem Profil nicht beobachtet werden und unvollständige Feedbackloops in nur 1,0 % der Einzelinteraktionen.

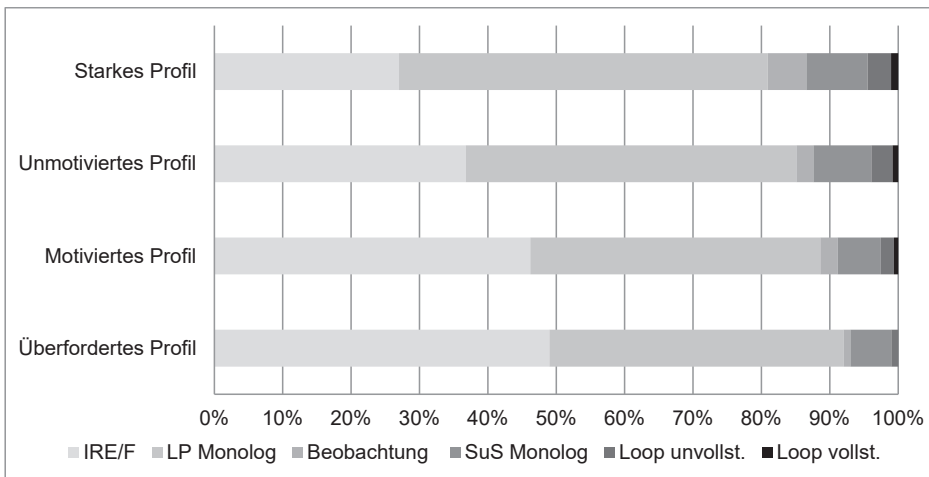


Abbildung 3: Verteilung der Interaktionsmuster nach Profil in Einzelinteraktionen

Tabelle 2: Interaktionsmuster nach Profiltypen

	Starkes Profil			Unmotiviertes Profil			Motiviertes Profil			Überfordertes Profil			Insgesamt ^a	
	N	%	Std. Res.	N	%	Std. Res.	N	%	Std. Res.	N	%	Std. Res.	N	%
IRE/F	181	27,0	-3.88*	178	36,8	1.25	73	46,2	2.42*	49	49,0	2.79*	481	34,1
LP-Monolog	361	53,9	2.35*	234	48,3	-0.77	67	42,4	-1.75	43	43,0	-1.37	705	49,9
Beobachtung	38	5,7	2.80*	12	2,5	-1.94	4	2,5	-1.17	1	1,0	-2.78°	55	3,9
SuS-Monolog	60	9,0	0.72	41	8,5	0.15	10	6,3	-1.01	6	6,0	-1.15	117	8,3
Unvollst. FB-Loop	23	3,4	0.86	15	3,1	0.18	3	1,9	-0.96	1	1,0	-2.10°	42	3,0
Vollst. FB-Loop	7	1,0	0.70	4	0,8	-0.07	1	0,6	-0.35	0	0,0	n.b.	12	0,8
Insgesamt	670	47,5		484	34,3		158	11,2		100	7,1		1412	100

Anmerkungen. N: Anzahl der Interaktionen; %: Spaltenprozentwerte und unterste Zeile (Insgesamt) Reihenprozentwerte; Std. Res.: Standardisierte adjustierte Residuen; fünf Zellen (20,8 %) weisen einen Erwartungswert < 5 auf. *Signifikante Unterschiede ($p < .05$; |standardisierte adjustierte Residuen| > 2); °nominelle Signifikanz ($p < .05$) aufgrund des Erwartungswerts < 5 fraglich; n.b.: nicht berechenbar.

^aDie von Tabelle 2 abweichenden Werte sind auf Missings in der Variable „Schüler:innenprofile“ zurückzuführen.

5 Diskussion

In der vorliegenden explorativen Untersuchung wurden Interaktionsmuster in Arbeitsphasen unter Berücksichtigung von Unterstützungsstrategien der Lehrperson (Fragen, Feedback) und Schüler:innenäußerungen identifiziert. Es wurde analysiert, inwieweit diese Interaktionsmuster in Einzelinteraktionen und Interaktionen der Lehrpersonen mit mehreren Lernenden unterschiedlich verteilt sind und inwieweit sie je nach den kognitiv-motivationalen Profilen der Schüler:innen variieren.

5.1 Interaktionsmuster in Arbeitsphasen

Die Ergebnisse zeigen, dass unterschiedliche Interaktionsmuster in Arbeitsphasen existieren. Analog zum Klassenunterricht überwiegen insgesamt lehrpersonendominierte Interaktionsmuster, die keine oder eine geringe Schüler:innenbeteiligung zulassen (Lehrpersonenmonologe, IRE/F-Sequenzen). Jedoch konnten auch Interaktionsmuster mit höherer Schüler:innenbeteiligung identifiziert werden (Schüler:innenmonologe;

unvollständige und vollständige Feedbackloops). Feedbackloops können als dialogisch beschrieben werden und zeichnen sich durch lernförderliche Unterstützungsstrategien der Lehrperson (aktivierende Fragen, z.T. anknüpfendes Feedback) sowie elaborierte Schüler:innenbeiträge aus.

Die Verteilung der beschriebenen Interaktionsmuster unterscheidet sich deutlich zwischen Einzelinteraktionen und Interaktionen der Lehrperson mit mehreren Lernenden. Einzelinteraktionen umfassten mehrheitlich (ca. 85 %) lehrpersonendominierte Interaktionsmuster (Lehrpersonenmonologe; IRE/IRF-Sequenzen). Dagegen lässt sich die Gesprächsführung in Interaktionen der Lehrperson mit mehreren Lernenden als deutlich offener beschreiben: Interaktionsmuster mit höherer Schüler:innenbeteiligung kamen zu gleichen Anteilen (ca. 50 %) wie lehrpersonendominierte Interaktionsmuster vor.

Diese Unterschiede können verschieden interpretiert werden: Studien zum kooperativen Lernen deuten darauf hin, dass Interaktionen zwischen Peers motivierend wirken und die verbale Beteiligung von Lernenden fördern können (Kyndt et al., 2013; Nemeth et al., 2019). Möglicherweise könnte es daher für Lehrpersonen bei Interaktionen mit mehreren Lernenden einfacher sein, an einen bereits bestehenden Dialog anzuknüpfen und so Feedbackloops zu realisieren. Des Weiteren könnten die Ergebnisse auf systematische Zusammenhänge zwischen den Sozialformen in Arbeitsphasen (Einzelarbeit versus Partner:innen- und Gruppenarbeit)⁶, als einer Oberflächenstruktur des Unterrichts, und dem tiefenstrukturellen Merkmal der Lernunterstützung hindeuten. Die Bedeutung des Zusammenhangs von Oberflächen- und Tiefenstrukturen wurde bereits in früheren Arbeiten hervorgehoben (Decristan et al., 2020; Hess & Lipowsky, 2020). Möglicherweise kommt den Sozialformen in Arbeitsphasen eine unterschiedliche Funktion für den Lernprozess zu. Es könnte beispielsweise sein, dass Lehrpersonen Partner:innen- und Gruppenarbeitsphasen eher für die selbstständige Exploration unbekannter, herausfordernder Fragestellungen konzipieren, während Einzelarbeitsphasen eher für das Üben, Festigen und Automatisieren zuvor erarbeiteter Inhalte geplant werden. Zu ähnlichen Überlegungen gelangten Hess und Lipowsky (2020) im Hinblick auf Unterschiede, die sie im Aktivierungsniveau von Fragen im Klassenunterricht versus der Arbeitsphase fanden.

Entlang dieser Interpretationen ergeben sich Implikationen für die Forschung: Beispielsweise sind videobasierte Ratings ein gängiges Verfahren zur Beurteilung der Qualität von Aspekten der Lernunterstützung (z. B. Feedback als Teil formativen Assessments). Einige weit verbreitete Ratinginstrumente (z. B. CLASS) unterteilen den Unterricht in Zeiteinheiten und berücksichtigen Oberflächenstrukturen nicht (Praetorius & Charalambous, 2018). Die vorliegenden Befunde legen allerdings nahe, dass Arbeitsphasen nicht als eine strukturell homogene Unterrichtsphase betrachtet werden können, da die Lernunterstützung in Abhängigkeit von Sozialformen systematisch

6 Ein Rückschluss auf die Sozialformen wird als legitim angenommen, weil Einzelinteraktionen fast ausschließlich in Phasen der Einzelarbeit beobachtet wurden und Interaktionen der Lehrperson mit mehreren sich weitestgehend mit Phasen der Partner:innen- und Gruppenarbeit decken.

variiert. Bei Analysen von Arbeitsphasen nach Sozialformen zu differenzieren, wäre somit angemessen. In zukünftiger Forschung müsste zudem vertiefend geklärt werden, inwieweit eine qualitätsvolle Lernunterstützung innerhalb unterschiedlicher Sozialformen, wie bislang angenommen, in derselben Art und Weise realisiert werden muss. Es ist beispielsweise denkbar, dass die gefundene Varianz auf legitimen didaktischen Entscheidungen von Lehrpersonen basiert. Eine hohe Unterstützungsqualität könnte dann ggf. in verschiedenen Phasen des Unterrichts nicht äquivalent sein, also könnte folglich je nach Sozialform unterschiedliche Merkmale aufweisen.

Auch für die Aus- und Weiterbildung von Lehrpersonen lassen sich Implikationen ableiten: Einerseits können Lehrpersonen für eine stärker dialogische Ausrichtung von Lernunterstützung sensibilisiert werden. Beispielsweise könnte ein Training von Strategien des formativen Assessments zum Anregen von Schüler:innenäußerungen (Eliciting Evidence) und lernförderlichen Anknüpfungsmöglichkeiten (Feedback), das Auftreten von Feedbackloops in Arbeitsphasen fördern. Implementierungsstudien deuten darauf hin, dass durch Trainings von Strategien des formativen Assessments Erfolge bei Lehrpersonen erzielt werden können (Hondrich et al., 2016). Gemäß den vorliegenden Ergebnissen könnten sich Partner:innen- und Gruppenarbeitsphasen als Einstieg in solche Trainings anbieten. Andererseits sollten Zusammenhänge zwischen der didaktisch-inhaltlichen Unterrichtsplanung und der Interaktionsebene bzw. -qualität des Unterrichts beleuchtet werden. Hier stellt sich beispielsweise die Frage nach Voraussetzungen einer gelingenden dialogischen Unterstützung: Welche Aufgaben und Lernsettings bieten sich an, um eine dialogische Lernunterstützung zu stärken? Wie können verschiedene Sozialformen eingesetzt werden, um diese zu realisieren? Grundsätzlich bietet sich für die Reflexion der eigenen Unterrichtspraxis zur professionellen Weiterentwicklung die Arbeit mit Unterrichtsvideos an (z. B. Krammer & Hugener, 2014; Krammer, 2020).

5.2 Unterschiede zwischen den kognitiv-motivationalen Profilen

Vergleiche zwischen den kognitiv-motivationalen Profilen zeigen vor allem Kontraste bezüglich der Unterstützung von Schüler:innen mit einem starken Profil und einem überforderten Profil in Einzelarbeitsphasen. Lernende mit einem starken Profil scheinen insgesamt häufiger nur beobachtet zu werden, ohne dass eine verbale Interaktion mit der Lehrperson stattfindet, und erhalten vergleichsweise oft kurze Inputs von der Lehrperson (Lehrpersonenmonologe), wohingegen sie signifikant seltener enggeführt werden (IRE/F-Sequenzen). Gleichzeitig ist eine Tendenz zu erkennen, dass starke Profile häufiger an dialogischen Interaktionen beteiligt sind (unvollständige und vollständige Feedbackloops). Im Vergleich zum starken Profil werden Schüler:innen mit einem überforderten Profil mehrheitlich enggeführt (IRE/F-Sequenzen). Unvollständige und vollständige Feedbackloops konnten hingegen bei diesem Profil fast nie beobachtet werden. Obwohl sich auch Unterschiede zum motivierten und unmotivierten Profil zeigen, sind diese weniger akzentuiert. Eine Erklärung dafür könnte sein, dass

es Lehrpersonen gemäß den Ergebnissen von Südkamp et al. (2018) schwerer zu fallen scheint, inkonsistente Profile zu erkennen. Inkonsistente Profile beschreiben die Autorinnen als divergent ausgeprägte kognitive und motivationale Voraussetzungen. Dies trifft auf das unmotivierte Profil (geringe Motivation bei mittlerem Vorwissen) und das motivierte Profil (hohe Motivation bei geringem Vorwissen) zu. Es könnte somit angenommen werden, dass Lehrpersonen ihr Verhalten eher am Vorwissen bzw. der Leistungsfähigkeit von Lernenden orientieren und motivationale Voraussetzungen weniger stark einbeziehen.

Die Ergebnisse veranschaulichen insgesamt eine Tendenz zum differenziellen Handeln der Lehrperson in Einzelinteraktionen mit unterschiedlichen kognitiv-motivationalen Profilen. Die Ergebnisse werfen jedoch Fragen für die zukünftige Forschung auf: Inwieweit ist das beobachtete differenzielle Handeln der Lehrpersonen vorteilhaft, also adaptiv, oder nachteilig für die entsprechenden Schüler:innengruppen? Forschungsbefunde zum Feedback legen nahe, dass leistungsschwache Lernende von eher direktivem, unmittelbarem und korrigierendem Feedback profitieren und leistungsstarke Lernende eher von zurückhaltendem, indirektem Feedback (Fyfe et al., 2022; Shute, 2008). Diese Befunde beziehen die Motivation von Lernenden nicht ein, könnten jedoch ein Indiz sein, dass IRE/F-Sequenzen bei Lernenden mit überfordertem Profil häufiger vorkommen, weil sie evtl. enggeführtes Scaffolding beinhalten. Wäre dies der Fall, könnte es als adaptives Verhalten von Lehrpersonen interpretiert werden. Um diese Annahme zu überprüfen, müssten in zukünftigen Untersuchungen der Lernzuwachs von Schüler:innen einbezogen und IRE/F-Sequenzen ggf. weiter ausdifferenziert werden.

Zusammengefasst legen die Befunde nahe, dass Schüler:innen mit verschiedenen kognitiven und motivationalen Voraussetzungen nicht dieselbe Lernunterstützung von Lehrpersonen während Arbeitsphasen erhalten. Daher muss hinterfragt werden, inwieweit die Qualität von Lernunterstützungsmerkmalen über Stichproben hinweg untersucht und allgemeine Schlussfolgerungen zum Lernzuwachs von „Standardschüler:innen“ gezogen werden können. Mit Ratings wird beispielsweise die Qualität der Lernunterstützung oftmals gesamthaft für eine Unterrichtsstunde (oder Ausschnitte davon) bestimmt. Wenn Lernende jedoch in den beobachteten Sequenzen nicht dieselbe Lernunterstützung erhalten, kann von einer gesamthaft erfassten Qualität der Lernunterstützung nicht auf individuelle Lernzuwächse geschlossen werden. Gemäß den vorliegenden Ergebnissen wären daher zukünftig differenzielle Analysen zur Wirkung von Lernunterstützung notwendig, welche die kognitiven und motivationalen Voraussetzungen von Lernenden bereits bei der Erfassung von Lernunterstützungsmerkmalen einbeziehen. Insgesamt heben die vorliegenden Befunde im Sinne des Angebots-Nutzungs-Modells hervor, dass multiple Faktoren bei der Untersuchung von Lernunterstützung einbezogen werden sollten und Schüler:innenfaktoren eine zentrale Bedeutung bei der Lernunterstützung zukommen.

5.3 Limitationen und weiterer Forschungsbedarf

Obwohl mit der vorliegenden Untersuchung die Bedeutung von Schüler:innenfaktoren bei der Lernunterstützung beleuchtet werden konnte, haben die Auswertungen einen explorativen Charakter. Folglich müssen sie durch weitere Untersuchungen abgesichert werden. Beispielsweise wurde die Leistungsentwicklung von Schüler:innen mit unterschiedlichen Profilen nicht erfasst, sodass mit den vorliegenden Befunden keine direkten Rückschlüsse über die Wirkung der Interaktionsmuster auf Leistungen möglich sind. In Anlehnung an existierende Befunde wurde angenommen, dass Feedback-loops ein lernförderliches, adaptives Interaktionsmuster darstellen (z. B. Ruiz-Primo & Furtak, 2007; Webb et al., 2014), während IRE/F-Sequenzen sich im Vergleich dazu deaktivierend und weniger lernförderlich auswirken (z. B. Decristan et al., 2020; Ufer et al., 2015; Reusser & Pauli, 2013). Wie im vorherigen Kapitel zum Aspekt der Adaptivität diskutiert wurde, bestehen allerdings Gründe, diese Annahmen im Hinblick auf Arbeitsphasen zu hinterfragen. Zukünftige Forschung würde daher davon profitieren, die Wirkung von Interaktionsmustern im Zusammenhang mit dem Leistungsfortschritt von Schüler:innen differenziell zu untersuchen.

In Anlehnung an die vorangegangene Forschung wurde eine begrenzte Anzahl grundlegender, als relevant angenommener Interaktionsmuster identifiziert. Im Rahmen dieser ersten explorativen Untersuchung wurden dabei nicht zwischen verschiedenen Formen von Feedback differenziert. Möglicherweise könnten weitere Submuster der gefundenen Interaktionsmuster unterschieden werden, indem verschiedene Feedbackformen ausdifferenziert werden. Auf diese Weise ließe sich gegebenenfalls überprüfen, ob enggeführtes Scaffolding in einigen IRE/F-Sequenzen vorkommt.

Durch den Fokus auf die Interaktionsebene und Schüler:innenvoraussetzungen wurden in den vorliegenden Analysen Unterschiede zwischen Lehrpersonen nicht modelliert. In zukünftiger Forschung wäre daher zu klären, inwieweit dem Handeln von Lehrpersonen bewusste didaktische Entscheidungen zugrunde liegen. Auch müsste untersucht werden, inwieweit Partner:innen- und Gruppenarbeiten, in denen häufiger Interaktionsmuster mit stärkerer Schülerbeteiligung festgestellt wurden, von Lehrpersonen kompensatorisch zu engführenden Einzelinteraktionen eingesetzt werden.

5.4 Fazit

Die vorliegende explorative Studie untersuchte die Lernunterstützung in Arbeitsphasen, indem anhand von Schüler:innenbeiträgen und Unterstützungshandlungen von Lehrpersonen Interaktionsmuster identifiziert wurden. Es konnte gezeigt werden, dass die Lernunterstützung nicht nur von Lehrpersonenhandlungen abhängt, sondern Schüler:innenäußerungen ein bedeutsamer Faktor sind, anhand dessen verschiedene Formen lernunterstützender Interaktionen differenziert werden können. Die Ergebnisse verdeutlichen ein großes Potential von Arbeitsphasen für die Realisierung einer dialogischen Lernunterstützung, die Schüler:innen bewusst interaktiv einbezieht und

sie zu Wort kommen lässt. Gleichzeitig beleuchten die Befunde auch die Komplexität von Lernunterstützung in wechselseitigen Interaktionen. Im Sinne des Angebots-Nutzungs-Modells verweisen sie auf die Notwendigkeit differenzierter Analysen, welche auch die kognitiven und motivationalen Voraussetzungen von Lernenden einbeziehen.

Anmerkungen:

Das Projekt wurde vom Schweizerischen Nationalfonds (Nr. 169771) unterstützt.

Literatur

- Bakker, A., Smit, J. & Wegerif, R. (2015). Scaffolding and dialogic teaching in mathematics education: introduction and review. *ZDM Mathematics Education*, 47(7), 1047–1065. <https://doi.org/10.1007/s11858-015-0738-8>
- Boyer, K. E., Vouk, M. A. & Lester, J. C. (2007). The influence of learner characteristics on task-oriented tutorial dialogue. *Frontiers in artificial intelligence and applications*, 158, 365–372.
- Buholzer, A., Baer, M., Zulliger, S., Torchetti, L., Ruelmann, M., Häfliger, A. & Lötscher, H. (2020). Formatives Assessment im alltäglichen Mathematikunterricht von Primarlehrpersonen: Häufigkeit, Dauer und Qualität. *Unterrichtswissenschaft*, 48(4), 629–661. <https://doi.org/10.1007/s42010-020-00083-7>
- Bürgermeister, A. & Saalbach, H. (2018). Theoretischer Beitrag: Formatives Assessment: Ein Ansatz zur Förderung individueller Lernprozesse. *Psychologie in Erziehung und Unterricht*, 65(3), 194. <https://doi.org/10.2378/peu2018.art11d>
- Chi, M. T. (2009). Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in cognitive science*, 1(1), 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T. & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25(4), 471–533. https://doi.org/10.1207/s15516709cog2504_1
- Chin, C. (2006). Classroom interaction in science: Teacher questioning and feedback to students' responses. *International Journal of Science Education*, 28(11), 1315–1346. <https://doi.org/10.1080/09500690600621100>
- Decristan, J., Fauth, B., Heide, E. L., Locher, F. M., Troll, B., Kurucz, C. & Kunter, M. (2020). Individuelle Beteiligung am Unterrichtsgespräch in Grundschulklassen: Wer ist (nicht) beteiligt und welche Konsequenzen hat das für den Lernerfolg? *Zeitschrift für Pädagogische Psychologie*, 34(3-4), 171–186. <https://doi.org/10.1024/1010-0652/a000251>
- Decristan, J., Hess, M., Holzberger, D. & Praetorius, A.-K. (2020). Oberflächen- und Tiefenmerkmale. Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. In A.-K. Praetorius, J. Grünkorn & E. Klieme (Hrsg.), *Zeitschrift für Pädagogik. Beiheft. 66. Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen* (S. 102–116). Beltz Juventa.
- Denn, A.-K. (2021). *Interaktionen von Lehrpersonen mit Mädchen und Jungen im Mathematikunterricht der Grundschule*. Springer. <https://doi.org/10.1007/978-3-658-32372-1>
- Denn, A.-K., Lotz, M., Heinzl, F. & Lipowsky, F. (2016). Interaktion im Mathematikunterricht mit Fokus Geschlecht (IMaGe): Eine videobasierte Studie im Mathematik-

- unterricht des zweiten Schuljahrs. In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler & R. Schmidt (Hrsg.), *Jahrbuch Grundschulforschung: Band 20. Facetten grundschulpädagogischer und -didaktischer Forschung* (S. 207–212). Springer VS. https://doi.org/10.1007/978-3-658-11944-7_33
- Furtak, E. M., Ruiz-Primo, M. A. & Bakeman, R. (2017). Exploring the utility of sequential analysis in studying informal formative assessment practices. *Educational Measurement: Issues and Practice*, 36(1), 28–38. <https://doi.org/10.1111/emip.12143>
- Fyfe, E. R., Borriello, G. A. & Merrick, M. (2022). A developmental perspective on feedback: How corrective feedback influences children's literacy, mathematics, and problem solving. *Educational Psychologist*, 1–16. <https://doi.org/10.1080/00461520.2022.2108426>
- Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1), 104–137. <https://doi.org/10.3102/00028312031001104>
- Haberman, S. J. (1973). The Analysis of residuals in cross-classified tables. *Biometrics*, 29(1), 205. <https://doi.org/10.2307/2529686>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Helmke, A. (2014). Forschung zur Lernwirksamkeit des Lehrerhandelns. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 807–821). Waxmann.
- Helmke, A., Helmke, T., Schrader, F.-W. & Wagner, W. (2007). *Die Basiscodierung der DESI-Videostudie*. Universität Koblenz-Landau. http://andreas-helmke.de/wordpress/wp-content/uploads/2015/01/Codebuch_Basiskodierung-final_Abk.pdf
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. SAGE Publications. <https://doi.org/10.4135/9781452219493>
- Hess, M. & Lipowsky, F. (2020). Zur (Un-)Abhängigkeit von Oberflächen- und Tiefenmerkmalen im Grundschulunterricht. Fragen von Lehrpersonen im öffentlichen Unterricht und in Schülerarbeitsphasen im Vergleich. In A.-K. Praetorius, J. Grünkorn & E. Klieme (Hrsg.), *Zeitschrift für Pädagogik. Beiheft. 66. Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen* (S. 117–131). Beltz Juventa.
- Hondrich, A. L., Hertel, S., Adl-Amini, K. & Klieme, E. (2016). Implementing curriculum-embedded formative assessment in primary school science classrooms. *Assessment in Education: Principles, Policy & Practice*, 23(3), 353–376. <https://doi.org/10.1080/0969594X.2015.1049113>
- Howe, C. & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, 43(3), 325–356. <https://doi.org/10.1080/0305764X.2013.786024>
- IBM Corporation. (2021). *IBM SPSS Complex Samples 28*. https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/de/IBM_SPSS_Complex_Samples.pdf
- Jurik, V., Gröschner, A. & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction*, 23, 33–42. <https://doi.org/10.1016/j.learninstruc.2012.09.002>
- Jurik, V., Gröschner, A. & Seidel, T. (2014). Predicting students' cognitive learning activity and intrinsic learning motivation: How powerful are teacher statements, student profiles, and gender? *Learning and Individual Differences*, 32, 132–139. <https://doi.org/10.1016/j.lindif.2014.01.005>
- Jurik, V., Häusler, J., Stubben, S. & Seidel, T. (2015). Interaction. Erste Ergebnisse einer vergleichenden Videostudie im Deutsch- und Mathematikunterricht [Interaction. First

- results of a comparative video study carried out in language and mathematics instruction]. *Zeitschrift für Pädagogik*, 61(5), 692–711. <https://doi.org/10.25656/01:15422>
- Kobarg, M. & Seidel, T. (2007). Prozessorientierte Lernbegleitung: Videoanalysen im Physikunterricht der Sekundarstufe I. *Unterrichtswissenschaft*, 35(2), 146–168.
- Krammer, K. (2009). *Individuelle Lernunterstützung in Schülerarbeitsphasen: Eine videobasierte Analyse des Unterstützungsverhaltens von Lehrpersonen im Mathematikunterricht*. Waxmann.
- Krammer, K. (2016). Lehren als Unterstützung des Lernens: Hilfreiche Formen der Lernunterstützung nach dem Prinzip der kognitiven Meisterlehre. *Friedrich Jahresheft 2016 zum Thema „Lehren“*, 76–77.
- Krammer, K. (2020). Videos in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 691–699). utb.
- Krammer, K. & Hugener, I. (2014). Förderung der Analysekompetenz angehender Lehrpersonen anhand von eigenen und fremden Unterrichtsvideos. *Journal für LehrerInnenbildung*, 14(1), 25–32.
- Krammer, K., Reusser, K. & Pauli, C. (2010). Individuelle Unterstützung der Schülerinnen und Schüler durch die Lehrperson während der Schülerarbeitsphasen. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 107–122). Waxmann.
- Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E. & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review*, 10, 133–149. <https://doi.org/10.1016/j.edurev.2013.02.002>
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Lazarides, R., Dicke, A.-L., Rubach, C. & Eccles, J. S. (2020). Profiles of motivational beliefs in math: Exploring their development, relations to student-perceived classroom characteristics, and impact on future career aspirations and choices. *Journal of Educational Psychology*, 112(1), 70–92. <https://doi.org/10.1037/edu0000368>
- Lazarides, R., Dietrich, J. & Taskinen, P. H. (2019). Stability and change in students' motivational profiles in mathematics classrooms: The role of perceived teaching. *Teaching and Teacher Education*, 79, 164–175. <https://doi.org/10.1016/j.tate.2018.12.016>
- Linnenbrink-Garcia, L., Pugh, K. J., Koskey, K. L. K. & Stewart, V. C. (2012). Developing conceptual understanding of natural selection: The role of interest, efficacy, and basic prior knowledge. *The Journal of Experimental Education*, 80(1), 45–68. <https://doi.org/10.1080/00220973.2011.559491>
- Lipowsky, F., Pauli, C. & Rakoczy, K. (2008). Schülerbeteiligung und Unterrichtsqualität. In M. Gläser-Zikuda & J. Seifried (Hrsg.), *Lehrerexpertise: Analyse und Bedeutung unterrichtlichen Handelns* (S. 67–90). Waxmann.
- Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K. & Klieme, E. (2007). Gleicher Unterricht – gleiche Chancen für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht. *Unterrichtswissenschaft*, 35(2).
- Lotz, M. (2016). *Kognitive Aktivierung im Leseunterricht der Grundschule: Dokumentation der Beobachtungsinstrumente*. Anhang zur Dissertation. Springer Fachmedien. <https://doi.org/10.1007/978-3-658-10436-8>
- Lotz, M. & Lipowsky, F. (2014). Jedem das Seine oder allen das Gleiche? Individuelle Lernunterstützung in Schülerarbeitsphasen im Leseunterricht der Grundschule. In B.

- Kopp, S. Martschinke, M. Munser-Kiefer, M. Haider, E.-M. Kirschhock, G. Ranger & G. Renner (Hrsg.), *Individuelle Förderung und Lernen in der Gemeinschaft* (S. 178–181). Springer. https://doi.org/10.1007/978-3-658-04479-4_28
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Harvard Univ. Press. <https://doi.org/10.4159/harvard.9780674420106>
- Nemeth, L., Denn, A.-K., Hirstein, A. & Lipowsky, F. (2019). Interaktionen von Schülerinnen in kooperativen Lernsituationen. In K. Verrière & L. Schäfer (Hrsg.), *Interaktion im Klassenzimmer: Forschungsgeleitete Einblicke in das Geschehen im Unterricht* (S. 51–73). Springer. https://doi.org/10.1007/978-3-658-23173-6_4
- Park, M., Yi, M., Flores, R. & Nguyen, B. (2020). Informal formative assessment conversations in mathematics: Focusing on preservice teachers' initiation, response and follow-up sequences in the classroom. *EURASIA Journal of Mathematics, Science & Technology Education*, 16(10), em1884. <https://doi.org/10.29333/ejmste/8436>
- Pauli, C. (2012). Kodierende Beobachtung. In H. de Boer & S. Reh (Hrsg.), *Beobachtung in der Schule – Beobachten lernen* (S. 45–63). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-18938-3_3
- Pauli, C., Drollinger-Vetter, B., Hugener, I. & Lipowsky, F. (2008). Kognitive Aktivierung im Mathematikunterricht. *Zeitschrift für Pädagogische Psychologie*, 22(2), 127–133. <https://doi.org/10.1024/1010-0652.22.2.127>
- Pauli, C. & Lipowsky, F. (2007). Mitmachen oder Zuhören? Mündliche Schülerinnen- und Schülerbeteiligung im Mathematikunterricht. *Unterrichtswissenschaft*, 35(2), 101–124.
- Pauli, C. & Reusser, K. (2015). Discursive cultures of learning in (everyday) mathematics teaching: A video-based study on mathematics teaching in German and Swiss classrooms. In C. S. C. Asterhan, L. B. Resnick & S. N. Clarke (Hrsg.), *Socializing intelligence through academic talk and dialogue* (S. 181–193). American Educational Research Association. https://doi.org/10.3102/978-0-935302-43-1_14
- Pfister, M., Moser Opitz, E. & Pauli, C. (2015). Scaffolding for mathematics teaching in inclusive primary classrooms: a video study. *ZDM Mathematics Education*, 47(7), 1079–1092. <https://doi.org/10.1007/s11858-015-0713-4>
- Pielmeier, M., Huber, S. & Seidel, T. (2018). Is teacher judgment accuracy of students' characteristics beneficial for verbal teacher-student interactions in classroom? *Teaching and Teacher Education*, 76, 255–266. <https://doi.org/10.1016/j.tate.2018.01.002>
- Pohlmann-Rother, S. & Kürzinger, A. (2015). Individuelle Lernunterstützung in Schülerarbeitsphasen im Schreibunterricht der Grundschule. In K. Liebers, B. Landwehr, A. Marquardt & K. Schlotter (Hrsg.), *Lernprozessbegleitung und adaptives Lernen in der Grundschule: Forschungsbezogene Beiträge* (S. 181–186). Springer. https://doi.org/10.1007/978-3-658-11346-9_22
- Pohlmann-Rother, S., Kürzinger, A. & Lipowsky, F. (2018). Individuelle Lernunterstützung im schriftsprachlichen Anfangsunterricht: Formen, Verteilungsmuster und Wirksamkeit. *Zeitschrift für Grundschulforschung*, 11(2), 315–332. <https://doi.org/10.1007/s42278-018-0024-2>
- Pohlmann-Rother, S., Kürzinger, A. & Lipowsky, F. (2020). Feedback im Anfangsunterricht der Grundschule – Eine Videostudie zum Feedbackverhalten von Lehrpersonen in der Domäne Schreiben. *Zeitschrift für Erziehungswissenschaft*, 23(3), 591–611. <https://doi.org/10.1007/s11618-020-00950-0>
- Praetorius, A.-K. & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM Mathematics Education*, 50(3), 535–553. <https://doi.org/10.1007/s11858-018-0946-0>
- Rao, J. N. K. & Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables.

- Journal of the American Statistical Association*, 76(374), 221–230. <https://doi.org/10.1080/01621459.1981.10477633>
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht. Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 9–32). Waxmann.
- Reusser, K. & Pauli, C. (2013). Verständnisorientierung in Mathematikstunden erfassen. Ergebnisse eines methodenintegrativen Ansatzes. *Zeitschrift für Pädagogik*, 59(3), 308–335. <https://doi.org/10.25656/01:11940>
- Rimm-Kaufman, S. E., Baroody, A. E., Larsen, R. A. A., Curby, T. W. & Abry, T. (2015). To what extent do teacher–student interaction quality and student gender contribute to fifth graders’ engagement in mathematics learning? *Journal of Educational Psychology*, 107(1), 170–185. <https://doi.org/10.1037/a0037252>
- Ruelmann, M., Torchetti, L., Zulliger, S., Buholzer, A. & Praetorius, A.-K. (2021). Kognitiv-motivationale Schüler*innenprofile und ihre Bedeutung für die Schüler*innenwahrnehmung der Lernunterstützung durch die Lehrperson. *Unterrichtswissenschaft*, 49(3), 395–422. <https://doi.org/10.1007/s42010-021-00100-3>
- Ruiz-Primo, M. A. & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers’ practices and student learning. *Educational Assessment*, 11(3-4), 205–235. <https://doi.org/10.1080/10627197.2006.9652991>
- Ruiz-Primo, M. A. & Furtak, E. M. (2007). Exploring teachers’ informal formative assessment practices and students’ understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84. <https://doi.org/10.1002/tea.20163>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/BF00117714>
- Schnebel, S. (2017). Unterstützungshandeln von Lehrpersonen und dessen Beziehung zu Lernausgangslagen und Lernzuwächsen von Schülerinnen und Schülern in einem naturwissenschaftlichen Lernangebot in der Grundschule. In F. Heinzl & K. Koch (Hrsg.), *Individualisierung im Grundschulunterricht* (S. 121–125). Springer. https://doi.org/10.1007/978-3-658-15565-0_19
- Schnitzler, K., Holzberger, D. & Seidel, T. (2021). All better than being disengaged: Student engagement patterns and their relations to academic self-concept and achievement. *European Journal of Psychology of Education*, 36(3), 627–652. <https://doi.org/10.1007/s10212-020-00500-6>
- Seidel, T. (2006). The role of student characteristics in studying micro teaching–learning environments. *Learning Environments Research*, 9(3), 253–271. <https://doi.org/10.1007/s10984-006-9012-x>
- Seidel, T. & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction*, 16(3), 228–240. <https://doi.org/10.1016/j.learninstruc.2006.03.002>
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Südkamp, A., Praetorius, A.-K. & Spinath, B. (2018). Teachers’ judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education*, 76, 204–213. <https://doi.org/10.1016/j.tate.2017.09.016>
- Ufer, S., Heinze, A. & Lipowsky, F. (2015). Unterrichtsmethoden und Instruktionsstrategien. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Wei-

- gand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 411–434). Springer. https://doi.org/10.1007/978-3-642-35119-8_15
- van de Pol, J., Volman, M. & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271–296. <https://doi.org/10.1007/s10648-010-9127-6>
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M. & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts: Ein kritischer Vergleich verschiedener Modellvarianten. *Zeitschrift für Pädagogik*, 62(1), 63–80. <https://doi.org/10.3262/ZPB2001063>
- Walshaw, M. & Anthony, G. (2008). The teacher’s role in classroom discourse: A review of recent research into Mathematics classrooms. *Review of Educational Research*, 78(3), 516–551. <https://doi.org/10.3102/0034654308320292>
- Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N. & Turrou, A. C. (2014). Engaging with others’ mathematical ideas: Interrelationships among student participation, teachers’ instructional practices, and learning. *International Journal of Educational Research*, 63, 79–93. <https://doi.org/10.1016/j.ijer.2013.02.001>
- Wirtz, M. A. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Hogrefe.

Formatives Feedback und Rubrics fördern mathematisches Argumentieren in der Primarschule

Robbert Smit¹ | Kurt Hess² | Patricia Bachmann¹ | Heidi Dober²

¹Pädagogische Hochschule St. Gallen | ²Pädagogische Hochschule Zug

Abstract: Mathematisches Argumentieren ist eine anspruchsvolle Kompetenz für Schüler:innen. Obwohl in neueren Bildungsstandards enthalten, wird das mathematische Argumentieren in den Klassenzimmern nur selten praktiziert. Die effektive Unterstützung von Schüler:innen während des Prozesses des Argumentierens ist eine Herausforderung für Lehrkräfte. Die Anwendung von formativem Feedback, allenfalls unterstützt mit einem Rubric (Beurteilungsraster), kann die Wirksamkeit der Lernbegleitung erhöhen. Es wurde im Rahmen einer mehrwöchigen Unterrichtsreihe geprüft, ob das von den Schüler:innen aus 71 Primarschulklassen ($N = 1261$) wahrgenommene formative Feedback der Lehrperson die Entwicklung des mathematischen Argumentierens unterstützte. Die erwarteten Zusammenhänge wurden mit einer Mehrebenenanalyse untersucht. Auf der Klassenebene sagte die Häufigkeit formativen Feedbacks die Kompetenzentwicklung im mathematischen Argumentieren voraus, auf der individuellen Ebene jedoch nicht. Zudem erteilten Lehrpersonen, die mit dem Rubric arbeiteten, häufiger formatives Feedback. Wir diskutieren Erklärungen für diese Ergebnisse und stellen Implikationen für das Unterrichten vor.

1 Einleitung

Mit der Einführung des Deutschschweizer Lehrplans 21 sollen mathematische Handlungsaspekte wie Erforschen und Argumentieren stärker in den unterrichtlichen Fokus rücken. Allerdings zeigt sich, dass entsprechende Tätigkeiten im Schulunterricht noch relativ selten vorkommen und diese für die Lehrpersonen auch schwierig zu organisieren und zu begleiten sind (Cramer & Knipping, 2018). Beim mathematischen Argumentieren spielen – insbesondere im Primarschulbereich – gute Aufgaben und Interaktionsprozesse zwischen Lernenden eine zentrale Rolle (Fetzer, 2011). Gute Aufgaben zur Erweiterung und Differenzierung der Argumentationskompetenz sollen Lernenden Freiräume für verschiedene Vorgehensweisen und unterschiedliche Darstellungen eröffnen. Aus didaktisch-methodischer Sicht eignen sich zudem kooperative oder dialogische Lernformen, in welchen Schülerinnen und Schüler ihre Lösungswege, Darstellungen und Antworten beschreiben, erklären oder begründen (Stein, Grover & Henningsen, 1996). Das Gelingen solcher anspruchsvollen Lernprozesse hängt massgeblich davon ab, ob und wie die Lehrperson die Schülerinnen und Schüler mit Fragen, Hinweisen oder Impulsen bzw. mit unterstützendem Feedback als Teil des formati-

ven Assessments (formative Beurteilung) begleitet (Herbert et al., 2022; Ruiz-Primo & Brookhart, 2018).

Es ist naheliegend und notwendig, unterrichtstaugliche Instrumente zur Optimierung der formativen Beurteilung des mathematischen Argumentierens zu entwickeln und deren Wirkung empirisch zu überprüfen (Schmidt, 2020). Diese Instrumente sollen Lehrpersonen und Schülerinnen und Schüler helfen, den Lernstand zu beobachten und einzuschätzen. Weiter sollen sie die Kriterien klären, auf welchen das Feedback beruht und bei der Steuerung des weiteren Vorgehens unterstützen. Rubrics (Beurteilungsraster) können solche Ansprüche insofern erfüllen, als sie z. B. die Möglichkeit bieten, Lernergebnisse entlang von inhaltlichen Kriterien und abgestuften Indikatoren einzuschätzen und davon nächste Lernschritte abzuleiten (Herbert et al., 2022). In der vorliegenden Studie soll die Wirkung des formativen Feedbacks auf die Entwicklung des mathematischen Argumentierens bei Schülerinnen und Schüler der oberen Primarstufen (5. und 6. Klasse) geprüft werden, wobei ein Teil der Lehrpersonen einen Rubric zur Unterstützung des Feedbacks verwendete.

2 Theoretischer Hintergrund

2.1 Formatives Feedback

Feedback wird spätestens seit Hatties (2008) umfangreichen Meta-Analysen zu den einflussreichsten Faktoren für den schulischen Lernerfolg gezählt, obschon deren Lernwirksamkeit von einigen differenzierten Kriterien abhängt, wie etwa dem Zeitpunkt oder der Ausführlichkeit des Feedbacks, der Komplexität der Aufgabenstellung oder den persönlichen Lernvoraussetzungen usw. (Kluger & DeNisi, 1996; Shute, 2008). Gemäss Ruiz-Primo und Brookhart (2018) ist ein formatives Feedback eine Unterstützung, welche über einen schriftlichen oder mündlichen Kommentar hinausgeht und die Form eines Lehr-Lerndialogs zwischen Lehrperson und Schüler:in annehmen kann, mit dem Ziel, den Lernprozess voranzubringen. Die Lehrperson bietet – im Sinne der cognitive apprenticeship (Collins et al., 1989) – stärkere und verbindlichere oder zurückhaltendere und offenere Hinweise für die weitere Bearbeitung der Aufgaben an. Die Rückmeldung erfolgt in einem Feedback-Dialog zwischen Schüler:in und Lehrperson im Sinne des sozio-konstruktivistischen Lernens (Ruf & Gallin, 1998; Vygotsky, 1974). Damit ist gemeint, dass die Lernenden eigenständig handelnd und auf graduell selbstregulierten Wegen in reichhaltigen Lernangeboten Kompetenzen erwerben (Reusser, 2009).

Formatives Feedback ist ein zentraler Bestandteil des Prozesses des formativen Assessments (Black & Wiliam, 2009) wie dargestellt in Abbildung 1. In der Schweiz wird für den englischen Begriff „formative assessment“ seit den neunziger Jahren der Begriff „formative Beurteilung/évaluation formative“ in Anlehnung an Perrenoud (1991) und Allal (1991) verwendet. Dabei wird die formative Beurteilung als Teil des didakti-

schen Planungshandelns angesehen. Im Rahmen dessen werden die Lernziele geklärt, der Lernprozess der Schülerinnen und Schüler durch die Lehrperson beobachtet und Schwierigkeiten diagnostiziert. Die Lehrperson bestimmt den Lernstand bezüglich der Lernziele und bietet danach unterstützendes Feedback ausgerichtet an diesen Zielen an (Hattie & Timperley, 2007; Wiliam & Thompson, 2007). In einer Studie zeigte sich, dass viele Lehrpersonen eine formative Beurteilung im Mathematikunterricht durchführen, diese jedoch mit Ausnahme von Feedback zeitlich eher kurz ausfällt; das Besprechen von Lernzielen oder die Selbstbeurteilung kommen kaum vor (Buholzer et al., 2020). Geschieht das Rückmelden auf eine lernförderliche Weise – bspw. mit Hilfe eines Rubrics – zeigte sich in einer weiteren Studie nicht nur eine positive Wirkung der formativen Beurteilung auf das fachliche Lernen, sondern auch auf die intrinsische Motivation (Harlen, 2006; Moss & Brookhart, 2010).



Abbildung 1: Zusammenhang zwischen Rubric, Beurteilung/Diagnose und Feedback als Teil des formativen Assessments (in Anlehnung an Black & Wiliam (2009))

2.2 Rubrics

Rubrics (Beurteilungsraster) können den Lernprozess mit einfachen, klaren Beschreibungen von Lernzielen leiten und die Lehrperson beim Beurteilen und Erteilen von Feedback bei komplexeren Aufgaben wie Schreibanlässen, mündlichen Präsentationen oder naturwissenschaftlichen Experimenten unterstützen (de Leeuw, 2016; Tang et al., 2015). Rubrics können Lernenden helfen, die Qualität des Lernens an transparenten Erwartungen auszurichten, d. h. eigene Lernergebnisse entlang von inhaltlichen Kriterien und abgestuften Indikatoren einzuschätzen und folglich, Lernprozesse entsprechend selbst zu steuern.

Zum Einsatz von Rubrics existiert international gesehen einiges an unterrichtspraktischer Literatur (Brookhart, 2013; Danielson, 1997), aber eher wenige empirische Erkenntnisse. Die meisten Studien konzentrierten sich auf die Hochschulstufe. Übersichten zur Wirkung von Rubrics im Unterricht finden sich beispielsweise bei Brookhart und Chen (2015), Panadero und Jonsson (2013) und spezifisch für unsere Zielgruppe (5. und 6. Jahrgangsstufe) bei English et al. (2022). In der Literaturübersicht von English et al. (2022) wird z. B. deutlich, dass Rubrics in der Klärung der Feedbackbotschaft oder dem Verstehen der Lernziele Wirkung zeigen. Werden die Schülerinnen und Schüler bei der Erstellung von Rubrics zudem miteinbezogen, zeigen sich verstärkt positive Effekte auf die Leistung (Ghaffar et al., 2020). Ausserdem wird damit die entscheidende Rolle der Schülerinnen und Schüler bei der Nutzung von Feedback anerkannt, da die Rückmeldung nur dann wirksam sein kann, wenn sich die Schülerinnen und Schüler aktiv damit befassen (Brooks et al., 2021). Kritische Hinweise bei der Nutzung von Rubrics finden sich für die teilweise fehlende Reliabilität und Validität bei der Bewertung von Leistungen (Cho et al., 2006; Gregori-Giralt & Menéndez-Varela, 2018; Meier et al., 2006). Letzteres spielt insbesondere dann eine Rolle, wenn der Rubric eine summative Funktion erhält.

2.3 Mathematisches Argumentieren

Mathematisches Argumentieren und Begründen gilt als bedeutsam für einen Unterricht, der das Verstehen ins Zentrum rückt bzw. Lernanlässe zur individuellen und sozialen Annäherung an mathematische Konzepte anbietet (Thompson & Schultzer-Ferrel, 2008). Mathematisches Argumentieren ist ein sozio-konstruktiver Prozess, bei dem Wissen entsteht, indem Lernende Behauptungen und Argumente austauschen und verifizieren (Ball & Bass, 2000). In diesem Prozess ist es gemäss Ball und Bass (2000) Aufgabe der Lehrperson, diesen Austausch zu leiten, klärende Fragen zu stellen, verschiedene Lösungen einander gegenüberzustellen und mathematische Sprachweisen für den Austausch zu fördern. In der vorliegenden Studie wird analog des HarmoS-Kompetenzmodells (Linneweber-Lammerskitten et al., 2010) nicht zwischen Argumentieren und Begründen unterschieden. Argumentieren beinhaltet in der Primarstufe etwa das Erklären von Vorgehensweisen, das Behaupten und Prüfen sowie das Vorhersagen und Verallgemeinern (Bezold, 2009). In der Grundschule und Sekundarstufe 1 erfolgt das mathematische Argumentieren beim Erforschen von Mustern und Strukturen oder beim Beschreiben von Beziehungen oder Zusammenhängen. Ende der 6. Primarschulklasse sollten Schülerinnen und Schüler einfache Aussagen durch Nachprüfen an einem konkreten Beispiel, durch Nutzen vorhandener Daten oder durch naheliegende Argumente begründen oder falsifizieren können (Linneweber-Lammerskitten et al., 2010). In Anlehnung an Bezold (2010) sehen wir das Beschreiben z. B. von Zahl- und Rechenmuster sowie das Begründen (warum etwas gilt) als das eigentliche Argumentieren an. Vorstufen des Argumentationsprozesses wie etwa das Verstehen des Sachverhaltes (Problems), das Anwenden von Strategien und das Durchführen von

Operationen stellen aber notwendige Schritte dar, um Schlussfolgerungen ziehen und damit den Gedankengang des Argumentierens abschliessen zu können (Lithner, 2000).

Lernanlässe zum mathematischen Argumentieren setzen gute Aufgaben voraus, die je nach Bedingungen und Möglichkeiten der Lernenden unterschiedlich komplexe und abstrakte Lösungswege zulassen sowie den Austausch unter Lernenden einfordern (Stein et al., 1996). Dies erfolgt vorzugsweise in kooperativen Lernanlässen, die der Lehrperson Gelegenheit bieten, mehr über die Denkprozesse der Kinder zu erfahren und darauf Einfluss zu nehmen (Ginsburg, 2009; Sfard, 2001). Die unterrichtliche Umsetzung des mathematischen Argumentierens erfordert von Lehrpersonen anspruchsvolle fachliche und auch fachdidaktische Kompetenzen, wie z. B. das Leiten von Diskursen und das Unterstützen bei der Ideenentwicklung der Lernenden (Blanton & Kaput, 2005). Lehrpersonen berichten jedoch von mangelnden zeitlichen Ressourcen, um längere mathematische Sachkontexte mit Gruppen von Schülerinnen und Schüler zu erforschen, die Ergebnisse schriftlich festzuhalten und diese dann zu diskutieren (Keiser & Lambdin, 1996). Zudem bekunden Kinder mit einem fremdsprachigen Hintergrund grössere Schwierigkeiten mit problemhaltigen Textaufgaben, wie sie z. B. beim mathematischen Argumentieren vorliegen, als diejenigen Kinder mit Deutscher Muttersprache (Kempert et al., 2011; Stanat & Christensen, 2006). Trotzdem weisen Kroesbergen und Van Luit (2003) ungeachtet möglicher Stolpersteine schwächerer Kinder eigenständiges Arbeiten im Bereich des Problemlösens als effektiver als direkte Instruktion aus. Als Unterstützung in selbstregulierten Arbeitsphasen können den Lernprozess stützende Werkzeuge eingesetzt werden. Dazu gehören ein formativ genutztes Feedback und Rubrics (Herbert et al., 2022; Turner, 2014).

Forschungsfragen

Es ergeben sich somit die folgenden Fragen, denen in der vorliegenden Studie aufgrund von Fragebogendaten von Schülerinnen und Schüler nachgegangen wurde:

- (1) Hat das von den Schülerinnen und Schüler wahrgenommene formative Feedback der Lehrperson einen Einfluss auf die Kompetenzentwicklung im mathematischen Argumentieren und
- (2) unterstützt ein kriteriengeleiteter Rubric diese Kompetenzentwicklung?

3 Methode

3.1 Forschungsdesign und Stichprobe

Die dem Beitrag zugrunde liegende Studie war Teil des Projektes „FEMAR – Formatives Feedback zum mathematischen Argumentieren“. Das dreijährige Projekt mit Beginn 2018 wurde von den Pädagogischen Hochschulen (PH) St. Gallen und Zug durchgeführt. Das quasi-experimentelle, längsschnittliche Forschungsdesign bestand aus zwei

Erhebungszeitpunkten. Datengrundlage für die vorliegende Studie sind Fragebogen sowie Leistungstests der Schülerinnen und Schüler. Durch die Anwendung eines analytischen Verfahrens, welches zwischen der Wahrnehmung von Individuum und Klasse differenziert, lässt sich ein gemeinsam geteilter Anteil an Unterrichtsqualität aus Sicht der Schüler:innen bestimmen. Diese geteilte Einschätzung von Unterrichtsqualität auf der Ebene der Lehrperson erlaubt wiederum zwischen solcher verschiedener Lehrpersonen unterscheiden (Ditton, 2002).

Tabelle 1: Forschungsdesign

T1 Test & Fragebogen August 2015/2019	Weiterbildung und 9 Lektionen praktische Umsetzung 2015/2019	T2 Test & Fragebogen November 2015/2019	Evaluation und Weiterbildung November 2015/2019
Fragebogen Schüler:innen	Lehrpersonen nehmen Teil an einer Weiterbildung zu - Mathematisches Argumentieren - Feedback - <i>Interventionsgruppe: Rubrics</i>	Fragebogen Schüler:innen	Evaluation der Implementation <i>Kontrollgruppe: Weiterbildung Rubrics</i>
Test Argumentieren		Test Argumentieren	

Die Gelegenheitsstichprobe umfasste 71 Lehrpersonen sowie deren 1261 Schülerinnen und Schüler (5. und 6. Jahrgangsstufe). Nach einer ersten Runde unseres Projekts im Jahr 2015 mit 45 Teilnehmenden haben wir die Studie 2019 mit weiteren 26 Lehrpersonen in identischer Form wiederholt, um Videomaterial zu erzeugen und die Stichprobe zu vergrößern. Rund die Hälfte der Klassen ($n = 664$ Schüler:innen) wurde der Interventionsgruppe zugewiesen. Bei der Bildung der zwei Gruppen wurde darauf geachtet, dass die Lehrpersonen bezüglich des Geschlechts, Alters, Jahrgangsklassen und Kantone gleichmässig verteilt waren. Alle Teilnehmenden haben sich freiwillig aufgrund einer Ausschreibung oder persönlicher Kontakte gemeldet.

3.2 Instrumente und Items

Die Schülerinnen und Schüler erhielten Fragebogen mit einer Auswahl an Konstrukten basierend auf dem theoretischen Modell von Panadero und Jonsson (2013, S. 139).

Die hier eingesetzte Skala „formatives Feedback der Lehrperson“ aus Sicht der Schüler:innen umfasste 5 Items ($t1/t2: \alpha = .75/.76$; Bsp. *Die Lehrperson bespricht mit mir, ob meine Lösungen zum Ziel führen.*). Die Items orientierten sich an den drei lernförderlichen von gesamthaft vier Feedbacktypen von Hattie und Timperley (2007): Feedback zur Aufgabe, zum Prozess und zur Selbstregulation. Die Items wiesen eine 6-stufige Likert-Skalierung auf („Wie oft geschieht...“ 6 = *immer*, 5 = *fast immer*, 4 = *oft*, 3 = *manchmal*, 2 = *selten* und 1 = *nie*).

Alle Items zur Messung des mathematischen Argumentierens wurden entweder aus anderen standardisierten Tests adaptiert oder für diese Studie entwickelt. Die Items orientierten sich an den nationalen Grundkompetenzen der Schweiz im Fach Mathematik (Schweizerische Konferenz der Erziehungsdirektoren (EDK), 2011) und bezogen sich auf drei der vier Inhaltsbereiche der Jahrgangsstufen 4–6: Zahlen und Variablen, funktionale Beziehungen sowie Grössen und Masse. Alle Testaufgaben wurden auf der Grundlage eines Korrekturmanuals auf vier Niveaustufen von geschulten Raterinnen eingeschätzt. Für jedes Bewertungsteam wurde eine zufriedenstellende Inter-Rater-Reliabilität erzielt: $Kappa > .70$. Insgesamt wurden 14 Items verwendet, die in zwei Mathematiktestheften zu je 10 Items verteilt wurden. Mittels IRT-Analysen wurden die endgültigen Kompetenzwerte aller Kinder auf der Grundlage von Bayes'schen Plausibilitätswerten (Von Davier et al., 2009) berechnet.

Als Kontextvariablen für die vorliegende Studie wurden „Geschlecht“, „Alter“, und „Familiensprache Deutsch/nicht Deutsch“ berücksichtigt. Auf der Klassenebene wurde die Gruppenvariable „mit/ohne Rubric“ als Dummy-Variable (1/0) miteinbezogen.

3.3 Praktische Umsetzung

Die Vermittlung des notwendigen Wissens zum mathematischen Argumentieren erfolgte in einem eintägigen Workshop. Die Interventionsgruppe wurde zusätzlich in den für das Projekt entwickelten Rubric eingeführt, während sich die Kontrollgruppe mit einem anderen mathematikdidaktischen Inhalt befasste. Beide Gruppen erhielten zudem einen Input zu formativem Feedback. In der neunwöchigen Implementationsphase arbeiteten die Interventions- und die Kontrollgruppe jeweils eine Lektion pro Woche an Aufgaben zum mathematischen Argumentieren. Dies entlang eines vorgegebenen Unterrichtsskripts, dessen Aufbau dem methodischen Vorgehen von Andrade, Du und Wang (2008, S. 6) entspricht. Das Skript beschrieb detailliert, welche Aufgaben in welcher Art und Weise zu bearbeiten waren. Dasjenige der Interventionsgruppe wies zusätzlich an, den Rubric einzusetzen, wiederum in Anlehnung an die Studie von Andrade et al. (2008, S. 6). Am Ende der Umsetzungsphase wurde die Weiterbildung mit den Lehrpersonen beider Gruppen evaluiert und die Teilnehmenden mit und ohne Einführung in die Rubrics auf einen vergleichbaren Wissensstand gebracht. In der Woche vor und nach der Implementationsphase erfolgte jeweils die Datenerhebung.

3.4 Analyse

Wir modellierten die Beziehung zwischen formativem Feedback und mathematischem Argumentieren mit Hilfe eines Pfadanalysemodells mit latenten Variablen. Dazu wurden Strukturgleichungsmodelle (SEM) mit Mplus 8 und einer „full information maximum likelihood“ Schätzung (FIML) gerechnet, also mit unvollständigen Datensätzen. Die Evaluation der Modellanpassung erfolgte mit komparativen und absoluten

Fit-Indices. Die Werte des komparativen Fit-Index mit $CFI > .95$ zeigen eine sehr gute Passung (Hu & Bentler, 1999). Dasselbe gilt für diejenigen des globalen Fit-Index Root Mean Square Error of Approximation (RMSEA) $< .05$. Bei den Analysen mit Daten der Schülerinnen und Schüler wurde die Clusterstruktur (Designeffekt, Adjustierung der Standardfehler) in Klassen mittels Mehrebenenanalyse berücksichtigt.

4 Ergebnisse

Bevor die Ergebnisse der Pfadanalyse vorgestellt werden, folgt ein Überblick über die Korrelationen, Mittelwerte (M) und Standardabweichungen (SD) der in der Studie verwendeten Variablen (Tabelle 2). Die Kompetenzwerte für mathematisches Argumentieren wurde beim ersten Erhebungszeitpunkt (T1) auf $M = 0$ festgelegt (Varianz = 1) und die plausiblen Werte für die Parameter der Personen wurden anhand einer Standardnormalverteilung transformiert. Entsprechend betrug der Mittelwert bei T1 = 0, bzw. $-.02$ durch Rundungsfehler, und die Standardabweichung (SD) betrug $.81$ zu beiden Messzeitpunkten. Beim zweiten Erhebungszeitpunkt (T2) war der Mittelwert des mathematischen Argumentierens höher: $M = .60$. Im Durchschnitt berichteten die Schülerinnen und Schüler über eine eher häufigere Wahrnehmung des formativen Feedbacks der Lehrkräfte ($M = 3.85$ bei T1). Die Wahrnehmung der Häufigkeit nahm aber im Laufe der Implementationsphase leicht ab ($M = 3.63$ bei T2).

Die Korrelationen auf der Ebene der Schülerinnen und Schüler zeigten nur stabile, positive Zusammenhänge über die Zeit und keine Zusammenhänge zwischen den Testleistungen und der Wahrnehmung der Feedbackhäufigkeit. Auf Klassenebene ergab sich eine stärkere, jedoch nicht signifikante Korrelation zwischen den Testwerten und der Wahrnehmung der Feedbackhäufigkeit bei T2 ($r = .14$).

Tabelle 2: Deskriptive Angaben und Korrelationen des von den Schüler:innen wahrgenommenen Lehrpersonenfeedbacks und der Kompetenzwerte im mathematischen Argumentieren

	<i>M</i>	<i>SD</i>	<i>ICC</i>	1. Argumentieren T1	2. Formatives Feedback T1	3. Argumentieren T2	4. Formatives Feedback T2
1. Argumentieren T1	-.02	.81	.16	-	-.02	.87**	-.03
2. Formatives Feedback T1	3.85	.77	.13	-.01	-	.11	.68*
3. Argumentieren T2	.60	.81	.15	.91**	-.01	-	.14
4. Formatives Feedback T2	3.63	.88	.18	.01	.46**	-.01	-

Anmerkungen: $N = 1261$; ** $p < .01$, * $p < .05$. *ICC* = Intraklassenkorrelation, Argumentieren: Standardisierte Kompetenzwerte. Feedback: 1-6 Likert-Skala; Pole: nie/immer. Die Korrelationen unterhalb der Diagonalen sind die Individualwerte und diejenigen oberhalb der Diagonalen sind die Klassenwerte.

Im Folgenden wurde in Bezug auf die erste Fragestellung geklärt, ob die von Seiten der Lernenden eingeschätzte formative Feedbackhäufigkeit der Lehrperson mit der Kompetenz im mathematischen Argumentieren zusammenhängt. Ergänzend wurde in Bezug auf die zweite Fragestellung mitberücksichtigt, ob sich die Interventionsgruppe mit Rubric von der Kontrollgruppe ohne Rubric bezüglich der Wahrnehmung des Feedbacks unterscheidet. Zum gerechneten Mehrebenen-Strukturgleichungsmodell (Abb. 2) wurden die drei bereits erwähnten Kontextvariablen (Alter, Geschlecht und Familiensprache Deutsch/nicht Deutsch) explorativ geprüft, aber keine Kontextvariable half zusätzlich bedeutsam die Varianz aufzuklären. Deshalb wurden sie nicht ins Modell aufgenommen. Indirekte Pfade wurden geprüft, aber es zeigten sich keine Effekte. Das finale Modell hat folgende Fit-Werte: $\chi^2(16) = 2178.05$, $p = .00$, CFI = 1.00, TLI = 1.00, RMSEA = .00, SRMRw = .00, SRMRb = .08.

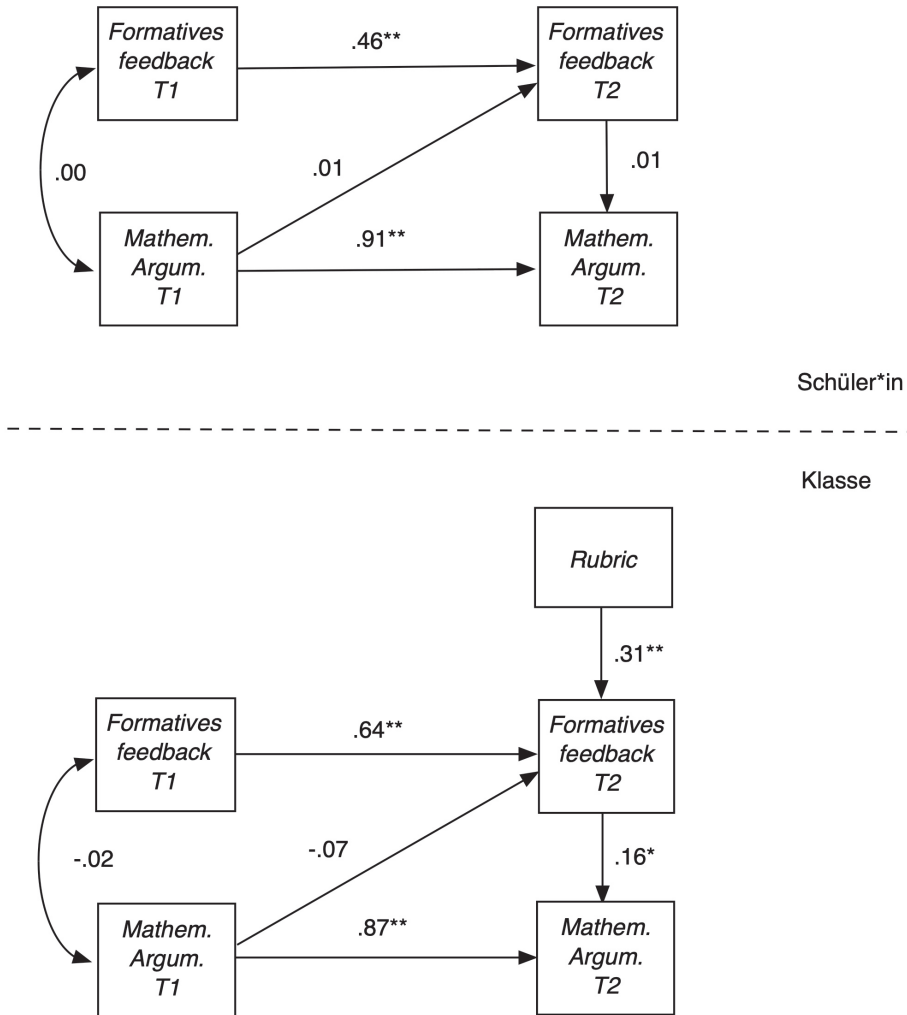


Abbildung 2: Strukturgleichungsmodell für den Zusammenhang zwischen formativem Feedback und Mathematischem Argumentieren auf Individual- und Klassenebene. Auf der Klassenebene wurde die Interventionsvariable „Gruppe mit Rubric“ mitberücksichtigt. $N_{between} = 71, N_{within} = 1261$; $**p < .01, *p < .05$.

In Bezug auf die erste Fragestellung, ob das von den Lernenden wahrgenommene formative Feedback einen Einfluss auf die Kompetenzentwicklung im mathematischen Argumentieren hat, wies die Häufigkeit von formativem Feedback T2 auf der Individualebene keine Wirkung auf das mathematische Argumentieren T2 auf, auf der Klassenebene schon. Es zeigten sich im Modell auf der Ebene der Schülerinnen und Schüler stabile Zusammenhänge über die Zeit (T1-T2): Häufigkeit von formativem

Feedback: $\beta = .46$ und mathematisches Argumentieren: $\beta = .91$. Während sich auf der Klassenebene ebenfalls starke zeitliche Zusammenhänge zeigten, ergaben sich hier auch zwei signifikante Pfade zwischen verschiedenen Konstrukten. Erstens zeigte sich ein signifikanter Effekt von der Häufigkeit von formativem Feedback T2 auf das mathematische Argumentieren T2 ($\beta = .16$) und zweitens fand sich im Modell in Bezug auf die zweite Fragestellung, ob ein kriteriengeleiteter Rubric die Kompetenzentwicklung unterstützt, ein signifikanter Pfad von der Interventionsgruppe (mit Rubric) auf die Häufigkeit von formativem Feedback T2 ($\beta = .31$). Das Modell klärt 83 % der Varianz der Leistungen beim mathematischen Argumentieren auf der Individual- und 76 % auf der Klassenebene. Dazu trägt in erster Linie die Variable mathematisches Argumentieren zu t1 bei. Der Beitrag des formativen Feedbacks ist eher gering, aber als eine Handlungsmöglichkeit für die Lehrperson relevant.

5 Zusammenfassung/Ausblick/Forschungsperspektiven

Im Schnitt haben sich die Schülerinnen und Schüler im mathematischen Argumentieren verbessert. Dabei verliefen diese Verläufe individuell und für die Klasse primär in Abhängigkeit von der Eingangsleistungen (T1). Beide Forschungsfragen konnten auf der Klassenebene positiv beantwortet werden. Erstens liess sich bezüglich des Einflusses des von den Schülerinnen und Schüler wahrgenommene formativen Feedbacks auf das mathematische Argumentieren mittels Mehrebenenanalyse zeigen, dass sich aus Sicht der Schülerinnen und Schüler die klassenbezogenen Argumentationskompetenzen direkt durch die berichtete Häufigkeit des formativen Feedbacks miterklären lassen. Zweitens zeigte sich bezüglich der Unterstützung der Kompetenzentwicklung durch den Rubric, dass Schülerinnen und Schüler aus Klassen, die einen Rubric nutzten, häufiger formative Feedbacks von Lehrperson erhielten als Schülerinnen und Schüler in Klassen, die keinen Rubric einsetzten. Auf der Individualseite konnte die Wirkung des formativen Feedbacks jedoch nicht nachgewiesen werden. Es lässt allenfalls auch darauf schliessen, dass es das richtige Feedback für alle in Bezug auf die Häufigkeit nicht gibt, sondern dass es individueller Anpassungen an den Lernstand bedarf (Fyfe & Brown, 2018). Kognitiv starke Schülerinnen und Schüler bedürfen etwa weniger Feedback, schwächere mehr. Die Wahrnehmung des Feedbacks kann zudem von affektiv-motivationalen Aspekten der Lernenden beeinflusst sein. Unmotivierte Lernende scheinen das Feedback der Lehrperson weniger aufnehmen und nutzen zu wollen (Ruelmann et al., 2021).

Für die Förderung der Kompetenz des mathematischen Argumentierens müssen Lehrpersonen wie in unserem Projekt reichhaltige Lernsituationen für einen Argumentationsdiskurs gestalten und Aufgaben mit offener Fragestellung für Schülerinnen und Schüler mit unterschiedlichem Wissensstand und unterschiedlichen Problemlösefähigkeiten stellen (Mercer, 2009). Ein entsprechender Unterricht sollte beinhalten, dass die Lehrperson die Schülerinnen und Schüler motiviert und anleitet, indem sie

gehaltvolles Feedback gibt, das auf die Zone der proximalen Entwicklung jedes einzelnen Kindes abgestimmt ist (Vygotsky, 1974). In solchen Lernanlässen kann die durch Rubrics ermöglichte Zieltransparenz wesentlich zur Sicherheit bezüglich Selbststeuerung und Umsetzen des Feedbacks beitragen, wie dies etwa von Wyatt-Smith und Adie (2019) als Teil einer professionellen Unterrichtspraxis gefordert wird. Wir stellten bei einer noch laufenden Videoanalyse grössere Unterschiede beim Einsatz des Rubrics fest. Wie wirksam Lehrpersonen schlussendlich Rubrics einsetzen, hängt vermutlich davon ab, inwiefern sie neben entsprechenden Weiterbildungen bereits positive Erfahrungen mit solchen Rastern gesammelt haben und Vorteile für ihren Unterricht erkennen können (Panadero et al., 2016). Für die Ausbildung von Lehrpersonen wäre es insofern sinnvoll, bereits mit Studierenden Rubrics zu entwickeln und diese in Praktika erproben zu lassen.

Die Wahrnehmung der Schülerinnen und Schüler kann Verzerrungen unterliegen. Es stellt sich somit die Frage, ob Schülerinnen und Schüler die Qualität des Feedbacks der Lehrkraft zuverlässig und valide bewerten können. Diesbezüglich berichteten van der Scheer et al. (2019) über grundsätzlich zuverlässige Werte für die Wahrnehmung von Lernenden auf der Primarstufe in Bezug auf unterrichtsbezogene Handlungen von Lehrpersonen. In einer früheren Studie haben wir zudem die Übereinstimmung zwischen den Wahrnehmungen von Lehrkräften und Schülerinnen und Schülern in Bezug auf formatives Feedback berechnet und einen zufriedenstellenden Zusammenhang ermittelt ($r = .40$) (Smit et al., 2017).

Unser Instrument zur Messung des formativen Feedbacks war ausreichend sensitiv, um einen Teil der Klassenvarianz vorherzusagen. Bezüglich der fehlenden Zusammenhänge auf der Schülerebene könnte eine weitere methodische Einschränkung darin bestehen, dass die Häufigkeit und Gestaltung des Feedbacks stärker auf die individuellen Voraussetzungen und den Fortschritt der Schülerinnen und Schüler beim Lernen ausgerichtet sein sollte. Schülerinnen und Schüler brauchen Feedback zu verschiedenen Aspekten, aber nicht gleichzeitig, sondern angepasst an den aktuellen Schritt im Problemlösungsprozess (Hattie & Gan, 2011). Es ist unter Umständen sinnvoll, dass mit fortschreitender Kompetenzentwicklung im mathematischen Argumentieren ein Wechsel von aufgabenbezogenem Feedback zu selbstregulationsbezogenem Feedback effektiver für das Lernen sein könnte. Es wäre – als Forschungsausblick – bedeutsam, solche zeitlichen Effekte des von Schülerinnen und Schülern wahrgenommenen optimalen formativen Feedbacks in einer zukünftigen Studie mittels häufigerer Messung während der Umsetzungsphase spezifischer zu erfassen.

Anmerkungen:

Das Projekt wurde vom Schweizerischen Nationalfonds (Nr. 149386) unterstützt.

Literatur

- Allal, L. (1991). *Vers une pratique de l'évaluation formative: matériel de formation continue des enseignants*. De Boeck.
- Andrade, H., Du, Y. & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 3–13. <https://doi.org/10.1111/j.1745-3992.2008.00118.x>
- Ball, D.L. & Bass, H. (2000). Making believe: The collective construction of public mathematical knowledge in the elementary classroom. In D. C. Philipps (Hrsg.), *Constructivism in education: Opinions and second opinions on controversial issues. Yearbook of the National Society for the Study of Education* (S. 193–224). University of Chicago Press. <https://doi.org/10.1177/016146810010200707>
- Bezold, A. (2009). *Förderung von Argumentationskompetenzen durch selbstdifferenzierende Lernangebote. Eine Studie im Mathematikunterricht der Grundschule*. Kovac. <https://doi.org/10.1007/BF03339083>
- Bezold, A. (2010). *Mathematisches Argumentieren in der Grundschule fördern. Was Lehrkräfte dazu beitragen können SINUS-Transfer Grundschule*. http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/Handreichung_Mathe_Bezold.pdf
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Blanton, M.L. & Kaput, J.J. (2005). Characterizing a classroom practice that promotes algebraic reasoning. *Journal for Research in Mathematics Education*, 36(5), 412–446.
- Brookhart, S.M. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: Ascd. <https://doi.org/10.4135/9781452218649.n15>
- Brookhart, S.M. & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343–368. <https://doi.org/10.1080/00131911.2014.929565>
- Brooks, C., Burton, R., van der Kleij, F., Ablaza, C., Carroll, A., Hattie, J. & Neill, S. (2021). Teachers activating learners: The effects of a student-centred feedback approach on writing achievement. *Teaching and Teacher Education*, 105, 103387. <https://doi.org/10.1016/j.tate.2021.103387>
- Buholzer, A., Baer, M., Zulliger, S., Torchetti, L., Ruelmann, M., Häfliger, A. & Lötscher, H. (2020). Formatives Assessment im alltäglichen Mathematikunterricht von Primarlehrpersonen: Häufigkeit, Dauer und Qualität. *Unterrichtswissenschaft*. <https://doi.org/10.1007/s42010-020-00083-7>
- Cho, K., Schunn, C.D. & Wilson, R.W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891. <https://doi.org/10.1037/0022-0663.98.4.891>
- Collins, A., Brown, J.S. & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Hrsg.), *Knowing, learning, and instruction. Essays in honor of Robert Glaser* (S. 453–494). Erlbaum. <https://doi.org/10.4324/9781315044408-14>
- Cramer, J.C. & Knipping, C. (2018). Participation in Argumentation. In U. Gellert, C. Knipping & H. Straehler-Pohl (Hrsg.), *Inside the Mathematics Class: Sociological Perspectives on Participation, Inclusion, and Enhancement* (S. 229–244). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-79045-9_11
- Danielson, C. (1997). *A Collection of Performance Tasks and Rubrics. Middle School Mathematics*. Eye on Education.

- de Leeuw, J. (2016). Rubrics and Exemplars in Writing Assessment. In S. Scott, E. D. Scott & F. C. Webber (Hrsg.), *Leadership of Assessment, Inclusion, and Learning* (S. 89–110). Springer International Publishing. https://doi.org/10.1007/978-3-319-23347-5_4
- Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik*, 48(2), 262–286.
- EDK (Schweizerische Konferenz der Erziehungsdirektoren). (2011). *Grundkompetenzen für die Mathematik. Nationale Bildungsstandards*. http://edudoc.ch/record/96784/files/grundkomp_math_d.pdf
- English, N., Robertson, P., Gillis, S. & Graham, L. (2022). Rubrics and formative assessment in K-12 education: A scoping review of literature. *International Journal of Educational Research*, 113, 101964. <https://doi.org/10.1016/j.ijer.2022.101964>
- Fetzer, M. (2011). Wie argumentieren Grundschul Kinder im Mathematikunterricht? Eine argumentationstheoretische Perspektive. *Journal für Mathematik-Didaktik*, 32(1), 27–51. <https://doi.org/10.1007/s13138-010-0021-z>
- Fyfe, E.R. & Brown, S.A. (2018). Feedback influences children's reasoning about math equivalence: A meta-analytic review. *Thinking & Reasoning*, 24(2), 157–178. <https://doi.org/10.1080/13546783.2017.1359208>
- Ghaffar, M.A., Khairallah, M. & Salloum, S. (2020). Co-constructed rubrics and assessment for learning: The impact on middle school students' attitudes and writing skills. *Assessing Writing*, 45, 100468. <https://doi.org/10.1016/j.asw.2020.100468>
- Ginsburg, H.P. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development*, 52, 109–128. <https://doi.org/10.1159/000202729>
- Gregori-Giralt, E. & Menéndez-Varela, J.-L. (2018). The reliability and sources of error of using rubrics-based assessment for student projects. *Assessment & Evaluation in Higher Education*, 43(3), 488–499. <https://doi.org/10.1080/02602938.2017.1360838>
- Harlen, W. (2006). The role of assessment in developing motivation for learning. In J. Gardner (Hrsg.), *Assessment and learning* (S. 61–80). Sage.
- Hattie, J. (2008). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9780203887332>
- Hattie, J. & Gan, M. (2011). Instruction based on feedback. In *Handbook of research on learning and instruction* (S. 263–285). Routledge. <https://doi.org/10.4324/9780203839089-22>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Herbert, S., Vale, C., White, P. & Bragg, L.A. (2022). Engagement with a formative assessment rubric: A case of mathematical reasoning. *International Journal of Educational Research*, 111, 101899. <https://doi.org/10.1016/j.ijer.2021.101899>
- Hu, L.-T. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Keiser, J.M. & Lambdin, D.V. (1996). The clock is ticking: Time constraint issues in mathematics teaching reform. *The journal of educational research*, 90(1), 23–31. <https://doi.org/10.1080/00220671.1996.9944440>
- Kempert, S., Saalbach, H. & Hardy, I. (2011). Cognitive benefits and costs of bilingualism in elementary school students: The case of mathematical word problems. *Journal of educational psychology*, 103(3), 547. <https://doi.org/10.1037/a0023619>

- Kluger, A.N. & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, 119(2), 254. <https://doi.org/10.1037/0033-2909.119.2.254>
- Kroesbergen, E.H. & Van Luit, J.E. (2003). Mathematics interventions for children with special educational neHrsg a meta-analysis. *Remedial and special education*, 24(2), 97–114. <https://doi.org/10.1177/07419325030240020501>
- Linneweber-Lammerskitten, H., Wälti, B., Smit, R., Marc, V., Pochon, L.-O., Frapolli, A., Cadorin, L., Moser Opitz, E., Hirt, U. & Keller, R. (2010). *Basisstandards für die Mathematik. Unterlagen für den Anhörungsprozess. Schweizerische Konferenz der Erziehungsdirektoren (EDK)*.
- Lithner, J. (2000). Mathematical Reasoning in Task Solving. *Educational studies in mathematics*, 41(2), 165–190. <https://doi.org/10.1023/A:1003956417456>
- Meier, S.L., Rich, B.S. & Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: Factors related to discrepancies in scoring. *Assessment in Education*, 13(01), 69–95. <https://doi.org/10.1080/09695940600563512>
- Mercer, N. (2009). Developing argumentation: Lessons learned in the primary school. In N. Muller Mirza & A.-N. Perret-Clermont (Hrsg.), *Argumentation and education* (S. 177–194). Springer. https://doi.org/10.1007/978-0-387-98125-3_7
- Moss, C.M. & Brookhart, S.M. (2010). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Alexandria, VA: ASCD.
- Panadero, E. & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129–144. <https://dx.doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., Jonsson, A. & Strijbos, J.-W. (2016). Scaffolding Self-Regulated Learning Through Self-Assessment and Peer Assessment: Guidelines for Classroom Implementation. In D. Laveault & L. Allal (Hrsg.), *Assessment for Learning: Meeting the Challenge of Implementation* (S. 311–326). Springer International Publishing. https://doi.org/10.1007/978-3-319-39211-0_18
- Perrenoud, P. (1991). Formative Schülerbeurteilung: Welcher Platz in der Didaktik? *Beiträge zur Lehrerbildung*, 9(3), 309–329.
- Reusser, K. (2009). Empirisch fundierte Didaktik—didaktisch fundierte Unterrichtsfor-schung. *Zeitschrift für Erziehungswissenschaft*, 10(Sonderheft 9), 219–237. https://doi.org/10.1007/978-3-531-91775-7_15
- Ruelmann, M., Torchetti, L., Zulliger, S., Buholzer, A. & Praetorius, A.-K. (2021). Kognitiv-motivationale Schüler*innenprofile und ihre Bedeutung für die Schüler*innenwahrnehmung der Lernunterstützung durch die Lehrperson. *Unterrichtswissenschaft*, 49(3), 395–422. <https://doi.org/10.1007/s42010-021-00100-3>
- Ruf, U. & Gallin, P. (1998). *Dialogisches Lernen in Sprache und Mathematik. Band 2: Spuren legen – Spuren lesen*. Kallmeyer.
- Ruiz-Primo, M.A. & Brookhart, S.M. (2018). *Using feedback to improve learning*. Routledge. <https://doi.org/10.4324/9781315627502>
- Schmidt, C. (2020). *Empirische Befunde zum formativen Assessment Formatives Assessment in der Grundschule: Konzept, Einschätzungen der Lehrkräfte und Zusammenhänge* (S. 63–81). Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-26921-0>
- Sfard, A. (2001). There is more to discourse than meets the ears: Looking at thinking as communicating to learn more about mathematical learning. *Educational studies in mathematics*, 46(1/3), 13–57. <https://doi.org/10.1023/A:1014097416157>
- Shute, V.J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>

- Smit, R., Bachmann, P., Blum, V., Birri, T. & Hess, K. (2017). Effects of a rubric for mathematical reasoning on teaching and learning in primary school. *Instructional science*, 45(5), 603–622. <https://doi.org/10.1007/s11251-017-9416-2>
- Stanat, P. & Christensen, G. (2006). *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003*. OECD.
- Stein, M.K., Grover, B.W. & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, 33(2), 455–488. <https://doi.org/10.3102/00028312033002455>
- Tang, X., Coffey, J. & Levin, D.M. (2015). Reconsidering the use of scoring rubrics in biology instruction. *The American Biology Teacher*, 77(9), 669–675. <https://doi.org/10.1525/abt.2015.77.9.4>
- Thompson, D.R. & Schultz-Ferrel, K. (2008). *Introduction to reasoning and proof. Grades 6–8*. Heinemann.
- Turner, S.L. (2014). Creating an assessment-centered classroom: Five essential assessment strategies to support middle grades student learning and achievement. *Middle School Journal*, 45(5), 3–16. <https://doi.org/10.1080/00940771.2014.11461895>
- van der Scheer, E.A., Bijlsma, H.J.E. & Glas, C.A.W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement*, 30(1), 30–50. <https://doi.org/10.1080/09243453.2018.1539015>
- Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI monograph series*, 2, 9–36.
- Vygotsky, L.S. (1974). *Denken und Sprechen*. Fischer.
- William, D. & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Hrsg.), *The future of assessment: Shaping teaching and learning* (S. 53–82). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781315086545-3>
- Wyatt-Smith, C. & Adie, L. (2019). The development of students' evaluative expertise: enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*, 1–21. <https://doi.org/10.1080/00220272.2019.1624831>

Unterrichtsgestaltung und Qualität von Peer-Assessment im alltäglichen Unterricht

Sandra Zulliger | Merle Ruelmann | Hanni Lötscher
Pädagogische Hochschule Luzern

Abstract: Es besteht ein breiter Konsens, dass Peer-Assessments am lernförderlichsten wirken, wenn sie formativ eingesetzt werden. Hinsichtlich der Qualität und Gestaltung von Peer-Assessment im alltäglichen Unterricht liegen jedoch kaum Forschungserkenntnisse vor.

Ziel der vorliegenden Studie ist es, Zusammenhänge zwischen der methodischen Umsetzung (Oberflächenstruktur) und der Qualität (Tiefenstruktur) von Peer-Assessment und seinem Unterrichtskontext zu untersuchen. Dazu wurden von 52 vierten Primarschulklassen der Deutschschweiz Einführungslektionen in die halbschriftliche Division videografiert und die Oberflächen- und Tiefenstrukturen mittels Codings und Ratings ausgewertet.

Die Ergebnisse zeigen, dass qualitative Peer-Assessments methodisch mit einem eigenständigen Peer-Assessment der Lernenden beginnen und mit einer lehrpersonenmoderierten Zusammenführung enden. Unterricht mit qualitativem Peer-Assessment zeichnet sich tiefenstrukturell durch Merkmale eines ko-konstruktiven Unterrichts und oberflächenstrukturell durch einen geringeren Anteil der Erarbeitung von Problemstellungen im Klassenunterricht aus und nutzt somit formative Assessment-Strategien als eine Lehr- und Lernstrategie.

1 Einleitung

Bildungspolitische Forderungen wie die Einführung des kompetenzorientierten Lehrplans 21 in der Schweiz (D-EDK, 2016) sowie Erkenntnisse aus der Lehr- und Lernforschung verlangen einen Unterricht, der die Schülerinnen und Schüler anleitet, selbstgesteuert und kooperativ zu lernen (Butler & Lee, 2010; Heritage, 2022). Dem liegt ein (sozial-)konstruktivistisches Lernverständnis zugrunde, das Lernen als aktiv-konstruierenden, auf Vorwissen aufbauenden, selbstregulierten und sozialen Prozess begreift (Reusser, 2016).

Peer-Assessment ist eine Unterrichtsmethode, die diese Forderungen einlösen kann und einen aktiven Einbezug der Schülerinnen und Schüler in den Lernprozess realisiert. Gemäss Topping (2009) kann Peer-Assessment folgendermassen definiert werden: Peer-Assessment „is an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners“ (S. 20). Peer-Assessment wird insofern auch als Teil formativen Assessments angesehen (Panadero & Brown, 2017), als die gemeinsame Reflexion mit Peers über Lernprozesse und -ergebnisse während des Unterrichts zu einer formativen Überprüfung und Optimierung des

Lernens beitragen kann (Heritage, 2022). Peer-Assessments sind zu fast jedem Zeitpunkt des Unterrichts einsetzbar und können von einer einfachen Anweisung, sich kurz mit dem Sitznachbarn oder der Sitznachbarin über eine Idee auszutauschen, bis hin zur Beurteilung eines Projekts einer anderen Gruppe mittels eines Beurteilungsrasters reichen (Wylie & Lyon, 2013). Trotz der vielen möglichen Varianten von Peer-Assessment besteht Konsens darüber, dass es sein höchstes Lernpotenzial entfaltet, wenn es im Sinne eines formativen Assessments umgesetzt wird (Yan et al., 2022) und die Informationen aus den generierten Peer-Feedbacks für die Optimierung der Lernprozesse genutzt werden (Heritage, 2022).

Die im Unterricht sichtbare methodische Umsetzung von Peer-Assessment, wie der vormals erwähnte Ideenaustausch unter Sitznachbarn, kann als Oberflächenstruktur des Unterrichts verstanden werden. Die Lernprozesse und das dahinterstehende pädagogische Handeln, wie z. B. die kognitive Aktivierung, sind auf einer darunterliegenden tiefenstrukturellen Ebene angesiedelt. Die Tiefenstruktur wird in der pädagogisch-psychologischen Forschung gleichgesetzt mit der Qualität des Unterrichts, welche die Lerneffekte voraussagt (Decristan et al., 2020).

Studien zu anderen Unterrichtsmerkmalen wie Lehrpersonenfragen (Hess & Lipowsky, 2020) und Inszenierungsmustern (Hugener, 2008) zeigen, dass Erkenntnisse zu den Zusammenhängen zwischen Tiefen- und Oberflächenstruktur des Unterrichts wertvolle Hinweise für die Umsetzung hochwertigen Unterrichts liefern können. Sie zeigen auf wie tiefenstrukturelle Qualität eines Merkmals auf der Unterrichtsoberfläche methodisch gestaltet werden kann (Pauli & Schmid, 2019).

Bislang sind Zusammenhänge zwischen allgemeinen Oberflächen- und Tiefenstrukturen des Unterrichts sowie Oberflächen- und Tiefenstrukturen spezifisch für das Peer-Assessments noch kaum erforscht. Mit der Unterscheidung zwischen Oberflächen- und Tiefenstrukturen wird im vorliegenden Beitrag anhand einer videobasierten Analyse die Umsetzung von Peer-Assessment im alltäglichen Unterricht untersucht. Ausgehend von der Qualität von Peer-Assessment werden die Zusammenhänge zur Oberfläche, d.h. der Gestaltung von Peer-Assessment veranschaulicht, um herauszufinden, wie potenziell lernförderliche Peer-Assessments im Unterricht gestaltet werden können. Des Weiteren wird der Zusammenhang zwischen der Qualität von Peer-Assessment und dem Unterricht, in den das Peer-Assessment eingebettet ist (gesamte Oberflächenstruktur), analysiert. Untersucht wurde alltäglicher Unterricht, weil einerseits Untersuchungen zum tatsächlichen Vorkommen formativen Assessments, hier Peer-Assessment, fehlen (Buholzer et al., 2020) und andererseits ökologisch valide Aussagen erarbeitet werden sollten (Schnell et al., 2013).

Aus den Ergebnissen können Schlussfolgerungen darüber gezogen werden, wie qualitätsvolles formatives Peer-Assessment im Unterricht gestaltet werden kann und ob sich gewisse Gelingensbedingungen für die Umsetzung von hochwertigen Peer-Assessments in den Oberflächen- oder Tiefenstrukturen von Unterricht zeigen.

Im folgenden Kapitel werden Begrifflichkeit, Nutzen, Einsatz und Qualitätsmerkmale von Peer-Assessment aufgezeigt. Das zweite Theoriekapitel verdeutlicht anhand des

Konzepts der Oberflächen- und Tiefenstrukturen, wie Unterrichtsqualität, die Unterrichtsgestaltung und Peer-Assessment zusammenhängen. Im anschließenden Kapitel wird die Fragestellung veranschaulicht. Es folgen Kapitel zur Methode sowie zu den Ergebnissen und die Diskussion.

2 Peer-Assessment im Unterricht

Die Begriffe ‚Oberflächenstruktur‘ und ‚Tiefenstruktur‘ des Unterrichts bilden ein hilfreiches Konzept, um die Verknüpfung der methodischen Gestaltung des Unterrichts und wirksamer Lehr- und Lernprozesse aufzuzeigen und zu analysieren.

Die Oberflächenstrukturen, teilweise auch Oberflächenmerkmale oder Sichtstrukturen genannt, umfassen die leicht erkennbaren organisatorischen und methodischen Gestaltungsmerkmale des Unterrichts (z. B. eingesetzte Sozialformen und inhaltliche Aktivitäten) und unmittelbar beobachtbares Verhalten der Lehrpersonen und der Schülerinnen und Schüler. Als Tiefenstrukturen des Unterrichts werden in der pädagogisch-psychologischen Forschung die Interaktionen zwischen Lehrperson und Schülerinnen und Schülern sowie die Auseinandersetzung mit den Lerninhalten bezeichnet, die zusammengenommen die Qualität der Unterrichtsprozesse abbilden. Aus der Verknüpfung von Unterrichtsmerkmalen der Tiefenstruktur und der Oberflächenstruktur können Aussagen zur lernwirksamen Ausgestaltung und Umsetzung der Inhalte im Unterricht generiert werden (Decristan et al., 2020).

2.1 Oberflächen- und Tiefenstrukturen von Peer-Assessment

Die eingangs dargelegte Definition von Topping (2009) umreißt Peer-Assessment als eine organisierte Aktivität, bei der gleichgestellte Lernende das Niveau, den Wert oder die Qualität eines Produkts oder einer Leistung anderer gleichgestellter Lernender berücksichtigen oder festlegen. Peer-Assessment kann zu zweit oder in Gruppen durchgeführt werden. Auch kann es in einer wechselseitigen verbalen Interaktion während des Unterrichts oder aus der Ferne schriftlich, ggf. auch anonymisiert, realisiert werden. Zudem können verschiedene Produkte von Schülerinnen und Schülern Gegenstand von Peer-Assessment sein, wie u. a. Leistungstests, Portfolios oder Unterrichtsergebnisse von Lernenden. Peer-Assessments können summativ, am Schluss des Lernprozesses, oder formativ, während des Lernprozesses, erfolgen (Topping, 2010). Zu den einzelnen Faktoren, welche die Praxis und die Implementation von Peer-Assessment bzw. formativem Assessment beeinflussen, liegen zahlreiche Studienergebnisse vor (Double et al., 2020). Die Organisation und die Gestaltung von Peer-Assessment durch die Lehrperson werden als Erfolgsfaktor für dessen Umsetzung hervorgehoben (Ploegh et al., 2009; Topping, 2010).

Effektives Feedback aus Peer-Assessment gibt den Schülerinnen und Schülern sowie den Lehrpersonen Hinweise zum Kompetenzstand im Vergleich zu den Lernzielen. Die

Informationen aus dem Peer-Assessment helfen, die Lücke zwischen dem derzeitigen Lernstand und den Lernzielen zu schliessen (Heritage, 2022). Die positive Wirkung von Peer-Assessments auf die Lernleistung der Schülerinnen und Schüler wurde bereits in mehreren Metastudien nachgewiesen (Double et al., 2020; Graham et al., 2015; Li et al., 2016; Sanchez et al., 2017, Yan et al., 2022). Schülerinnen und Schüler können Kritik und Hinweise voneinander eher akzeptieren und ernst nehmen als von der Lehrperson (Andrade, 2010). Die Schülerinnen und Schüler nehmen beim Peer-Assessment auch die Rolle des Lehrenden ein und werden dabei als instruktionale Ressourcen füreinander aktiviert (Topping, 2010).

Die Qualität der Umsetzung von Peer-Assessment lässt sich anhand der folgenden drei Dimensionen beschreiben.

Anforderungsniveau des Peer-Assessments: Hochwertiges Peer-Assessment verlangt anspruchsvolle Aufgabenstellungen (Black et al., 2004; Wylie & Lyon, 2013), die die Lernenden zu tiefgreifenden und metakognitiven Denkprozessen herausfordern (Topping, 2009). Die Anleitungen und Aufgabenstellungen regen die Lernenden an, ihre eigenen oder die Lernprozesse und -ergebnisse anderer differenziert zu beschreiben sowie zu diskutieren und sie auf der Grundlage von Lernzielen und Bewertungskriterien explizit zu überprüfen und zu bewerten. Peer-Assessment sollte ausführlich und spezifisch sein und zu einem Dialog zwischen den Teilnehmenden führen (Kollar & Fischer, 2010).

Qualität der Anleitung, Durchführung und Unterstützung: Lehrerinnen und Lehrer, vor allem in der Grundschule, haben die wichtige Aufgabe, die Umsetzung von Peer-Assessment sorgfältig anzuleiten und zu überwachen und die Lernenden bei notwendigen Reflexions- und Bewertungsprozessen zu unterstützen (Harris & Brown, 2013; Ploegh et al., 2009). Spezifische Anweisungen und ein sorgfältiges Management der zwischenmenschlichen Beziehungen sind wichtig für die erfolgreiche Umsetzung von Peer-Assessment (Harris & Brown, 2013), denn so können Lernende entsprechende Fähigkeiten entwickeln, um Peer-Assessment anzuwenden (Black et al., 2004). Reusser (2016) betont die Bedeutung der Ko-Konstruktion im Unterricht, d.h., Lernprozesse sollen interaktiv gestaltet und in soziale Kontexte eingebettet sein.

Grad der Nutzung der gewonnenen Informationen: Peer-Assessment wird dann lernwirksam, wenn die Lernenden ermutigt werden, die Erkenntnisse aus dem Peer-Assessment (metakognitiv) zu nutzen, um ihre Lernprozesse zielbezogen zu steuern. Eine effektive Nutzung der gewonnenen Informationen bedeutet, dass die Erkenntnisse aus dem Peer-Assessment im Unterricht ausgetauscht, diskutiert und systematisiert werden. Oder dass die Lernenden angeleitet werden, die aus ihrem eigenen Feedback und dem Feedback anderer gewonnenen Informationen für das weitere Lernen zu berücksichtigen (Andrade & Valtcheva, 2009). Für die Lehrkräfte besteht die Herausforderung darin, Peer-Assessment in die Gestaltung der Lehr- und Lernprozesse einzubinden, sodass die nachfolgenden Unterrichtsphasen darauf Bezug nehmen.

Zur Nutzung und der Qualität von Peer-Assessment im alltäglichen Unterricht der Primar- und Sekundarstufe sind für den deutschsprachigen Raum wenige Studienergebnisse vorhanden, die ausschliesslich auf Lehrpersonenbefragungen beruhen. Der

Einsatz von Peer-Assessment im Unterricht ist dabei eher gering (Bürgermeister, 2014; Schmidt, 2020). In einer Befragung aus Spanien geben die Lehrpersonen ebenfalls eine nur gelegentliche Verwendung von Peer-Assessment in ihrem Unterricht an (Panadero & Brown, 2017). Zur Qualität von Peer-Assessment liegen einzelne Video- und Beobachtungsstudien aus dem amerikanischen Raum vor, die auf eher geringe qualitative Umsetzungen hindeuten (Gotwals et al., 2015; Oswald, 2013; Ruiz-Primo & Furtak, 2006). Studien, welche die Nutzung im Zusammenhang mit der Qualität von Peer-Assessment im alltäglichen Unterricht untersuchen, sind der Autorenschaft nicht bekannt.

2.2 Oberflächen- und Tiefenstrukturen des Unterrichts

Die Studien, welche Oberflächenstrukturen des Unterrichts in Bezug zu den Tiefenstrukturen des Unterrichts untersuchen, verwenden z.B. eine Aufteilung in öffentlichen Unterricht und Arbeitsphasen (Hess & Lipowsky, 2020) oder die unterschiedlichen methodisch-didaktischen Inszenierungsmuster des Unterrichts (Hugener, 2008) als Oberflächenstruktur. In der Studie von Hugener (2008) wurden mittels Videoanalyse verschiedene inhaltsbezogene Aktivitäten erfasst und auf dieser Basis zehn didaktische Inszenierungsmuster identifiziert. Es konnten Zusammenhänge zwischen diesen oberflächenstrukturellen Inszenierungsmustern und der kognitiven Aktivierung (Tiefenstruktur) gefunden werden. Eine problemlösend-entdeckende Inszenierung wurde im Vergleich zu anderen Inszenierungsmustern als kognitiv aktivierender eingeschätzt (Hugener, 2008).

Da Leistungseffekte hauptsächlich den Tiefenstrukturen, d. h. der Unterrichtsqualität, zugeschrieben werden, beziehen sich Forschungsdesiderate im Wesentlichen auf die Tiefenstrukturen des Unterrichts. Unterrichtsqualität gibt in der quantitativen Unterrichtsforschung an, wie effektiv Schülerinnen und Schüler bestimmte Ziele erreichen. Im deutschsprachigen Raum haben sich zur Charakterisierung der Unterrichtsqualität die drei Basisdimensionen „Klassenführung“, „konstruktive Unterstützung“ und „kognitive Aktivierung“ (Praetorius et al., 2021) etabliert. Die Dreiteilung der Unterrichtsqualität und deren Effektivität hinsichtlich Leistung und Motivation der Schülerinnen und Schüler konnten empirisch nachgewiesen werden. Insbesondere aufgrund der nicht einheitlichen Operationalisierung von Unterrichtsqualität in den jeweiligen Studien variieren jedoch die gemessenen Effekte (Praetorius et al., 2021). Im amerikanischen Raum wurde parallel zu den Basisdimensionen von Pianta, Hamre und Mintz (2012) ebenfalls ein Unterrichtsqualitätsmodell mit drei Bereichen und ein entsprechendes Instrument entwickelt. Das Instrument, das Classroom Assessment Scoring System, kurz CLASS, unterscheidet zwischen den unterschiedlichen Schulstufen, ist jedoch fachunspezifisch. Die drei Bereiche „emotionale Unterstützung“, „Klassenmanagement“ und „Lernunterstützung“ sind dem deutschsprachigen Ansatz ähnlich. Aufgrund des Vorliegens eines elaborierten Beobachtungsinstruments (Pianta et al., 2012), das auch in der vorliegenden Studie verwendet wurde, wird im Folgenden auf das Unterrichtsqualitätsmodell von CLASS Bezug genommen.

Peer-Assessment ist gemäss CLASS eng mit dem Qualitätsbereich Lernunterstützung verknüpft. Herausfordernde Peer-Assessments aktivieren das Vorwissen der Schülerinnen und Schüler, binden sie in konstruktive Lernprozesse ein und leisten somit einen Beitrag zu einem kognitiv aktivierenden Unterricht (Pinger et al., 2018). Peer-Assessment wird somit weitgehend als Aspekt der Lernunterstützung in Instrumenten zur Unterrichtsqualitätsmessung abgebildet. Einzig das im vorangehenden Kapitel erwähnte Qualitätsmerkmal von Peer-Assessment – der Einbezug und die Nutzung der Informationen aus dem Peer-Assessment für das weitere Lernen – wird in CLASS und auch den Basisdimensionen nicht abgebildet (Buholzer et al., 2023).

Empirisch wurde die Verbindung von formativem Assessment, wovon Peer-Assessment ein Teil ist, und Unterrichtsqualität in einer Studie von Decristan et al. (2015) untersucht. Die Wirkung formativen Assessments auf die Leistungen der Schülerinnen und Schüler war im Vergleich zur Kontrollgruppe grösser, wenn die kognitive Aktivierung und die Lernunterstützung auf einem hohen Niveau rangierten. Diese positive Wechselwirkung zeigt sich jedoch nicht für das Klassenmanagement. Daraus schlussfolgern Decristan et al. (2015), dass Unterricht am effektivsten ist, wenn spezifische Lehrmethoden mit einer hohen Unterrichtsqualität kombiniert werden.

2.3 Die vorliegende Studie

Qualitätskriterien, welche gut umgesetztes formatives Peer-Assessment im Unterricht beschreiben, können, wie vorangehend aufgezeigt, aus der Forschungsliteratur abgeleitet werden als eine kognitiv aktivierende Aufgabenstellung, Hilfestellung durch die Lehrperson und die Nutzung der aus den Peer-Assessments gewonnenen Informationen für die Optimierung der folgenden Lehr- und Lernprozesse. Wie gute formative Peer-Assessments im Unterricht methodisch ausgestaltet werden, ist jedoch nicht bekannt.

In Bezug auf die Tiefen- und Oberflächenstrukturen des Unterrichts in den qualitativen formativen Peer-Assessment eingebettet ist, kann aufgrund der Ergebnisse von Decristan et al. (2015) vermutet werden, dass ein Zusammenhang zwischen formativem Peer-Assessment und der Unterrichtsqualität besteht. Ob sich der Einsatz guten Peer-Assessments auch auf der Oberflächenstruktur des Unterrichts zeigt, kann aufgrund der bislang vorliegenden Erkenntnisse nicht aufgezeigt werden.

Mit dem Ziel mehr über die Realisierung und die Unterrichtsgestaltung von qualitativem Peer-Assessment im alltäglichen Unterricht in Erfahrung zu bringen, wurden zwei Fragestellungen formuliert:

- a.) Wie werden Peer-Assessments mit einer hohen Qualität im alltäglichen Mathematikunterricht gestaltet?
- b.) Welche Zusammenhänge zeigen sich zwischen der Umsetzung einer hohen Peer-Assessment-Qualität und der Unterrichtsgestaltung (Oberflächenstruktur) sowie der Unterrichtsqualität insgesamt?

3 Methodisches Vorgehen

3.1 Forschungsdesign

Für die Untersuchung der Fragestellung wurden die Daten der Videobeobachtungsstudie TUFA verwendet. Die Videoerfassung erfolgte im ersten Halbjahr 2017 in Deutschschweizer Schulklassen der vierten Primarstufe. Von allen Klassen wurde eine Doppelstunde im Fach Mathematik auf Video aufgezeichnet. Um vergleichbare Rahmenbedingungen zu gewährleisten, wurde das Unterrichtsthema „Einführung in die halbschriftliche Division“ vorgegeben. Da keine weiteren Vorgaben zur methodisch-didaktischen Gestaltung bestanden, kann davon ausgegangen werden, dass die Lehrpersonen weitgehend ihren alltäglichen Unterricht realisierten und daraus ökologisch valide Aussagen generiert werden können. Die auf Video aufgezeichneten Stunden repräsentieren folglich den Mathematikunterricht im Allgemeinen, da subjektive Theorien und Unterrichtsskripte als relativ stabil angesehen werden können (Groeben et al., 1988; Pianta & Hamre, 2009). Die Lehrpersonen wurden erst nach der Datenerhebung über das Projektthema informiert.

Geschulte Forschungsgruppenmitglieder und Hilfskräfte führten die Videoaufnahmen nach einem detaillierten Drehbuch durch. Es wurde eine Kamera verwendet, die der Lehrperson folgte. Die Lehrperson trug zudem ein Ansteckmikrofon, um auch geflüsterte Interaktionen mit den Schülerinnen und Schülern zu erfassen.

3.2 Stichprobe

Rekrutiert wurde die Stichprobe der Schulklassen über Anfragen bei Schulleitenden und Schulbehörden in sechs Deutschschweizer Kantonen (Aargau, Luzern, Nidwalden, Obwalden, Uri und Zug). Insgesamt nahmen 52 Schulklassen mit 52 Lehrpersonen und 634 Schülerinnen und Schülern an der Studie teil. Von den ursprünglich 922 Schülerinnen und Schülern mussten diejenigen aus anderen Klassenstufen (205), Teilnahmeverweigernde (24) und Schülerinnen und Schüler, die krank oder hauptsächlich von der schulischen Heilpädagogin, dem schulischen Heilpädagogen unterrichtet wurden, ausgeschlossen werden.

Teilgenommen haben 40 Lehrerinnen und 12 Lehrer im Alter von durchschnittlich 36 Jahren ($SD = 10.6$; Range: 25–60 Jahre) mit einer durchschnittlichen Berufserfahrung von 10,6 Jahren ($SD = 10.3$; Range: 1,5–39 Jahre). Sie unterrichteten im Durchschnitt 22,1 Lektionen pro Woche in der gefilmten Klasse ($SD = 5.1$; Range: 4–29 Lektionen).

Aus den Schulklassen nahmen 315 Mädchen und 315 Jungen an der Studie teil, bei vier Kindern gab es keine Angaben zum Geschlecht. 35,6 % der 634 Schülerinnen und Schüler sprachen mit mindestens einem Elternteil eine andere Sprache als (Schweizer) Deutsch. Das Durchschnittsalter betrug 10,5 Jahre ($SD = 0.49$; Range: 9–12,6 Jahre).

Die Teilnahme an der Studie war freiwillig. Sowohl die Lehrpersonen als auch die Erziehungsberechtigten der teilnehmenden Schülerinnen und Schüler gaben ihr akti-

ves Einverständnis zur Teilnahme. Schülerinnen und Schüler ohne Einwilligung zur Studienteilnahme sassen ausserhalb des Erfassungsraums der Kamera.

3.3 Instrumente

Für die Analyse der 52 durchschnittlich 81 Minuten langen Unterrichtsvideos wurden verschiedene Coding- und Ratinginstrumente verwendet. Wie in anderen Videostudien (z. B. Hess et al., 2019; Hugener et al., 2006) wurde ein mehrschichtiges, hierarchisch strukturiertes Beobachtungsverfahren eingesetzt, welches mit der Erfassung der Oberflächenstruktur beginnt. Für die Erfassung der Oberflächenstruktur des Unterrichts eignen sich niedrig inferente Codierungen (Pauli & Reusser, 2006), d. h. innerhalb meist kurzer Analyseabschnitte werden einfach und nahezu objektiv beobachtbare Ereignisse einzelnen Kategorien zugewiesen. Die Kategorienzuordnung erfordert ein geringes pädagogisch-didaktisches Verständnis sowie wenige schlussfolgernde Kognitionen (Clausen et al., 2003).

Im TUFA-Projekt wurde als Grundlage eine erste Ebene mit beschreibenden niedrig inferenten Codierungen gesetzt, um Lektionsdauer, mathematische Unterrichtszeit, Sozialform und Unterrichtsgestaltung zu bestimmen. Für jedes 10-Sekunden-Intervall des Unterrichts wurde ein Code den entsprechenden Kategorien zugeordnet. Mittelinferente Codings wurden anschliessend verwendet, um ausgewählte Codes zu Peer-Assessments und anderen formativen Assessment-Strategien zu präzisieren.

Die anschliessenden Qualitätseinschätzungen beziehen sich entweder nur auf vorangehende codierte Videoausschnitte oder auf den gesamten Unterricht. Für die Erfassung von Tiefenstrukturen wurden Ratings eingesetzt. Als hoch inferente Verfahren gehen Ratings über die Ebene der konkret beobachtbaren Ereignisse hinaus und erfassen komplexe Prozesse und Merkmale auf einer höheren Abstraktionsebene. Dies erfordert mehr interpretative Schlussfolgerungen und ganzheitlichere Einschätzungen (Clausen et al., 2003). Die meist längeren Zeitabschnitte werden von trainierten Beobachtenden analysiert (Lotz et al., 2013). Die Instrumente bestehen meistens aus zu Skalen zusammengestellten Items. Zu den Items werden Verhaltensindikatoren definiert und für jedes Item gesamthaft beurteilt (Praetorius, 2013). Mit hoch inferenten Ratingssystemen werden u. a. Qualitätseinschätzungen möglich (Pauli & Reusser, 2006). Tabelle 1 gewährt einen Überblick über die Anwendung der einzelnen Instrumente.

Tabelle 1: Erfassungsinstrumente von Peer-Assessment (PA) sowie Tiefen- und Oberflächenstrukturen von Unterricht

Instrument	Unterrichtsgestaltung	Gestaltung von Peer-Assessment	Peer-Assessment Qualität	Unterrichtsqualität
<i>Inferenz</i>	Mittel	Mittel	Hoch	Hoch
<i>Verfahren</i>	Codierung/ Kategoriensystem	Codierung/ Kategoriensystem	Rating	Rating
<i>Inhalt</i>	Lehr- und Lernphasen im Unterricht	Art und Dauer von Anleitung und Aktivität zum Peer-Assessment	Qualität der Umsetzung des Peer-Assessments	Unterrichtsqualität
<i>Zeitabschnitt und Analyseeinheit</i>	Gesamter mathematikbezogener Unterricht 10-Sekunden-Intervalle	Zeitabschnitt Peer-Assessment aufgeteilt in 10-Sekunden-Intervalle	Gesamte Zeitintervalle Peer-Assessment	Gesamter Unterricht, Analyse von Abschnitten von 15 bis 20 Minuten
<i>Vorgehen Daten-gewinnung</i>	Jedem 10-Sekunden-Intervall wird eine der acht Kategorien zugeordnet	Bestimmung der Art der Anleitung und Aktivität, Abgrenzung der einzelnen PA-Aufgaben.	Bewertung der drei Items auf einer vierstufigen Skala von „nicht vorhanden“ (0) über „gering“ (1) und „mittel“ (2) bis „hoch“ (3) und Bestimmung des Mittelwerts	Bewertung der 12 Dimensionen auf einer Skala von 1 bis 7 für jeden Unterrichtsabschnitt; pro Schulklasse mindestens vier Unterrichtsabschnitte raten und mitteln
<i>Quelle</i>	Theorie- und datengeleitete Eigenentwicklung, Erstellung eines Manuals für jedes Instrument			Classroom Assessment Scoring System – CLASS (Pianta et al., 2012)
<i>Interrater</i>	Cohens Kappa = 0.88	Cohens Kappa = 0.95	ICC = 0.93	ICC = 0.83

Die Instrumente in den ersten drei Spalten in Tabelle 1 sind eigenentwickelt und bauen aufeinander auf. Im folgenden Kapitel werden die Inhalte und der Aufbau dieser Instrumente aufgezeigt (für weitere Ausführungen siehe auch Buholzer et al., 2020). In einem weiteren Unterkapitel wird das CLASS-Instrument kurz vorgestellt.

3.3.1 Unterrichtsgestaltung und Peer-Assessment

Unterrichtsgestaltung: Für die gesamte mathematische Unterrichtszeit wurde in Anlehnung an Hugener (2008) für die einzelnen 10-Sekunden-Intervalle entschieden, welche Unterrichtsaktivität im Zentrum des Lernprozesses steht (siehe Tabelle 5 im Anhang). Es wurde dabei darauf geachtet, mit welchen Aktivitäten die Mehrheit der Schülerinnen und Schüler beschäftigt ist (Hugener, 2008).

Gestaltung von Peer-Assessment: Die Phase „6 Präsentieren, erheben und beurteilen der Ergebnisse/des eigenen Könnens“ (siehe Tabelle 5 im Anhang) diente u.a. als Grundlage für die Codierung von Peer- und Self-Assessment. In einem weiteren Codiervorgang wurden die Zeitabschnitte mit Peer- und Self-Assessment separat codiert.

In Anlehnung an Wylie und Lyon (2013) wurde der Operationalisierung von Peer-Assessment ein breites Verständnis zugrunde gelegt. Als Peer-Assessment wurden alle Zeitabschnitte codiert, während derer sich die Schülerinnen und Schüler in Gruppen eigenständig oder unter Moderation der Lehrperson rückblickend zu Produkten (Rechenwege, Lösungen und Fehler) anderer austauschen, diese diskutieren, vergleichen oder beurteilen. Das Arrangement muss dabei nicht wechselseitig sein, sondern kann sich auch einzig auf das Produkt von einem oder wenigen Lernenden konzentrieren.

Für die als Peer-Assessment codierten Zeitabschnitte wurden die einzelnen Peer-Assessment-Aufgaben genauer bestimmt. Eine Aufgabe besteht aus einer Anleitung und einer nachfolgenden Aktivität, bzw. Aufgabebearbeitung. Die Anleitung beinhaltet Formulierung und Hinweise zur nachfolgenden Bearbeitung einer Aufgabe (Kürzinger et al., 2018). Eine Aufgabe zielt darauf ab, sich mit einem bislang unbekanntem Aspekt zu befassen (Hugener et al., 2006) und beginnt, sobald die Lehrperson eine Anleitung zu einer PA-Aufgabe gibt, und endet, wenn die Lehrperson die PA-Aktivität abschliesst und einen neuen Auftrag erteilt oder der Unterricht endet.

Für die Codierung wurde in einem ersten Schritt entschieden, ob es sich bei den vormals codierten PA-Intervallen um eine Anleitung oder eine Aktivität handelt. Anschliessend wurde die Art der Anleitung und Aktivität genauer bestimmt. Die Beschreibungen der einzelnen Kategorien und Abgrenzungen listet Tabelle 6 (im Anhang) auf.

Qualität von Peer-Assessment: Für die vormals als Peer-Assessment codierten Zeitintervalle wurde für jede Lehrperson ein Wert zur Qualität des Peer-Assessments vergeben. Das zu diesem Zweck eingesetzte Ratinginstrument wurde theorie- und datengeleitet entwickelt. Die Qualität von Peer-Assessment wurde anhand von drei Items bewertet und gemittelt (siehe auch Buholzer et al., 2020 und Zulliger et al., 2022):

- (1) **Anspruchsniveau:** Inwieweit sind die PA-Aufgabenstellungen kognitiv aktivierend und anspruchsvoll und regen die Schülerinnen und Schüler dazu an, den Lernprozess der anderen Peers zu beschreiben und zu bewerten? Die Schülerinnen und Schüler müssen ihre Einschätzungen begründen.
- (2) **Anleitung, Durchführung und Unterstützung durch die Lehrperson:** Die Lehrperson unterstützt und leitet die PA-Aufgabe der Schülerinnen und Schüler an. Die Anweisungen sind eindeutig formuliert.
- (3) **Nutzung der gewonnenen Informationen für das weitere Lernen und Lehren:** Es findet eine Diskussion der Bewertungsergebnisse statt. Die Resultate werden für das weitere Lernen genutzt.

3.3.2 Unterrichtsqualität

Das Beobachtungsinstrument *Classroom Assessment Scoring System* (CLASS) ist ein hoch inferentes fachunabhängiges Beobachtungsinstrument, das insbesondere in den USA weite Verbreitung findet (Buell et al., 2017). Das CLASS teilt die Klassenzimmer-Interaktionen in die drei Bereiche emotionale Unterstützung, Klassenmanagement und Lernunterstützung (*emotional support, classroom organization, instructional support*). Diese Bereiche sind als latente Struktur für die Organisation des Lehrpersonenverhaltens vorhanden. Die Bereiche werden weiter differenziert in Dimensionen, welche die konkreten Merkmale und Verhaltensmarker umfassen (Pianta et al., 2012).

Zusammenfassend kann mit den CLASS-Werten veranschaulicht werden, wie produktiv die Lernumgebung bezüglich Zeit und Aktivitäten genutzt wird, wie sensitiv die Lehrperson auf die Schüleranliegen reagiert, wie gut die Qualität von Instruktionen und Feedback und der effektive Umgang mit Unterrichtsstörungen und wie kognitiv stimulierend die jeweiligen Aktivitäten und Interaktionen sind (Pianta et al., 2005). In Tabelle 7 sind die einzelnen Bereiche und Dimensionen des CLASS aufgezeigt.

Die Beobachtung mit CLASS umfasst pro Schulklasse/Lehrperson ca. 80 Minuten, aufgeteilt in vier Zyklen (je 20 Min.). Gemäss dem *Implementation Guide* von CLASS (Hamre et al., 2009) kann mit vier Beobachtungszyklen ein reliabler Wert erreicht werden. Jeder Dimension wird ein Wert von 1, 2 (tief), 3, 4 oder 5 (mittel) bis 6 oder 7 (hoch) gegeben und für die Bestimmung der übergeordneten Bereiche gemittelt (Pianta et al., 2012).

4 Ergebnisse

Die Ergebnisse zu den zwei Fragestellungen nach der Oberflächenstruktur qualitativen formativen Peer-Assessments und den Zusammenhängen qualitativen formativen Peer-Assessments mit den Oberflächen- und Tiefenstrukturen des Unterrichts insgesamt, werden im Folgenden aufgezeigt.

4.1 Gestaltung qualitätsvollen Peer-Assessments im Unterricht

Sechsvierzig Lehrpersonen (88 %) setzten Peer-Assessments in den 90-minütigen Einführungslektionen alltäglichen Mathematikunterrichts ein. Peer-Assessments beanspruchten durchschnittlich 9.3 Minuten ($SD = 7.2$; Range: 0.8–33.2 Minuten) Unterrichtszeit. Die Ratings verweisen auf eine eher mittelmässige Qualität der durchgeführten Peer-Assessment ($n = 46$; $M = 1.7$; $SD = 0.7$) (siehe auch Buholzer et al., 2020).

Für die vorliegende Studie wurden die für das Peer-Assessment aufgewendeten Zeitabschnitte hinsichtlich der Art der Aufgaben analysiert. Von den 46 Lehrpersonen, die Peer-Assessment im Unterricht umsetzten, haben 31 im Laufe der 90-minütigen Einführungslektion einmal eine PA-Aufgabe durchgeführt. Zwölf Lehrpersonen haben

zwei PA-Aufgaben realisiert und zwei Lehrpersonen haben dreimal eine PA-Aufgabe ermöglicht. Eine Lehrperson hat zwar Peer-Assessment durchgeführt, jedoch ohne erkennbare Anleitung (Kategorie „Rest“).

In Tabelle 2 sind die Aufgaben absteigend nach Anspruchsgehalt der Anleitung und Anzahl der Aktivitäten geordnet.

Bezüglich Häufigkeit und Dauer der einzelnen Aufgaben zeigt sich deutlich eine starke Häufung bei den Aufgaben „Beschreibung mit moderierter Aktivität“. Fast 60 % aller Aufgaben konnten dieser Kategorie zugeordnet werden. Aufgaben mit eigenständigen Gruppenarbeiten sind weniger häufig, dauern jedoch länger.

Tabelle 2: Häufigkeit und Dauer der Aufgaben zu Peer-Assessment

Aufgabe Peer-Assessment	Häufigkeit		Dauer in Minuten	
	Anzahl	%	M (SD)	Min.–Max.
1 Beschreibung und Bewertung mit eigenständiger und moderierter Aktivität	2	2,9	17.7 (17.4)	5.3–30.0
2 Beschreibung mit eigenständiger und moderierter Aktivität	9	13,0	12.2 (9.2)	3.2–30.0
3 Beschreibung und Bewertung mit eigenständiger Aktivität	2	2,9	8.9 (2.2)	7.3–10.5
4 Beschreibung und Bewertung mit moderierter Aktivität	2	2,9	2.8 (1.3)	1.8–3.7
5 Beschreibung mit eigenständiger Aktivität	5	7,2	6.8 (3.1)	3.8–11.7
6 Beschreibung mit moderierter Aktivität	41	59,4	5.0 (3.8)	0.8–18.8
7 Rest (nicht zuordenbar)	8	11,6	2.3 (1.4)	0.2–4.0

$n = 46$

In Abbildung 1 sind die Schulklassen mit PA-Zeitanteilen in absteigender Rangfolge der Qualitätseinschätzungen aufgeführt. Dauer und Art der PA-Aufgaben wurden in den Balken abgetragen. Wie bereits in Buholzer et al. (2020) erwähnt zeigt sich eine signifikante Korrelation zwischen der Dauer und der Qualität von Peer-Assessment von $r = .80, p < .01$ – in der Abbildung ist dies deutlich zu erkennen. Lehrpersonen mit den höchsten Qualitätsausprägungen (Höchstwert von drei) weisen wesentlich längere Anteile von Peer-Assessment auf als die Lehrpersonen mit geringen Qualitätseinschätzungen. Einführungslektionen mit Anteilen eigenständiger wie auch moderierter PA-Aktivitäten wurden mit höheren Qualitätswerten beurteilt.

Peer-Assessments der Art „moderierte Beschreibung“ sind am häufigsten, werden bezüglich ihrer Qualität jedoch eher geringwertig eingeschätzt.

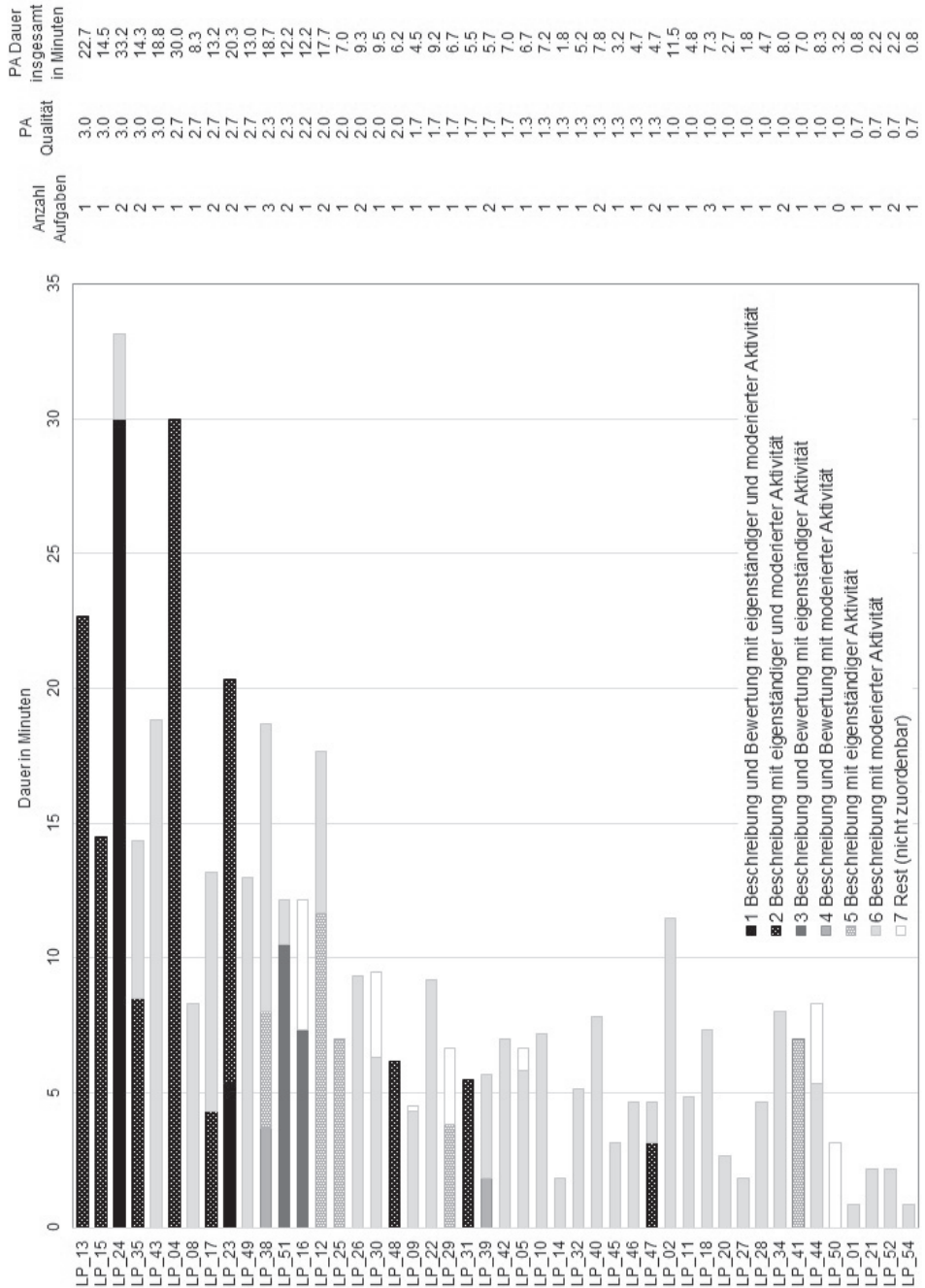


Abbildung 1: Gestaltung und Qualität von Peer-Assessment
 n = 46

4.2 Unterrichtsqualität und Qualität von Peer-Assessment

Die Tiefenstruktur und damit die Unterrichtsqualität wurden mit dem *Classroom Assessment Scoring System* (CLASS) erfasst und mit den Werten zur Qualität von Peer-Assessment korreliert. In Tabelle 3 sind die Mittelwerte, Standardabweichungen, Minimal- und Maximalwerte der einzelnen Variablen, die Skalenwerte und die Korrelationen aufgeführt. Die Dimensionen von CLASS wurden zu den vormals mit grösseren Stichproben geprüften Bereichen (Pianta et al., 2012) zusammengefasst und deren Skalenwerte ermittelt.

Wie in vorangehenden Studien (Decristan et al., 2015) zeigen sich auch hier Zusammenhänge zwischen der Qualität von Peer-Assessment und den Bereichen „emotionale Unterstützung“ ($r = .32, p < .01$) und „Lernunterstützung“ ($r = .40, p < .001$). Keinerlei Zusammenhang besteht zwischen der Qualität von Peer-Assessment und dem Klassenmanagement.

Eine genauere Betrachtung des Zusammenhangs zwischen der Qualität von Peer-Assessment und der emotionalen Unterstützung auf der Ebene der Items und Dimensionen offenbart eine mittlere Korrelation zwischen der Nutzung der Informationen aus dem Peer-Assessment für den weiteren Unterricht und dem positiven Klima. Des Weiteren korrelieren alle Items von PA-Qualität mit der CLASS-Dimension „Berücksichtigung der Schülerperspektiven“ ($r = .48, p < .001$). In Bezug auf die einzelnen Items der Qualität von Peer-Assessment korreliert insbesondere das Item „Nutzung“ mit der Berücksichtigung der Schülerperspektive. Dieser Zusammenhang kann damit erklärt werden, dass die Schülerinnen und Schüler am Schluss des Peer-Assessments anhand der gewonnenen Informationen teilweise wählen dürfen, wo sie weiterarbeiten wollen.

Eine mittlere Korrelation zeigt sich auch zwischen der PA-Qualität und dem Bereich Lernunterstützung ($r = .40, p < .001$). Von den fünf Dimensionen der Lernunterstützung korrelieren die drei Dimensionen „Analyse und Erkundung“, „Qualität des Feedbacks“ und „Unterrichtsgespräche“ mit der PA-Qualität. Qualitativ gutes Peer-Assessment geht somit einher mit einem Unterricht, der durch Analyse und Erkundung Denkprozesse anregt und sich durch gutes Feedback und strukturierte, kumulative Unterrichtsgespräche auszeichnet.

Tabelle 3: Zusammenhänge zwischen Unterrichtsqualität und Qualität von Peer-Assessment

Skalen (α)	M ^a (SD)	Min- Max.	Korrelationen zwischen den Skalen bzw. Items/Dimensionen														
Items / Dimensionen			5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Qualität PA (ω = 0,92)	1.5 (.9)	0– 3.0	.32*	.28*	.06	.48**	.14	.17	.22	.08	-.18	.40**	.18	.09	.47**	.42**	.41**
2. Anleitung	1.8 (1.0)	0– 3.0	.27	.20	.06	.42**	.15	.20	.30	.08	-.19	.38**	.21	.13	.39**	.42**	.35*
3. Anspruchsniveau	1.5 (.9)	0– 3.0	.25	.25	-.03	.39**	.12	.13	.16	.06	-.13	.33*	.12	.04	.45**	.33*	.32*
4. Nutzung	1.2 (1.0)	0– 3.0	.38**	.33*	.12	.51**	.11	.14	.18	.07	-.18	.41**	.17	.07	.46**	.40**	.47**
5. Emotionale Unterstützung (0,833)	4.8 (.6)	3.5– 5.9	–	.93**	.87**	.81**	.61**	.45**	.37**	.49**	-.45**	.82**	.77**	.55**	.56**	.71**	.72**
6. Positives Klima	5.3 (.8)	3.6– 6.8	–	–	.79**	.62**	.53**	.33*	.25	.40**	-.38**	.73**	.67**	.47**	.50**	.71**	.59**
7. Sensitivität der Lehrperson	6.0 (.5)	4.7– 6.9	–	–	.52**	.72**	.61**	.51**	.68**	.68**	-.49**	.69**	.72**	.58**	.35**	.63**	.56**
8. Berücksichtigung Schülerperspektive	2.9 (.6)	1.2– 4.4	–	–	–	.38**	.27*	.26	.24	.24	-.33*	.72**	.64**	.42**	.60**	.51**	.75**
9. Schulisches Engagement ^b	6.2 (.5)	4.6– 7.0	–	–	–	–	.84**	.78**	.81**	.81**	-.53**	.58**	.64**	.44**	.39**	.46**	.45**
10. Klassenmanagement (0,860)	6.4 (.4)	4.6– 7.0	–	–	–	–	–	–	.96**	.93**	-.59**	.47**	.44**	.45**	.27*	.41**	.34*
11. Verhaltensmanagement	6.5 (.5)	4.0– 7.0	–	–	–	–	–	–	–	.79**	-.64**	.40**	.34*	.35*	.31*	.35**	.28*
12. Produktivität	6.4 (.4)	5.2– 7.0	–	–	–	–	–	–	–	–	–	.49**	.54**	.52**	.19	.43**	.38**
13. Negatives Klima ^b	1.1 (.2)	1.0– 1.7	–	–	–	–	–	–	–	–	–	-.34*	-.22	-.28*	-.33*	-.32*	-.24
14. Lernunterstützung (0,866)	3.8 (.5)	2.6– 4.8	–	–	–	–	–	–	–	–	–	–	.76**	.76**	.73**	.87**	.90**
15. Lehr- und Lernformate	5.2 (.5)	3.8– 6.3	–	–	–	–	–	–	–	–	–	–	–	.55**	.43**	.55**	.65**
16. Inhaltsverständnis	3.7 (.5)	2.5– 4.8	–	–	–	–	–	–	–	–	–	–	–	–	.35*	.64**	.63**
17. Analyse und Erkundung	3.1 (.7)	2.0– 5.7	–	–	–	–	–	–	–	–	–	–	–	–	–	.53**	.63**
18. Qualität des Feedbacks	3.9 (.8)	2.3– 5.7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	.74**
19. Unterrichtsgespräche	3.2 (.7)	1.9– 4.7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

n = 52

^a Einschätzung der Items von PA-Qualität auf vierstufiger Skala (0–3) und der Dimensionen von CLASS auf siebenstufiger Skala (1–7).

^b Schulisches Engagement (9) und Negatives Klima (10) wurden von der Skalenbildung ausgeschlossen.

* Die Pearson-Produkt-Moment-Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

** Die Pearson-Produkt-Moment-Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant.

4.3 Unterrichtsgestaltung und Qualität von Peer-Assessment

Die Ergebnisse zu den Angaben der Zeitdauer der Unterrichtsphasen und den Zusammenhängen mit der Qualität von Peer-Assessment sind in Tabelle 4 aufgeführt.

Die Angaben zur Zeitdauer der Unterrichtsphasen verweisen auf grosse Unterschiede bezüglich der Gestaltung der 90-minütigen Einführungslektion und eine beträchtliche Streuung bezüglich der Zeitdauer für die einzelnen Unterrichtsphasen. Die Minimal- und Maximalwerte zeigen zudem, dass nur die Unterrichtsphasen Konfrontation mit neuem Problem und zusammenfassender Rückblick bei allen Lehrpersonen vorkommen. Die übrigen Unterrichtsphasen kommen nicht bei allen Lehrpersonen vor.

Die Qualität von Peer-Assessment korreliert erwartungsgemäss am stärksten ($r = .82$, $p < .001$) mit der Unterrichtsphase „Präsentieren, erheben und beurteilen der Ergebnisse/des eigenen Könnens“, welche u.a. die Codierungen zu Peer- und Self-Assessment enthält. Eine negative Korrelation von $r = -.37$, $p < .001$ besteht zwischen der Qualität von Peer-Assessment und dem „Problem-/Aufgabenstellung im Klassenunterricht erarbeiten“, welche auf eine eher traditionell transmissive Unterrichtsgestaltung hinweist.

5 Diskussion

Mittels Coding- und Ratingverfahren wurden 52 auf Video aufgezeichnete 90-minütige Einführungslektionen alltäglichen Mathematikunterrichts hinsichtlich der methodischen Umsetzung und Qualität von formativem Peer-Assessment untersucht und mit weiteren Unterrichtsmerkmalen der Oberflächen- und Tiefenstruktur des Unterrichts verglichen. Ziel war es, zu analysieren wie gutes Peer-Assessment auf der Oberfläche gestaltet werden kann und welche Zusammenhänge zwischen gutem Peer-Assessment und der Qualität und der Gestaltung des Unterrichts insgesamt bestehen. Daraus lassen sich Hinweise ableiten, wie effektives Peer-Assessment im Unterricht gestaltet und implementiert werden kann.

Von den 52 Lehrpersonen hat die Mehrheit (46 Lehrpersonen) Peer-Assessment in den untersuchten 90-minütigen Einführungslektionen durchgeführt. Überwiegend wurde von diesen Lehrpersonen nur eine Aufgabenstellung zu Peer-Assessment durchgeführt. Am häufigsten besprachen die Lehrperson mit den Schülerinnen und Schülern zusammen den Lernprozess oder ein Lernergebnis eines Peers (moderierte Beschreibung). Seltener kamen zeitintensivere Peer-Assessments mit zwei Aktivitäten vor, d. h. einer eigenständigen Bearbeitung durch die Schülerinnen und Schüler und einer durch die Lehrperson moderierten Phase.

Die Ergebnisse zu der ersten Frage nach der Ausgestaltung von gutem Peer-Assessment im alltäglichen Mathematikunterricht verweisen auf die Bedeutung der Realisierung von eigenständigen Aktivitäten der Schülerinnen und Schülern mit anschliessenden lehrpersonenmoderierenden Aktivitäten. Länger andauernde Peer-Assessments, die sowohl das eigenständige Peer-Assessment der Lernenden als auch eine Zusammen-

Tabelle 4: Dauer Unterrichtsphasen (Minutenangaben) und Zusammenhang mit PA-Qualität

	Dauer in Minuten		Korrelationen									
	M (SD)	Min.-Max.	5	6	7	8	9	10	11	12		
<i>1. Qualität Peer-Assessment</i>												
1. Anleitung	-	-	-0,15	-0,20	-0,37**	,24	-0,13	,82**	,16	-0,29*		
2. Anspruchsniveau	-	-	-0,17	-0,15	-0,39**	,18	-0,08	,76**	,17	-0,22		
3. Nutzung	-	-	-0,14	-0,25	-0,28*	,22	-0,10	,76**	,06	-0,22		
Unterrichtsphasen	-	-	-0,10	-0,15	-0,37**	,29*	-0,18	,78**	,20	-0,38**		
5. Konfrontation mit neuem Problem/Aufgabe	3,5 (3,1)	3-14,0	-	,27*	,03	,02	-0,40**	-0,05	,04	,18		
6. Vorläuferfertigkeiten trainieren	5,3 (9,4)	0-50,7	-	-	,04	-0,09	-0,55**	-0,19	,24	,13		
7. Problem-/Aufgabenstellung im Klassenunterricht erarbeiten	14,3 (6,8)	0-28,0	-	-	-	-0,11	-0,09	-0,50**	-0,007	,11		
8. Problem-/Aufgabenstellung eigenständig bearbeiten	10,3 (8,1)	0-33,0	-	-	-	-	-0,51**	,12	-0,03	,10		
9. Bekannte Problem-/Aufgabenstellung durcharbeiten/üben	35,1 (14,2)	0-62,5	-	-	-	-	-	-0,17	-0,22	-0,25		
10. Präsentieren, erheben oder beurteilen der Ergebnisse/des eigenen Könnens ^a	9,4 (7,4)	0-33,3	-	-	-	-	-	-	,10	-0,23		
11. Zusammenfassender Rückblick	2,1 (1,4)	0,3-6,7	-	-	-	-	-	-	-	,12		
12. Sonstiges	1,6 (1,8)	0-7,7	-	-	-	-	-	-	-	-		

n = 52

^a Peer- und Self-Assessment sind Teil dieser Unterrichtsphase.

* Die Pearson-Produkt-Moment-Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

** Die Pearson-Produkt-Moment-Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant.

führung dieser Assessments in einem moderierten Setting ermöglichten, erfüllten alle Qualitätskriterien, während die häufig durchgeführten und durchschnittlich nur fünf Minuten dauernden Peer-Assessments mit ausschliesslich moderierten Beschreibungen eher tiefe Qualitätseinschätzungen erhielten.

Die zweite Frage nach den Zusammenhängen zwischen der Qualität des Peer-Assessments und der mittels CLASS erfassten Unterrichtsqualität und der Unterrichtsgestaltung zeigen erwartungsgemäss eine Korrelation zwischen der PA-Qualität und den Bereichen emotionale Unterstützung und Lernunterstützung und keinen Zusammenhang zum Bereich Klassenmanagement. Genauer betrachtet korrelieren aus dem Bereich der emotionalen Unterstützung die Dimensionen „positives Klima“ und „Berücksichtigung der Schülerperspektiven“ mit der Qualität von Peer-Assessment. Von den fünf Dimensionen der Lernunterstützung korrelieren die drei Dimensionen „Analyse und Erkundung“, „Qualität des Feedbacks“ und „Unterrichtsgespräche“ mit der Peer-Assessment-Qualität. Zusammengefasst zeigen die Ergebnisse, dass gutes Peer-Assessment mit einem respektvollen, freundlichen und positiven Unterrichtsklima und einem Unterricht einhergeht, der auf den Einbezug und den Austausch mit den Schülerinnen und Schülern setzt und den Lernenden Gelegenheit zu selbstständigen Erkundungen gibt. Die Ergebnisse zum Zusammenhang zwischen der Qualität von Peer-Assessment und der Gestaltung des Unterrichts gehen in die gleiche Richtung. Im Unterricht mit fehlendem oder qualitativ schlechtem Peer-Assessment dominiert die lehrpersonenzentrierte Erarbeitung von Problem-/Aufgabenstellungen im Klassenunterricht, was einem eher transmissiven Verständnis von Unterricht entspricht, wo im Klassenunterricht mit engen Fragen und Schritt für Schritt ein Vorgehen beim halbschriftlichen Dividieren „erarbeitet“ wird. Ein solches Vorgehen benötigt viel Zeit, die dann für eine intensive individuelle Auseinandersetzung mit eignen Strategien und Vorgehensweisen fehlt.

Die Zusammenhänge der Qualität von formativem Peer-Assessment mit der Unterrichtsqualität stehen in einer Linie mit den Ergebnissen von Decristan et al. (2015) und theoretischen Überlegungen (Pinger et al., 2018; Schildkamp et al., 2020). Unterricht mit qualitativ gutem Peer-Assessment ist auch ein Unterricht mit einem positiven Lernklima, der die Perspektiven der Schülerinnen und Schüler berücksichtigt und sie zu Analysen und Erkundungen herausfordert, ihnen gutes Feedback gibt und sie in qualitativ gute Lehr-Lern-Gespräche involviert. Diese Merkmale verweisen auf ein kognitiv-sozial-konstruktivistisches Lernverständnis, welches entdeckendes und forschendes Lernen zulässt und in dem sich Lehrpersonen auch als Begleitende von Lern- und Moderierende von Austauschprozessen verstehen (Reusser, 2016).

Hochwertiges formatives Peer-Assessment wurde mit dem Rating der folgenden drei Items erfasst: Anspruchsniveau der Aufgabe, Anleitung und Support durch die Lehrperson und Nutzung der gewonnenen Informationen für den weiteren Unterricht. Die Ergebnisse zur Gestaltung formativen Peer-Assessments und zum Zusammenhang mit der Qualität von Peer-Assessment verweisen auf die Bedeutung der Begleitung und Anleitung von schülerzentrierten Lernformaten wie dem Peer-Assessment. Eine hohe

Qualität formativen Peer-Assessments wird eher dann erreicht, wenn die Lehrperson nach einer eigenständigen Schülerarbeitsphase noch eine lehrpersonenmoderierte abschliessende Aktivität integriert, die die Ergebnisse aus der eigenständigen Schülerarbeitsphase aufnimmt. Die Einplanung ausreichender Zeit und eines lehrpersonmoderierten Abschlusses können Voraussetzungen sein für die Durchführung qualitativ guten formativen Peer-Assessments. Dieses Ergebnis geht einher mit den Forderungen von Reusser (2016), demgemäss es nicht ausreicht, Schüler und Schülerinnen mit minimaler Anleitung selbstorganisiert lernen zu lassen, vielmehr brauche es eine engagierte und aktiv den Unterricht gestaltende Lehrperson zur Optimierung des Lernerfolgs. Für den Einsatz von gutem, d.h. potentiell effektivem Peer-Assessment heisst dies, dass die Lehrpersonen neben dem Aufstellen einer anspruchsvollen Aufgabenstellung, die Lernenden zum Peer-Assessment anleiten und begleiten können müssen und auch Kompetenzen zum Moderieren von Unterrichtsgesprächen mitbringen müssen. Gemäss Chen et al. (2020) fehlen den Lehrpersonen häufig diese Fähigkeiten für einen produktiven Klassendiskurs.

Die Ergebnisse zur Gestaltung von qualitativem Peer-Assessment und den Zusammenhängen mit der Unterrichtsqualität und -gestaltung zeigen einmal mehr, dass gutes formatives Peer-Assessment nicht als „add-on activity“ (Schildkamp et al., 2020, S. 3) des Unterrichts betrachtet werden kann, sondern einen Unterricht bedingt, der Peer-Assessment und allgemein Strategien formativen Assessments als eine Lehr- und Lernmethode ansieht, welche die gesamte Unterrichtspraxis durchdringt (Heritage, 2022). Der Einsatz von Peer-Assessment erfordert ein umfassendes Unterrichtsverständnis, bei dem die formative Beurteilung und die entsprechenden formativen Assessment-Strategien in den Lehr- und Lernprozess integriert und die Erkenntnisse für die weitere Steuerung dieser Prozesse genutzt werden (Lötscher & Roos, 2022; Black & William, 2018).

Limitationen

Die Auswertungen beziehen sich auf einen Ausschnitt von 90-minütigen Mathematik-ektionen. Für die Bestimmung der Unterrichtsqualität mit CLASS ist dieser Zeitabschnitt ausreichend. Bezüglich des Vorhandenseins und der Qualität von Peer-Assessment kann die Eingrenzung auf einen 90-minütigen Einführungsunterricht eventuell zu kurz greifen. Zudem wurden ausschliesslich Einführungs-ektionen untersucht. Aus der TIMS-Studie ist bekannt, dass Lehrpersonen des deutschsprachigen Teils der Schweiz zwischen Einführungs- und Übungs-ektionen unterscheiden (Pauli & Schmidt, 2019). Inwieweit sich Einführungs- und Übungs-ektionen hinsichtlich des Einsatzes von Peer-Assessment unterscheiden, müsste weiter untersucht werden.

Die Erhebungsinstrumente beziehen sich tw. auf die gleichen Zeitabschnitte und erheben ähnliche Konstrukte. So sind die Zeitabschnitte zum Peer-Assessment auch in den Analysen zur Qualität und Gestaltung des gesamten videographierten Unterrichts enthalten. Die Ratinginstrumente zeigen tw. inhaltliche Überlappungen, wodurch sich die hohen Korrelationen zum Teil erklären lassen.

Die Unterscheidung in Oberflächen- und Tiefenstrukturen ist nicht immer eindeutig, da es sich um einen fließenden Übergang handelt.

Formative Assessments werden je nach Fach unterschiedlich gestaltet mit dementsprechend unterschiedlichen Effekten (Kingston & Nash, 2011). Die vorliegenden Ergebnisse beziehen sich auf das Unterrichtsfach Mathematik. Inwieweit sich für andere Fächer übereinstimmende Ergebnisse einstellen, müsste geprüft werden.

Für die Zusammenhänge wurden Korrelationen berichtet, die anfällig sind für Ausreisser. Zudem konnte keine Wirkrichtung aufgezeigt werden.

Anmerkungen:

Das Projekt wurde vom Schweizerischen Nationalfonds (Nr. 169771) unterstützt.

Literatur

- Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 90–105). Routledge. <https://doi.org/10.4324/9780203874851>
- Andrade, H. L. & Valtcheva, A. (2009). Promoting Learning and Achievement Through Self-Assessment. *Theory into Practice*, 48(1), 12–19. <https://doi.org/10.1080/00405840802577544>
- Black, P. & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 1–25. <https://doi.org/10.1080/0969594X.2018.1441807>
- Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2004). Working inside the Black Box: Assessment for Learning in the Classroom. *Phi Delta Kappan*, 86(1), 8–21. <https://doi.org/10.1177/003172170408600105>
- Buell, M., Han, M. & Vukelich, C. (2017). Factors affecting variance in Classroom Assessment Scoring System scores: season, context, and classroom composition. *Early Child Development and Care*, 187(11), 1635–1648. <https://doi.org/10.1080/03004430.2016.1178245>
- Buholzer, A., Baer, M., Ruelmann, M. & Zulliger, S. (2023). Formatives Assessment – eine vernachlässigte Dimension von Unterrichtsqualität? *Bildung und Erziehung*, 76 (1), 43–60. <https://doi.org/10.13109/buer.2023.76.1.43>
- Buholzer, A., Baer, M., Zulliger, S., Torchetti, L., Ruelmann, M., Häfliger, A. & Lötscher, H. (2020). Formatives Assessment im alltäglichen Mathematikunterricht von Primarlehrpersonen: Häufigkeit, Dauer und Qualität. *Unterrichtswissenschaft*, 48(4), 629–661. <https://doi.org/10.1007/s42010-020-00083-7>
- Bürgermeister, A. (2014). *Leistungsbeurteilung im Mathematikunterricht. Bedingungen und Effekte von Beurteilungspraxis und Beurteilungsgenauigkeit*. Waxmann.
- Butler, G. Y. & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5–31. <https://doi.org/10.1177/0265532209346370>
- Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N. & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. *Journal of The Learning Sciences*, 29(4–5), 642–880. <https://doi.org/10.1080/10508406.2020.1783269>

- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen. Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31, 122–141. <https://doi.org/10.25656/01:6775>
- Decristan, J., Hess, M., Holzberger, D. & Praetorius, A.-K. (2020). Oberflächen- und Tiefenmerkmale – eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. In A.-K. Praetorius, J. Grünkorn & E. Klieme (Hrsg.), *Empirische Forschung zu Unterrichtsqualität: Theoretische Grundfragen und quantitative Modellierungen* (S. 102–116). Beltz Juventa.
- Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., Hondrich, A. L., Rieser, S., Hertel, S. & Hardy, I. (2015). Embedded Formative Assessment and Classroom Process Quality. *American Educational Research Journal*, 52(6), 1133–1159. <https://doi.org/10.3102/0002831215596412>
- D-EDK (2016). *Lehrplan 21. Gesamtausgabe*. <https://v-fe.lehrplan.ch/downloads.php>
- Double, K. S., McGrane, J. A. & Hopfenbeck, T. N. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educational Psychology Review*, 32(2), 481–509. <https://doi.org/10.1007/s10648-019-09510-3>
- Gotwals, A. W., Philhower, J., Cisterna, D. & Bennett, S. (2015). Using video to examine formative assessment practices as measures of expertise for mathematics and science teachers. *International Journal of Science and Mathematics Education*, 13(2), 405–423. <https://doi.org/10.1007/s10763-015-9623-8>
- Graham, S., Hebert, M. & Harris, K. R. (2015). Formative Assessment and Writing. *The Elementary School Journal*, 115(4), 523–547. <https://doi.org/10.1086/681947>
- Groeben, N., Wahl, D., Schlee, J. & Scheele, B. (1988). *Das Forschungsprogramm subjektive Theorien: eine Einführung in die Psychologie des reflexiven Subjekts*. Francke.
- Hamre, B. K., Goffin, S. G. & Kraft-Sayre, M. (2009). *Classroom Assessment Scoring System Implementation Guide: Measuring and Improving Classroom Interactions in Early Classroom Settings*. <http://www.teachstone.org/wpcontent/uploads/2010/06/CLASSImplementationGuide.pdf>
- Harris, L. R. & Brown, G. T. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101–111. <https://doi.org/10.1016/j.tate.2013.07.008>
- Heritage, M. (2022). *Formative Assessment: Making It Happen in the Classroom* (2. Aufl.). SAGE Publications. <https://doi.org/10.4135/9781071813706>
- Hess, M., Denn, A.-K. & Lipowsky, F. (2019). *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE)*. *Technischer Bericht zu den PERLE-Videostudien. Band 2: Beobachtungssysteme zur Beschreibung und Qualität von Grundschulunterricht*. GFPP; DIPF. <https://doi.org/10.25656/01:17081>
- Hess, M. & Lipowsky, F. (2020). Zur (Un-)Abhängigkeit von Oberflächen- und Tiefenmerkmalen im Grundschulunterricht. In A.-K. Praetorius, J. Grünkorn & E. Klieme (Hrsg.), *Empirische Forschung zu Unterrichtsqualität: Theoretische Grundfragen und quantitative Modellierungen* (S. 117–131). Beltz Juventa.
- Hugener, I. (2008). *Inszenierungsmuster im Unterricht und Lernqualität: Sichtstrukturen schweizerischen und deutschen Mathematikunterrichts in ihrer Beziehung zu Schülerwahrnehmung und Lernleistung – eine Videoanalyse*. Waxmann.
- Hugener, I., Pauli, C. & Reusser, K. (2006). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lern-*

- verhalten und mathematisches Verständnis“. 3. Videoanalysen. GPF u.a. <https://doi.org/10.25656/01:3130>
- Kingston, N. & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Kollar, I. & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, 20(4), 344–348. <https://doi.org/10.1016/j.learninstruc.2009.08.005>
- Kürzinger, A., Pohlmann-Rother, S., Hess, M. & Lipowsky, F. (2018). Aufgabenstellungen und ihre Bedeutung für die Textqualität von Schülerinnen und Schülern im Anfangsunterricht. *Unterrichtswissenschaft*, 46(3), 327–343. <https://doi.org/10.1007/s42010-018-0016-x>
- Li, H., Xiong, Y., Zang, X., L. Kornhaber, M., Lyu, Y., Chung, K. S. & K. Suen, H. (2016). Peer assessment in the digital age: a meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245–264. <https://doi.org/10.1080/02602938.2014.999746>
- Lotz, M., Gabriel, K. & Lipowsky, F. (2013). *Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung*. Beltz Juventa. <https://doi.org/10.25656/01:11942>
- Lötscher, H. & Roos, M. (2022). Grundlagen kompetenzorientierter Beurteilung. In H. Lötscher, M. Naas & M. Roos (Hrsg.), *Kompetenzorientiert beurteilen* (S. 12–128). Hep Verlag.
- Oswalt, S. G. (2013). *Identifying formative assessment in classroom instruction* [Doctoral thesis, Boise State University]. Boise State University Scholar Works. <https://bit.ly/3qF256K>
- Panadero, E. & Brown, G. T. L. (2017). Teachers’ reasons for using peer assessment: positive experience predicts use. *European Journal of Psychology of Education*, 32, 133–156. <https://doi.org/10.1007/s10212-015-0282-5>
- Pauli, C. & Reusser, K. (2006). Von international vergleichenden Video Surveys zur video-basierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52, 774–798. <https://doi.org/10.25656/01:4488>
- Pauli, C. & Schmid, M. (2019). Zur Didaktik guten Unterrichts: Qualitätsvollen Unterricht gestalten lernen. In U. Steffens & R. Messner (Hrsg.), *Beiträge zur Schulentwicklung: Bd. 3. Unterrichtsqualität: Konzepte und Bilanzen gelingenden Lehrens und Lernens* (S. 167–182). Waxmann.
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D. & Barbarin, O. (2005). Features of Pre-Kindergarten Programs, Classrooms, and Teachers: Do They Predict Observed Classroom Quality and Child-Teacher Interactions? *Applied Developmental Science*, 9(3), 144–159. https://doi.org/10.1207/s1532480xads0903_2
- Pianta, R. C. & Hamre, B. K. (2009). Classroom processes and positive youth development: Conceptualizing, measuring, and improving the capacity of interactions between teachers and students. *New Directions for Youth Development*, 2009(121), 33–46. <https://doi.org/10.1002/yd.295>
- Pianta, R. C., Hamre, B. K. & Mintz, S. (2012). *Classroom Assessment Scoring System. Upper Elementary*. Teachstone.
- Pinger, P., Rakoczy, K., Besser, M. & Klieme, E. (2018). Interplay of formative assessment and instructional quality – interactive effects on students’ mathematics achievement. *Learning Environments Research*, 21(1), 61–79. <https://doi.org/10.1007/s10984-017-9240-2>
- Ploegh, K., Tillema, H. H. & Segers, M. S. (2009). In search of quality criteria in peer assessment practices. *Studies in Educational Evaluation*, 35(2–3), 102–109. <https://doi.org/10.1016/j.stueduc.2009.05.001>

- Praetorius, A.-K. (2013). *Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter. Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis*. <https://doi.org/10.25656/01:13845>
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Praetorius, A.-K., Martens, M. & Brinkmann, M. (2021). Unterrichtsqualität aus Sicht der quantitativen und qualitativen Unterrichtsforschung. In T. Hascher, T.-S. Idel, & W. Helsper (Hrsg.), *Handbuch Schulforschung* (S. 1–20). Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-24734-840-1>
- Reusser, K. (2016). Jenseits der Beliebigkeit. „Konstruktivistische Didaktik“ auf dem Prüfstand der empirischen Unterrichtsforschung. *Journal für LehrerInnenbildung*, 16(2), 40–48.
- Ruiz-Primo, M. A. & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3–4), 237–263. <https://doi.org/10.1080/10627197.2006.9652991>
- Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H. & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, 109(8), 1049–1066. <https://doi.org/10.1037/edu0000190>
- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B. & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 101602. <https://doi.org/10.1016/j.ijer.2020.101602>
- Schmidt, C. (2020). *Formatives Assessment in der Grundschule*. Springer. <https://doi.org/10.1007/978-3-658-26921-0>
- Schnell, R., Hill, P. B. & Esser, E. (2013). *Methoden der empirischen Sozialforschung* (10. Aufl.). Oldenbourg Verlag.
- Topping, K. J. (2009). Peer Assessment. *Theory into Practice*, 48(1), 20–27. <https://doi.org/10.1080/00405840802577569>
- Topping, K. J. (2010). Peers as a Source of Formative Assessment. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 61–74). Routledge.
- Wylie, C. & Lyon, C. (2013). *Using the formative assessment rubrics, reflection and observation tools to support professional reflection on practice*. https://www.pblconnections.com/uploads/2/6/7/2/26722395/formative_assessment_rubrics_and_observation_tools_document.pdf
- Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L. & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, 37, 100484. <https://doi.org/10.1016/j.edu-rev.2022.100484>.
- Zulliger, S., Buholzer, A. & Ruelmann, M. (2022). Observed Quality of Formative Peer and Self-Assessment in Everyday Mathematics Teaching and its Effects on Student Performance. *European Journal of Educational Research*, 11(2), 663–680. <https://doi.org/10.12973/eu-jer.11.2.663>

Anhang

A Coding- und Ratinginstrumente

Tabelle 5: Kategoriensystem Unterrichtsgestaltung

Kategorien	Beschreibung der Kategorie (Zusammenfassung)
1 <i>Konfrontation mit neuem Problem/Aufgabe</i>	Lernziel oder Thema wird von der Lehrperson genannt, mit der Klasse besprochen. Situierung des Themas mittels Geschichten oder Bildern. Erklären einer mathematischen Situation (ohne Aufgabenstellung). Direkte Konfrontation mit Aufgabe (ohne Bearbeitung).
2 <i>Vorläuferfertigkeiten trainieren</i>	Vorläuferfertigkeiten wie z. B. einfache Division werden erhoben, trainiert oder repetiert. Die Kategorie umfasst die Anleitung und die Durchführung.
3 <i>Problem-/Aufgabenstellung im Klassenunterricht erarbeiten</i>	Unter Anleitung der Lehrperson werden Aufgaben erarbeitet, gemeinsam als Klasse gelöst oder die Lehrperson zeigt vor und erklärt.
4 <i>Problem-/Aufgabenstellung eigenständig bearbeiten</i>	Eigenständige Bearbeitung von Aufgaben (allein oder in Gruppen). Die Kategorie umfasst die Anleitung und die Durchführung.
5 <i>Bekannte Problem/Aufgabenstellung durcharbeiten/üben</i>	Das bislang Gelernte wird in Schülerarbeitsphasen geübt. Die Kategorie umfasst die Anleitung und die Durchführung.
6 <i>Präsentieren, erheben oder beurteilen der Ergebnisse/des eigenen Könnens*</i>	Die Schülerinnen und Schüler präsentieren Ergebnisse aus vorangehenden Schülerarbeitsphasen und die Lehrperson moderiert, systematisiert die Ergebnisse. Lernkontrolle oder Korrektur von Aufgaben. Self-Assessment. Peer-Assessment.
7 <i>Zusammenfassender Rückblick</i>	Die Lehrperson nimmt einen organisatorischen, fachlichen, überfachlichen oder emotional-motivationalen Rückblick vor.
8 <i>Sonstiges</i>	Mathematikbezogene Unterrichtsphasen, die nicht den anderen Unterrichtsphasen zugeordnet werden können, wie allgemein organisatorische Hinweise oder Hausaufgabenkontrolle.

* Die Phase „6 Präsentieren, erheben und beurteilen der Ergebnisse/des eigenen Könnens“ gilt u.a. als Grundlage für die Codierung von Peer- und Self-Assessment.

Tabelle 6: Kategoriensystem Anleitung und Aktivität von Peer-Assessment (PA)

Facette	Kategorien (Zusammenfassung)
<p><i>Anleitung:</i> Die Facette „PA-Anleitung“ umfasst sämtliche Abschnitte des Unterrichts, in denen die Lehrperson den Lernenden Anweisungen zu den in der Folge zu bearbeitenden PA-Aktivitäten erteilt. Dazu zählen u.a. Erläuterungen zum Material, Hinweise zum Ablauf, zur Gruppenzusammensetzung und zum Arbeitsort.</p>	<p><i>Beschreibung und Bewertung:</i> Die Lehrperson fordert alle oder die Mehrheit der Schülerinnen und Schüler dazu auf, sich Produkte vorangegangener Arbeitsphasen zu präsentieren/erklären/besprechen und explizit zu bewerten.</p> <hr/> <p><i>Bewertung:</i> Die Lehrperson fordert alle oder die Mehrheit der Schülerinnen und Schüler explizit dazu auf, Produkte vorangegangener Arbeitsphasen zu überprüfen und zu bewerten.</p> <hr/> <p><i>Beschreibung:</i> Die Lehrperson fordert alle oder die Mehrheit der Schülerinnen und Schüler dazu auf, sich Produkte vorangegangener Arbeitsphasen zu präsentieren/erklären/besprechen.</p>
<p><i>Aktivität:</i> Die Facette „PA-Aktivität“ umfasst sämtliche Abschnitte des Unterrichts, in denen alle oder die Mehrheit der Schülerinnen und Schüler die Möglichkeit (das Potenzial) hat, sich mit Produkten (Rechenwege, Lösungen, Fehler) anderer zu befassen. Zur PA-Aktivität gehören sowohl Lehrpersonen- als auch Schülerinnen- und Schüleräußerungen und -aktivitäten.</p>	<p><i>Moderierte Aktivität:</i> Alle oder die Mehrheit der Klasse macht ein PA und die Lehrperson übernimmt die Steuerung der PA-Aktivitäten. Die Lehrperson moderiert im Klassenunterricht den Austausch und die Beurteilung anderer Produkte (Lösungen, Rechenwege, Fehler). Sie ruft einzelne Beiträge der Schülerinnen und Schüler auf, hört zu, fragt nach, hat erklärende Anteile, stellt Vergleiche an usw.</p> <hr/> <p><i>Eigenständige Aktivität:</i> Alle oder die Mehrheit der Klasse macht ein PA. Die Steuerung der PA-Aktivitäten liegt mehrheitlich bei den Schülerinnen und Schülern. Die Lehrperson nimmt eine beobachtende Rolle ein und unterstützt mit kurzen Beiträgen einzelne Gruppen. Die Verantwortung des PA liegt bei den Schülerinnen und Schülern.</p> <hr/> <p><i>Rest:</i> Es ist ersichtlich, dass ein Teil der Schülerinnen und Schüler Peer-Assessments macht, aber nicht erkennbar, was genau sie tun.</p>

Tabelle 7: Übersicht Bereiche und Dimensionen des Classroom Assessment Scoring System CLASS

Bereiche	Dimensionen
<p><i>Emotionale Unterstützung:</i> Sensitivität der Lehrperson gegenüber Anliegen der Schülerinnen und Schüler.</p>	<p>Positives Klima: Freude und emotionale Verbundenheit der Lehrperson mit den Schülerinnen und Schülern sowie die Art der Interaktionen im Klassenzimmer.</p> <p>Sensitivität der Lehrpersonen: Berücksichtigung der akademischen und sozial-emotionalen Bedürfnisse und Niveaus der einzelnen Schülerinnen und Schüler durch die Lehrperson.</p> <p>Berücksichtigung der Schülerperspektiven: Ausmass, in dem die Lehrperson Mitentscheidungen und Autonomie der Schülerinnen und Schüler zulässt.</p> <p>Wertschätzung der Meinung der Schülerinnen und Schüler und sinnvolle Interaktionen mit Gleichaltrigen.</p> <p>Schulisches Engagement: Aktive Teilnahme der Schülerinnen und Schüler an den Lerngelegenheiten.</p>
<p><i>Klassenmanagement:</i> Effektiver Umgang mit Unterrichtsstörungen und produktive Nutzung von Unterrichtszeit und Unterrichtsaktivitäten.</p>	<p>Verhaltensmanagement: Wie gut Lehrpersonen positives Verhalten fördern und Fehlverhalten überwachen, verhindern und umlenken.</p> <p>Produktivität: Wie routiniert der Unterricht abläuft und inwieweit die Lehrpersonen Aktivitäten und Anweisungen geben, damit möglichst viel Zeit für Lernaktivitäten zur Verfügung steht.</p> <p>Negatives Klima: Ausmass geäussertter Negativität wie Wut, Feindseligkeit, Aggression oder Respektlosigkeit, die von Lehrpersonen und/oder Schülerinnen und Schülern im Klassenzimmer gezeigt wird.</p>
<p><i>Lernunterstützung:</i> Qualität von Instruktionen und Feedback und Grad der kognitiven Stimulierung der jeweiligen Aktivitäten und Interaktionen.</p>	<p>Lehr- und Lernformate: Ausmass der Einbindung der Schülerinnen und Schüler in lernförderliche Aktivitäten.</p> <p>Inhaltsverständnis: Inwieweit die Lehrpersonen den Schülerinnen und Schülern helfen, sowohl den allgemeinen Rahmen als auch die Schlüsselideen in einem akademischen Fach zu verstehen.</p> <p>Analyse und Erkundung: Wie fördern die Lehrkräfte die Denkprozesse der Schülerinnen und Schüler (z. B. Analyse und Integration von Informationen, Hypothesentests, Metakognition) und bieten Möglichkeiten zur Anwendung in neuen Kontexten?</p> <p>Qualität des Feedbacks: Wie die Lehrpersonen das Lernen der Schülerinnen und Schüler durch ihre Antworten und ihre Teilnahme an Aktivitäten erweitern und vertiefen.</p> <p>Unterrichtsgespräche: Wie die Lehrpersonen strukturierte kumulative Fragen und Diskussionen einsetzen, um das Verständnis der Schülerinnen und Schüler für den Inhalt zu fördern und zu unterstützen.</p>

Ein digitales Tool zur Unterstützung von formativem Peer-Feedback bei Lehramtsstudierenden

Inga Glogger-Frey¹ | Anika Bürgermeister² | Henrik Saalbach²

¹Universität Erfurt | ²Universität Leipzig

Abstract: In diesem Kapitel stellen wir ein digitales Tool vor, das Lehramtsstudierende unterstützt, sich gegenseitig formatives Peer-Feedback auf die Verwendung von Lernstrategien zu geben. Die Lernstrategien sollen von den Studierenden in regelmäßig zu verfassenden Lerntagebucheinträgen eingesetzt werden, wodurch sie für ein formatives Feedback zugänglich gemacht werden. Die Unterstützung des Peer-Feedbacks besteht aus zwei Teilen, einer Assessmentunterstützung und einer Feedbackunterstützung. Das Assessment der Lernstrategien unterstützt das digitale Tool durch Bewertungsrubriken (sogen. *Rubrics*). Die Formulierung des Feedbacks unterstützt das Tool durch Satzanfänge im Sinne des *Procedural Facilitation*. Die theoretische Fundierung für die Unterstützungsmaßnahmen und die Ergebnisse aus Bürgermeister et al. (2021) zu einer experimentellen Untersuchung des Tool-Einsatzes werden zusammenfassend dargestellt. Hauptziel des Kapitels ist es, konkrete Einblicke in das Tool und seine Verwendung im Rahmen eines Seminarkonzepts zu gewähren. Abschließend werden weiterführende Fragen, sowie Möglichkeiten und Grenzen in der Anwendbarkeit des Tools betrachtet.

1 Einführung

Man stelle sich folgendes Szenario vor: Studierende schreiben wöchentlich zwei Seiten Reflexionen zur Vertiefung von Vorlesungsinhalten und zur Förderung ihres selbst-regulierten Lernens. Viele Studierende sind verunsichert, ob das Geschriebene den Anforderungen entspricht und bitten den Dozierenden um Feedback. Ein elaboriertes formatives Feedback auf die schriftlichen Arbeiten zu geben ist für den Dozierenden nicht leistbar.

Hochschullehrende stehen häufig vor dieser Situation, wenn es um anspruchsvolle Lernaufgaben (z. B. Essays, Präsentationen) geht. Dieses Kapitel soll dazu beitragen, elaboriertes formatives Feedback in der Hochschullehre umsetzen zu können – auch in großen Studiengängen wie den Lehramtsstudiengängen. Wir stellen ein digitales Tool vor, das Studierende unterstützt, ihren Peers ausführliches, lernförderliches Feedback zu geben – hier bezüglich des Einsatzes von Lernstrategien beim Reflektieren in Lerntagebüchern.

Formatives Feedback kann – neben der Anpassung der Instruktion – eine Maßnahme in Folge von formativem Assessment sein. Formatives Assessment wird als prozessorientierte Diagnostik verstanden, die die Förderung des weiteren Lernens zum Ziel hat (Bürgermeister & Saalbach, 2019). Formatives Feedback kann verschiedene Informationen aus dem vorangehenden Assessmentprozess an den Lernenden kommunizieren (Narciss, 2006), beispielsweise ausschließlich, ob eine Aufgabenlösung richtig oder falsch ist, oder ausführlich, wo der Lernende in Bezug auf verschiedene Aspekte des Lernziels steht. Als besonders lernförderlich wird letzteres, nämlich „elaboriertes“ Feedback angesehen (Shute, 2007), das nach Hattie und Timperley (2007) folgende Aspekte einbezieht: (1) „Feed up“: Rückbezug auf Lernziele und damit auf den anvisierten Zielzustand (vgl. Ramaprasad, 1983), (2) „Feed back“: Rückmeldung über den aktuellen Lernstand bzw. Lernprozess des Lernenden, (3) „Feed forward“: Vorschlag zu nächsten Schritten beim Lernen, um die ggf. identifizierte Lücke zwischen aktuellem Stand und Lernziel zu schließen. Die Umsetzung eines solchen elaborierten Feedbacks ist jedoch ressourcenintensiv. Beispielsweise beinhaltet es eine individuelle, sprachlich ausformulierte Rückmeldung darüber, wo der/die Lernende in Bezug auf das Anwenden von Lernstrategien in einem Lerntagebuch steht und was er/sie in einem nächsten Schritt zur Verbesserung der Lernstrategien tun kann (vgl. Hattie & Timperley, 2007). Gerade in der Lehramtsausbildung mit ihren häufig sehr großen Studiengängen ist ein solches formatives Feedback für Dozierende nicht leistbar. *Peer-Feedback* bietet die Möglichkeit, ein solches Feedback dennoch zu realisieren. Indem die Lernenden sich gegenseitig (d. h. ihren Peers) Feedback geben, werden die Dozierenden entlastet. Gut angeleitet, kann Peer-Feedback zudem aus folgenden Gründen lernförderlich wirken: die Lernenden setzen sich potentiell intensiv mit den Lernzielen, Beurteilungskriterien sowie möglichen Lösungsstrategien einer Aufgabenbearbeitung auseinander; darüber hinaus können ihre selbstregulativen Kompetenzen gestärkt werden (Gielen et al., 2010). Indem formatives Peer-Feedback in der Lehramtsausbildung eingesetzt wird, können künftige Lehrkräfte außerdem in ihrem Studium Kompetenzen erwerben, die sie zur Umsetzung formativen Assessments und Feedbacks in ihrer späteren Lehrtätigkeit befähigen (Bransford & Darling-Hammond, 2005).

Im vorliegenden Beitrag stellen wir ein digitales Tool vor, das mehrere Prozesse des formativen Peer-Feedbacks unterstützt. Wir gehen insbesondere auf Feedback bzgl. Lernstrategien ein, halten jedoch andere Einsatzbereiche für denkbar (s. Abschnitt 6). Um formatives Feedback auf Ebene von Lernprozessen bzw. Strategien des selbstregulierten Lernens erstellen zu können, müssen zunächst Lernprozesse *sichtbar* gemacht werden, welche dann qualitativ *beurteilt* werden (*Assessment-Prozesse*); auf dieser Basis kann geeignetes lernförderliches Feedback *formuliert* werden (Glogger-Frey & Herppich, 2017). Entsprechend betrachten wir im Folgenden, mit welcher Methode wir Lernstrategien sichtbar und damit für das Feedback zugänglich machen konnten (Abschnitt 2). Abschnitt 3 beschreibt, wie das digitale Tool die Assessment-Prozesse und die Feedbackformulierungsprozesse unterstützen kann. Abschnitt 4 fasst die empirischen

Befunde aus Bürgermeister et al. (2021) zu Wirkungen der Unterstützungsmaßnahmen des digitalen Tools zusammen. Abschnitt 5 gibt Einblicke, wie die Assessment-Methode und formatives Peer-Feedback in ein Seminarkonzept an der Hochschule eingebettet werden kann. Abschließend werden in Abschnitt 6 nach einer Zusammenfassung weiterführende Fragen der Anwendbarkeit des Tools betrachtet.

2 Das Lerntagebuch als Methode, Lernstrategien für formatives Assessment und Feedback zugänglich zu machen

Die Methode *Lerntagebuch* kann als Instrument genutzt werden, Lernprozesse bzw. Lernstrategien bei Lernenden sichtbar zu machen (Glogger et al., 2012; Berthold et al., 2007; Nückles et al., 2020). In Lerntagebucheinträgen vertiefen Lernende die Lerninhalte schriftlich. Nückles und Kolleg:innen (2020) fassen Anregungen und Fragen, sogenannte *Lernstrategie-Prompts* zusammen, die sich in Studien des Freiburg Self-Regulated-Journal-Writing Approach als wirksam erwiesen haben. Solche Prompts sind wirksam, indem sie das Verwenden der entsprechenden Strategien anregen (z. B. Anregung von Elaboration: „Welche Beispiele fallen Ihnen ein, die das Gelernte illustrieren, bestätigen oder ihm widersprechen?“; Glogger et al., 2009; Nückles et al., 2009) und darüber vermittelt den Lernerfolg erhöhen (Berthold et al., 2007; Roelle et al., 2017). Studierende können nach einer Einführung und ggf. weiteren instruktionalen Maßnahmen (z. B. Beispielinträge im Sinne von Lösungsbeispielen; Hübner et al., 2010) mit solchen Prompts Lerntagebucheinträge als Fließtext verfassen. Sie vertiefen auf diese Weise unter Anwendung der Strategien *schreibend* ihr Verständnis des Lernstoffs.

Tabelle 1 zeigt Beispielinträge aus Lerntagebüchern und darin sichtbare, d. h. diagnostizierbare Lernstrategien. Grundsätzlich lassen sich drei Arten von Lernstrategien unterscheiden: die kognitiven Strategien Elaboration und Organisation, sowie metakognitive Strategien (Nückles et al., 2020; cf. Fiorella & Mayer, 2015). Elaborationsstrategien zeigen sich immer dann, wenn das neu Gelernte anhand von eigenem Vorwissen für die Lernenden bedeutungsvoller gemacht wurde. ‚Eigene Beispiele erdenken‘ oder ‚Hinterfragen auf Grundlage eigener Erfahrungen‘ erfordert jeweils das Heranziehen von Vorwissen und gehört demnach zur Elaboration. Organisationsstrategien zielen auf das sinngebende Strukturieren der Lerninhalte sowie auf deren Reduktion auf das Wesentliche. Beim Organisieren findet der/die Lernende Zusammenhänge innerhalb des neu Gelernten und trennt Wichtiges von Unwichtigerem. Metakognitive Strategien dienen dem Überwachen des eigenen Verständnisses und dem Steuern des kognitiven Lernprozesses. Beim Schreiben eines Lerntagebuchs können Lernende etwa von einem Verständnisproblem ausgehen und im Schreiben das Problem aufzulösen suchen (z. B. sich den Unterschied zweier Begriffe anhand eines Beispiels klarmachen).

Empirische Ergebnisse zeigen, dass die im Lerntagebuch messbaren Lernstrategien mit inhaltlichem Verständnis des Gelernten zusammenhängen. Glogger et al. (2012) ließen Neuntklässler der Realschule über mehrere Wochen hinweg als Hausaufgabe

Lerntagebuch schreiben. In Mathematik und in Biologie fanden sich substantielle Zusammenhänge zwischen kodierten Lernstrategien und den Ergebnissen von Verständnistests. Diese Zusammenhänge sind höher als solche, die beim Einsatz üblicher Lernstrategie-Fragebögen zu finden sind (z. B. Artelt, 2000).

In mehreren Studien wurde die Intensität der Lernstrategienutzung der Lernenden durch experimentelle Variation der Art der Lernstrategieprompts erhöht. Gruppen, die metakognitive und kognitive Lernstrategien (Organisation, Elaboration) im Lerntagebuch stärker verwendeten, zeigten auch höhere Lernerfolge (Berthold et al., 2007; Glogger et al., 2012; Nückles et al., 2009). Es ist also davon auszugehen, dass die im Lerntagebuch diagnostizierbaren Strategien tatsächlich von Lernenden eingesetzt werden und sich verständnisförderlich auswirken. Das heißt, Lerntagebucheinträge von Studierenden können als Grundlage für ein formatives Feedback zu Lernstrategien genutzt werden. Mit entsprechender Unterstützung können Studierende sich auch gegenseitig solches Feedback, also Peer-Feedback, geben.

Tabelle 1:

	Lerntagebuchausschnitt	diagnostizierbare Lernstrategie
1	<p>„Was mir bei Piaget allerdings fehlt, ist die Beachtung interindividueller Unterschiede in der Entwicklung. Ich finde, dass Heranreifen auch immer durch äußere Einflüsse wie der Erziehung bestimmt ist.</p> <p>Ein Kind eines alkoholabhängigen Elternteils wird sicherlich nicht so gefördert wie das von ehrgeizigen Elternteilen, die nicht mit persönlichen Problemen belastet ist und viel Energie für die Förderung des Kindes aufwenden.</p> <p>Während meines Freiwilligendienstes an einer Grundschule mit sozial benachteiligten Kindern gab es ein Mädchen mit einer alleinerziehenden Mutter mit Drogenproblemen. Ihr Entwicklungsstand war weit hinter dem ihrer Mitschüler, d. h. sie konnte z. B. – mit 11 Jahren! - nicht fließend lesen. So kann Entwicklung also – entgegen der Theorie von Piaget individuell unterschiedlich ablaufen (wobei das natürlich nur ein Einzelbeispiel ist).“</p>	(ausführliche) Elaborationsstrategie: Kritische Beleuchtung unter Rückgriff auf Vorwissen (u.a. eigene Erfahrungen)
2	<p>„In der letzten Seminarstunde waren für mich viele Begriffe und Erläuterungen neu und ergaben noch keinen Zusammenhang und wenig Sinn, so zum Beispiel der Strukturalismus und Piaget als Vertreter.“</p>	Metakognition: Überwachen des eigenen Verständnisses
3	<p>„Mit ein bisschen Recherche habe ich allerdings erkannt, dass die Annahme von Schemata, die sich immer weiter differenzieren, und letztlich je nach Stadium unterschiedliche Denk-Strukturen bilden, ein Zeichen für den Strukturalismus von Piagets Theorie ist. Beim Übergang von einem zum nächsten Entwicklungsstadium verändern sich die Denkstrukturen des Kindes.“</p>	Metakognitive Strategie: Regulation des eigenen Lernens Organisationsstrategie: Zusammenfassung des (neu) Verstandenen

3 Tool zur Unterstützung von formativem Peer-Feedback zu Lernstrategien (FormFeed-Tool)

Lehramtsstudierende sollten angeleitet und unterstützt werden, wenn sie Peer-Feedback geben. Feedback von einem Peer wird als hilfreich empfunden, wenn es spezifische Informationen dazu enthält wie man Fehler korrigiert und weiterlernt (Strijbos & Sluijsmans, 2010) und wenn es vollständig verstanden wird (Kollar & Fischer, 2010). Qualitativ hochwertiges, formatives Feedback zu geben, ist eine anspruchsvolle Aufgabe, und die Forschung zeigt, dass Studierende diesen Prozess als eher unangenehm und schwierig empfinden (Hanrahan & Isaacs, 2001; Kaufman & Schunn, 2011). Nückles et al. (2005) fanden, dass Peers ohne Unterstützung dazu neigten, ihr Feedback im Laufe der Zeit aneinander anzugleichen. Das heißt, sie tendierten dazu, in Länge und Stil immer ähnlicheres Feedback zu geben. So reduzierten sie ihr Feedback auf bestimmte Aspekte und vernachlässigten wichtige andere Aspekte, oder entwickelten sehr knappes Feedback. Daher ist es wichtig, Lehramtsstudierende zu unterstützen, wenn sie die Arbeit eines Peers bewerten und angemessenes, klares Feedback geben sollen (Hanrahan & Isaacs, 2001; Walker, 2015).

Unterstützung sollte dabei an den zwei zentralen Teilprozessen von formativem Peer-Feedback ansetzen: (a) am Diagnostizieren der Peer-Leistung, zu der es Feedback zu geben gilt; und (b) am Formulieren des Feedbacks: auf Grundlage der gesammelten diagnostischen Informationen wird ein Feedback an den Peer in einem zweiten Teilprozess in Worte gefasst. Die theoretische Unterscheidung dieser Prozesse findet sich beispielsweise in dem Modell von Herppich et al. (2018) in Form einer Unterscheidung von „assessment competence“ und „instructional competence“ (vgl. auch van Ophuysen, 2010). Das Diagnostizieren setzt Kenntnisse über den zu diagnostizierenden Gegenstand voraus, wie etwa Kenntnisse über mögliche Lernstrategien und Kriterien zu ihrer Qualitätsbeurteilung. Für die Formulierung des Feedbacks müssen die gesammelten Informationen genutzt werden. Die Feedbackformulierung setzt zudem Kenntnisse darüber voraus, welche Informationen lernförderliches Feedback enthalten sollte und wie eine strukturierte, verständliche Formulierung gefunden werden kann. Daher enthält das hier vorgestellte Peer-Feedback-Tool „FormFeed“ eine Unterstützungsmaßnahme für das Diagnostizieren (Assessment) und eine Unterstützungsmaßnahme für das Formulieren von Feedback.

Die *Assessmentunterstützung* erfolgt mittels „Rubrics“ (Reddy & Andrade, 2010), wodurch Lehramtsstudierende angeleitet werden, die Umsetzung der angewandten Lernstrategien theoriegeleitet einzuschätzen (siehe Abbildung 1). Solche Bewertungsrubriken informieren über die Erwartungen an eine Aufgabe bzw. in unserem Fall an die Umsetzung von Lernstrategien beim Lerntagebuchschreiben, indem sie die Bewertungskriterien formulieren und die Qualitätsstufen in Bezug auf jedes dieser Kriterien auflisten (Panadero & Jonsson, 2013). In unserem Tool sind die Rubriken folgendermaßen umgesetzt (siehe Abbildung 1): Aussagen über das Lerntagebuch werden auf einer Skala von 1 (*trifft überhaupt nicht zu*) bis 6 (*trifft voll und ganz zu*) eingeschätzt. Eine

solche Aussage, wie bspw. „Es wird ein Beispiel erläutert, in nachvollziehbarer Weise, und eng bezogen auf das Thema“ benennt eine mögliche Art der Lernstrategie (hier: Elaboration) und formuliert ein Qualitätskriterium (hier: nachvollziehbare Weise, enge Bezugnahme auf das Thema der Literatur). Studierende schätzen Aussagen zu jeder Strategieart ein (in Abbildung 1 sind Aussagen zur Elaboration angezeigt; wie durch den Balken ganz oben in der Abbildung visualisiert, führt das Tool den Studierenden von Organisations-, zu Elaborations- und schließlich zu Metakognitionsstrategien.)

Organisation	Elaboration I			Elaboration II			Metakognition
Elaboration I	Trifft überhaupt nicht zu ←			Trifft voll und ganz zu →			
Es wird, in nachvollziehbarer Weise, eine Verknüpfung zu eigenen Erfahrungen (z.B. aus Schulzeit, Praktika, Studium) hergestellt.	●	●	●	●	●	●	
Es wird, in nachvollziehbarer Weise, ein Bezug zum Vorwissen (bisher erworbenes Wissen aus Schulzeit, Studium oder anderen Kontexten) hergestellt.	●	●	●	●	●	●	
Es wird ein Beispiel erläutert, in nachvollziehbarer Weise, und eng bezogen auf das Thema.	●	●	●	●	●	●	

Abbildung 1: Auszug aus dem FormFeed-Tool: Rubriken unterstützen das Assessment von Elaborationsstrategien

Tatsächlich konnte bereits gezeigt werden, dass sich Rubrics vorteilhaft auf Peer-Assessment (Hafner & Hafner, 2003; Saddler & Andrade, 2004) und Lernen (z.B. Reddy & Andrade, 2010) auswirken können. Dabei sind Randbedingungen wie Kompetenz der Feedbackgebenden und Domänenunterschiede zu beachten (z.B. Strijbos et al., 2010). Das kriterienbasierte Einschätzen von Peer-Leistungen kann die *feedbackgebenden* Peers die zugrundeliegenden Lernziele und die relevanten Bewertungskriterien verinnerlichen lassen. Außerdem erleichtern Rubriken eine *genaue* und *effiziente* Beurteilung der Arbeit der Peers (Reynolds-Keefer, 2010), was eine solide Grundlage für angemessenes und genaues formatives Feedback bieten kann. Sadler und Good (2006) zeigten in diesem Zusammenhang eine hohe Korrelation zwischen Schüler- und Lehrerbeurteilung ($r = 0.91-0.94$), nachdem sie Schülerinnen und Schülern der Mittelstufe in der Verwendung von Bewertungsrubriken (für einen spezifischen Test) geschult hatten.

Die *Feedbackunterstützung* im hier vorgestellten Tool erfolgt durch Textbausteine im Sinne eines „*procedural facilitation*“ (Scardamalia et al., 1984). *Procedural facilitation* bezieht sich auf instruktionale Unterstützung, die Prozesse, wie etwa das Schreiben, erleichtern. So wird ein Schreibprozess unterstützt, indem eine quasi unendliche Zahl an Möglichkeiten des Formulierens auf eine begrenzte Zahl reduziert wird. Zudem werden

Prozesse strukturiert und Gedächtnisstützen gegeben (Baker et al., 2002). Scardamalia et al. (1984) nutzten eine Auswahl an möglichen Satzanfängen, um Lernende darin zu unterstützen, einen Aufsatz zu entwickeln und zu schreiben. Die Satzanfänge bezogen sich etwa auf die Ziele des Aufsatzes (z. B. „A goal I think I could write to ...“) oder das Vertiefen einer Idee (z. B. „I could develop this idea by adding ...“, „An example of this ...“).

Angelehnt an diese Arbeit werden im hier vorgestellten Tool Satzanfänge zum unterstützten Schreiben von Feedback genutzt. Die Satzanfänge werden strukturiert in drei Abschnitten angezeigt (siehe Abbildung 2): (1) *Ziel der Lernstrategie*, kurz: Ziel (2) *Qualität der Umsetzung der Strategie*, kurz: Stand – Stärken, Schwächen (3) *Strategieempfehlungen für den weiteren Lernprozess*, kurz: Strategien. Diese Abschnitte orientieren sich an den drei Fragen, die lernförderliches Feedback nach Hattie und Timperley (2007) beantworten sollte und die sich jeweils auf spezifische Aspekte von Feedback beziehen: (1) Wohin gehe ich? („Feed up“: Rückbezug auf die Lernziele), (2) Wie komme ich voran? („Feed back“: Rückmeldung über den aktuellen Lernstand bzw. Lernprozess des Lernenden), (3) Wohin als nächstes? („Feed forward“: Rückmeldung, die sich auf das weitere Vorgehen bzw. die Strategien des Lernenden beim weiteren Lernen bezieht).

Satzanfänge

1 Ziel

- Das Ziel ist es, ...
- Um diese Lernstrategie gut umzusetzen, sollte(n) ...
- Das Verfassen des Lerntagebuchs zielte darauf ab ...

2 Stand

2.1 Stärken

- Du hast versucht, ...
- Bereits gelungen ist dir ...
- Es ist gut erkennbar, dass ...
- Du hast diese Strategie teils umgesetzt, indem du ...

2.2 Schwächen

- Noch besser wäre es, wenn du ...
- Es wäre noch wichtig, dass ...
- Du hast dich sehr bemüht, aber ...
- Nicht ganz nachvollziehbar ist ...
- Es ist weniger gut erkennbar, dass ...

3 Strategien

- Vielleicht hilft es dir, ...
- Eine Möglichkeit wäre ...
- Alternativ könntest du in Zukunft ...
- Versuche beim nächsten Mal ...
- Achte in Zukunft darauf, dass ...
- Das zeigt, dass du beim nächsten Mal, ...
- Eine bessere Lernstrategie in diesem Zusammenhang könnte sein ...
- Für dich förderlich könnte sein ...
- Diese Strategie wird dann lernförderlich, wenn du ...

Abbildung 2: Auszug aus dem FormFeed-Tool: Satzanfänge

Inhaltlich sind die Satzanfänge meist generisch (z. B. Ziel: „Das Ziel ist es, ...“; Stärken: „Bereits gut gelungen ist dir ...“), beziehen sich teils aber ausdrücklich auf Lernstrategien (z. B. Ziel: „Um diese Lernstrategie gut umzusetzen, sollte(n) ...“; Strategien: „Diese Lernstrategie wird dann lernförderlich, wenn du ...“). Die Studierenden können per Klick Satzanfänge auswählen und in eine Textbox übertragen lassen. Der Satzanfang kann dann bei Bedarf editiert und der Satz vervollständigt werden. Die Studierenden verfassen das Feedback getrennt für jede Strategieart (Organisation, Elaboration, Metakognition). Je nachdem, wie gut die Strategieart im Assessmentschritt mittels der Rubriken bewertet wurde (d. h. gut, mittel, oder schlecht), werden leicht angepasste Satzanfänge angezeigt. Abbildung 2 zeigt Satzanfänge für eine mittelmäßige Umsetzung der Strategien. Zum Abschluss des Feedbackprozesses wird ein Gesamtdokument (PDF) mit dem Feedback für alle Lernstrategiearten ausgegeben (per Betätigung eines Buttons).

Eine Meta-Analyse von Wisniewski et al. (2020) zeigt, dass Lernende besonders von Feedback profitieren, wenn es Informationen zur Aufgaben-, Prozess- und Selbstregulationsebene sowie zum weiteren Lernen enthält. Der Fokus auf Lernstrategien im FormFeed-Tool setzt insbesondere auf der eher selten fokussierten Ebene der Selbstregulation an.

Die Meta-Analyse von Wisniewski et al. (2020) zeigt zudem, dass Feedback umso effektiver ist, je mehr Informationen es enthält. Textboxen zu jeder geprompteten Lernstrategie (Elaboration, Organisation, Metakognition) sowie die Satzanfänge, die nach den drei Fragen lernförderlichen Feedbacks gegliedert sind, sollten reichhaltige Informationen im Feedback der Studierenden sicherstellen.

4 Zusammenfassung von Ergebnissen zum FormFeed-Tool

In einer ersten Studie der Autor:innen (Bürgermeister et al., 2021) mit einer Stichprobe von $N = 129$ Lehramtsstudierenden wurden Effekte des FormFeed-Tools auf die Selbstwirksamkeit und auf die Qualität des Feedbacks überprüft. Es wurde zudem der Effekt der Feedbackunterstützung sowie der Assessmentunterstützung einzeln und in Kombination geprüft. Studierende nutzten das Tool je nach experimenteller Gruppe mit (1) Feedbackunterstützung, (2) Assessmentunterstützung, (3) mit beiden Unterstützungen, oder (4) keiner der Unterstützungen (d. h. nur leere Textfelder und Arbeitsauftrag im Tool), über ein Semester hinweg zu drei Zeitpunkten. Es wurden Fragebogendaten erhoben sowie Feedbacks der Studierenden kodiert¹. Bürgermeister et al. (2021) fanden dabei, dass die Studierenden ihre *Selbstwirksamkeit* in Bezug auf die Bewertung und das Geben von Feedback zu Lernstrategien in Lerntagebüchern als höher einschätzten,

¹ Weitere Daten, nämlich Lerntagebücher und Abschlusstestaufgaben, wurden ausgewertet, jedoch zum Zeitpunkt dieser Publikation noch nicht veröffentlicht (s. Zeeb et al., in Begutachtung).

nachdem sie wiederholt Peer-Feedback, unterstützt durch das FormFeed-Tool, erhalten haben. Die Ergebnisse zeigen, dass die Studierenden von der kontinuierlichen Arbeit mit dem digitalen Tool über das Semester hinweg sowie vom Erstellen und Erhalten von Feedback profitierten – unabhängig davon, welche Unterstützung (Feedbackunterstützung, Assessmentunterstützung, beide oder keine) sie erhielten. Dies steht im Einklang mit theoretischen Arbeiten, die davon ausgehen, dass Selbstwirksamkeit durch Übung gefördert werden kann (Schunk, 1995). Darüber hinaus zeigten unsere empirischen Ergebnisse in Bürgermeister et al. (2021), dass die Lehramtsstudierenden *unmittelbar* vom Procedural Facilitation profitieren – also der Unterstützung, um das Peer-Feedback zielgerichtet zu formulieren (1. Messzeitpunkt). Über das Semester hinweg zeigte sich, dass die Kombination aus Unterstützung des Assessmentprozesses (Rubrics) und der Unterstützung der Formulierung des Feedbacks durch Satzanfänge besonders vorteilhaft für die Selbstwirksamkeit der Lernenden war (2. Messzeitpunkt). D.h. nach mehrmaliger – nicht aber nach einmaliger – Anwendung der Rubriken trauten sich die Studierenden eher zu, die Lernstrategien ihrer Peers angemessen zu bewerten und die diagnostischen Informationen beim Schreiben des Feedbacks zu nutzen. Da Rubriken viele Informationen enthalten, zeigt dieses Ergebnis möglicherweise, dass die Studierenden zunächst etwas Übung mit der Arbeit mit den Rubriken als Bewertungsinstrument benötigen. Mit etwas Übung und Zeit können möglicherweise bessere Lernergebnisse erzielt werden (Andrade, 2005; Luft, 1999).

Wir fanden zudem einen positiven Effekt der Feedbackunterstützung auf die *Gesamtqualität des Feedbacks*, d.h. die Lehramtsstudierenden sind in der Lage, mehr nützliches Feedback zu schreiben, wenn sie Procedural Facilitation erhielten; d. h. wenn sie über spezifische Formulierungsmöglichkeiten eine Struktur und Gedächtnisstützen bekamen, was wirksames Feedback enthalten sollte. Die Studierenden konzentrierten sich mehr auf das *zugrundeliegende Lernziel* sowie auf *Stärken der Peers* als die Lehramtsstudierenden, die keine Hilfestellung erhielten.

Für die Hochschulpraxis impliziert dieses Ergebnis, dass das Peer-Feedback-Tool möglichst wiederholt eingesetzt (wie im oben beschriebenen Seminarkonzept) bzw. die Nutzung der Rubriken eingeübt werden sollte. Ist nur Zeit, um ein solches Tool ein einziges Mal einzusetzen, ist eher zur Nutzung von Formulierungshilfen wie den Satzanfängen zu raten.

5 Einbindung des FormFeed-Tools in ein Seminarkonzept

Die folgende Beschreibung gibt Einblick in eine mögliche Einbindung des FormFeed-Tools in ein Seminarkonzept in der universitären Lehre. Um die Handhabbarkeit sowie die Wirkungen des FormFeed-Tools auf die Studierenden zu prüfen, wurde es in ein Seminarkonzept der Entwicklungspsychologie im Rahmen der Lehramtsausbildung der Universität Leipzig implementiert und wie oben zusammenfassend dargestellt eva-

luiert. Das Seminarkonzept setzte sich aus drei zentralen Teilen zusammen: den Lerntagebüchern, dem Peer-Feedback und einem Workshop. Die Lehramtsstudierenden verfassten *wöchentlich Lerntagebucheinträge*. Die Einträge dienten der Vertiefung des Verständnisses von pädagogisch-psychologischen Grundlagentexten, die die Studierenden zu den verschiedenen Seminarinhalten bearbeiteten. Die Studierenden reichten ihren Lerntagebucheintrag im Laufe des Semesters kontinuierlich jeweils vor den entsprechenden Seminarsitzungen über eine Lernplattform ein. Dies ermöglichte, dass die Studierenden das Seminar mit einem grundlegenden Verständnis über das Thema besuchten und sie dieses durch die weitere Bearbeitung des entsprechenden Inhaltes im Rahmen eines Workshops vertiefen konnten. Zudem wurden die Studierenden angehalten, ihren Lerntagebucheintrag insgesamt, insbesondere aber ihre im Tagebuch formulierten Fragen vor dem Hintergrund der Seminarsitzung zu reflektieren bzw. zu beantworten. Dies sollte die Studierenden darin unterstützen, die Inhalte der Sitzung mit den von ihnen im Lerntagebuch organisierten, elaborierten und metakognitiv reflektierten Inhalten der Grundlagenliteratur zu verbinden bzw. zu integrieren und somit das Lernergebnis zu sichern.

Um die Qualität der im Lerntagebuch eingesetzten Lernstrategien zu steigern, erstellten die Studierenden zu drei Zeitpunkten im Semester zudem ein *Peer-Feedback* mit Hilfe des FormFeed-Tools. Der erste Zeitpunkt wurde so gewählt, dass die Studierenden das Schreiben des Lerntagebuchs anhand der gegebenen Prompts mindestens einmal ausprobieren konnten. Der dritte Zeitpunkt lag am Ende des Semesters, sollte aber selbstverständlich noch ein Umsetzen des erhaltenen Feedbacks in einem folgenden Lerntagebucheintrag ermöglichen. Die Zuteilung zu einem Peer erfolgte dabei automatisiert durch die Lernplattform und der Feedbackprozess war anonymisiert.

Die Studierenden gestalteten zudem einen *Workshop* (d. h. eine auf 45 Minuten begrenzte Lernsequenz innerhalb einer Seminarsitzung), in dem sie eines der Seminarthemen mit Übungen weiter vertieften. Sie waren dabei in Kleingruppen verantwortlich für die Planung, Durchführung und Reflexion dieses Workshops und erhielten zur Vorbereitung vertiefende empirische und theoretische Literatur sowie Lernziele und Impulse zur Umsetzung. Ziel des Workshops war es, durch den Einsatz verschiedener Methoden an die im Lerntagebuch bearbeiteten Inhalte anzuknüpfen (Vorwissensaktivierung), vertiefende und weiterführende Inhalte zu vermitteln und vor allem durch verschiedene Transferübungen den Erwerb von Handlungskompetenzen zu ermöglichen.

6 Zusammenfassung und Ausblick

Der vorliegende Beitrag beschreibt ein Tool, das Studierende dabei unterstützt, lernförderliches Feedback zum Einsatz von Lernstrategien beim Schreiben von Lerntagebüchern zu geben. Es wurde im Rahmen von universitären psychologischen Lehrveranstaltungen für Lehramtsstudierende evaluiert (s. Bürgermeister et al., 2021). Die Ergebnisse dieser Evaluationsstudie zeigen positive Effekte des Tool-Einsatzes (1) auf

die Selbstwirksamkeit der Studierenden im Hinblick auf das Verfassen von Feedback und (2) auf die Qualität des Feedbacks. Die Struktur und Inhalte des Tools basieren auf evidenzbasierten Modellen des Feedbacks und der Forschung zu Lerntagebüchern. Die positiven Ergebnisse der Implementation und die gute Anwendbarkeit des Tools empfehlen es für eine breitere Anwendung in der Hochschullehre. Damit kann auf das Bedürfnis von Studierenden nach individuellen bzw. personalisiertem Feedback eingegangen und das Potential des formativen Feedbacks als eine der lernwirksamsten Lehrmethoden (vgl. Hattie, 2009) Rechnung getragen werden.

Für die Anwendbarkeit des Tools in anderen Bereichen sehen wir im Folgenden diskutierte Möglichkeiten und Grenzen. Die Anwendung des Tools bzw. der Unterstützungsmaßnahmen beschränkt sich derzeit auf das formative Feedback zur Verwendung von Lernstrategien. Neben dem beschriebenen, evaluierten Einsatz wurde das Tool auch im ersten Semester des Bachelorstudiengangs *Psychologie* der Universität Freiburg (Seminar Entwicklungspsychologie) sowie einem Seminar zum selbstregulierten Lernen im psychologischen Bachelorstudiengang der Universität Erfurt, jeweils in Verbindung mit der Methode Lerntagebuch, eingesetzt. Es ist auf Grundlage der Befunde der Lerntagebuchforschung davon auszugehen, dass der effektive Einsatz von Lernstrategien und damit die Stärkung selbstregulierten Lernens das fachliche Lernen unterstützte. Direktes Feedback auf fachliche Aspekte (z. B. korrektes Verständnis einer Theorie) wurde jedoch nicht gefordert.

Für eine Übertragung auf das fachliche Lernen bedarf es deutlichen Anpassungen und Erweiterungen. Um eine lernförderliche Rückmeldung etwa auf konzeptuelle Fehlvorstellungen, beispielsweise im Bereich Entwicklungspsychologie zu geben, braucht es umfangreiches fachliches und fachdidaktisches Wissen. In umschriebenen Bereichen mit klar formulierten Lernzielen und Bewertungskriterien ist der Einsatz des Tools jedoch denkbar. Dies könnte ein spezifischer Essay sein, eine Präsentation zum Zwischenstand einer Projektarbeit oder Abschlussarbeit, oder auch eine von Studierenden für Studierende gehaltene Seminarstunde bzw. Workshop. Um eine Anwendung des Tools in der Hochschullehre in (solchen) anderen Bereichen zu ermöglichen, müssten Rubriken entsprechend der jeweiligen Lernziele formuliert und die nicht-generischen Satzanfänge angepasst werden.

Mit der Vorstellung des FormFeed-Tools in diesem Kapitel soll ein Beitrag geleistet werden, formatives Peer-Feedback auf hohem Niveau in der Hochschullehre umsetzen zu können. Diese Art des digital unterstützten Peer-Feedbacks eröffnet eine ökonomische und wirksame Möglichkeit, formatives Feedback in der Lehre anzubieten.

Literatur

Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27–31. <https://doi.org/10.3200/CTCH.53.1.27-31>

- Artelt, C. (2000). Wie prädiktiv sind retrospektive Selbstberichte über den Gebrauch von Lernstrategien für strategisches Lernen? *Zeitschrift für Pädagogische Psychologie*, 14(2), 72–84. <https://doi.org/10.1024//1010-0652.14.23.72>
- Baker, S., Gersten, R. & Scanlon, D. (2002). Procedural facilitators and cognitive strategies: Tools for unraveling the mysteries of comprehension and the writing process, and for providing meaningful access to the general curriculum. *Learning Disabilities Research & Practice*, 17(1), 65–77. <https://doi.org/10.1111/1540-5826.00032>
- Berthold, K., Nückles, M. & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, 17(5), 564–577. <https://doi.org/10.1016/j.learninstruc.2007.09.007>
- Bürgermeister, A., Glogger-Frey, I. & Saalbach, H. (2021). Supporting peer feedback on learning strategies: Effects on self-efficacy and feedback quality. *Psychology Learning & Teaching*, 20(3), 383–404. <https://doi.org/10.1177/14757257211016604>
- Bürgermeister, A. & Saalbach, H. (2019). Formatives Assessment. Ein Ansatz zur Förderung individueller Lernprozesse. *Psychologie in Erziehung und Unterricht*, 65(3), 194–205. <http://dx.doi.org/10.2378/peu2018.art11d>
- Bransford, J.D. & Darling-Hammond, L. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. Jossey-Bass.
- Fiorella, L. & Mayer, R.E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28, 717–741. <https://doi.org/10.1007/s10648-015-9348-9>
- Gielen, S., Peeters, E., Dochy, F., Onghena, P. & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304–315. <https://doi.org/10.1016/j.learninstruc.2009.08.007>
- Glogger, I., Holzäpfel, L., Schwonke, R., Nückles, M. & Renkl, A. (2009). Activation of learning strategies in writing learning journals: the specificity of prompts matters. *Zeitschrift für Pädagogische Psychologie*, 23(2), 95–104. <https://doi.org/10.1024/1010-0652.23.2.95>
- Glogger, I., Schwonke, R., Holzäpfel, L., Nückles, M. & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology*, 104(2), 452–468. <https://doi.org/10.1037/a0026683>
- Glogger-Frey, I. & Herppich, S. (2017). Formative Diagnostik und ihre Anforderungen an diagnostische Kompetenz. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (S. 13–18). Waxmann.
- Hafner, J. & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509–1528. <https://doi.org/10.1080/0950069022000038268>
- Hanrahan, S. J. & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research and Development*, 20(1), 53–70. <https://doi.org/10.1080/07294360123776>
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9780203887332>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.1007/s11159-011-9198-8>
- Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J. & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Journal of Teaching and Teacher Education*, 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>

- Hübner, S., Nückles, M. & Renkl, A. (2010). Writing learning journals: instructional support to overcome learning-strategy deficits. *Learning and Instruction*, 20(1), 18–29. <https://doi.org/10.1016/j.learninstruc.2008.12.001>
- Kaufman, J. H. & Schunn C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, 39(3), 387–406. <https://doi.org/10.1007/s11251-010-9133-6>
- Kollar, I. & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, 20(4), 344–348. <https://doi.org/10.1016/j.learninstruc.2009.08.005>
- Luft, J. A. (1999). Rubrics: Design and use in science teacher education. *Journal of Science Teacher Education*, 10(2), 107–121. <https://doi.org/10.1023/A:1009471931127>
- Narciss, S. (2006). *Informatives tutorielles Feedback. Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse*. Waxmann.
- Nückles, M., Hübner, S. & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction*, 19(3), 259–271. <https://doi.org/10.1016/j.learninstruc.2008.05.002>
- Nückles, M., Renkl, A. & Fries, S. (2005). Wechselseitiges Kommentieren und Bewerten von Lernprotokollen in einem Blended Learning Arrangement. *Unterrichtswissenschaft*, 33(3), 227–243. <https://doi.org/10.25656/01:5795>
- Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J. & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*, 32, 1089–1126. <https://doi.org/10.1007/s10648-020-09541-1>
- Panadero, E. & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Ramaprasad, A. (1983). On the definition of feedback. *Systems Research and Behavioral Science*, 1983(1), 4–13. <https://doi.org/10.1002/bs.3830280103>
- Reddy, Y. M. & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448. <https://doi.org/10.1080/02602930902862859>
- Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research, and Evaluation*, 15(1), 1–9. <https://doi.org/10.7275/psk5-mf68>
- Roelle, J., Nowitzki, C. & Berthold, K. (2017). Do cognitive and metacognitive processes set the stage for each other? *Learning and Instruction*, 50, 54–64. <https://doi.org/10.1016/j.learninstruc.2016.11.009>
- Saddler, B. & Andrade, H. (2004). The writing rubric. *Educational Leadership*, 62(2), 48–52.
- Sadler, P. M. & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1–31. https://doi.org/10.1207/s15326977ea1101_1
- Scardamalia, M., Bereiter, C. & Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science*, 8(2), 173–190. [https://doi.org/10.1016/S0364-0213\(84\)80016-6](https://doi.org/10.1016/S0364-0213(84)80016-6)
- Schunk, D. H. (1995). Self-efficacy and education and instruction. In J. E. Maddux (Hrsg.), *Self-efficacy, adaptation, and adjustment: Theory, research, and application. Plenum series in social/ clinical psychology* (S. 281–303). Springer. https://doi.org/10.1007/978-1-4419-6868-5_10
- Shute, V. J. (2007). Focus on formative feedback. *ETS Research Report Series*, 2007(1), i–47. <https://doi.org/10.3102/0034654307313795>

- Strijbos, J. W. & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265–269. <https://doi.org/10.1016/j.learninstruc.2009.08.002>
- Strijbos, J. W., Narciss, S. & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*, 291–303. <https://doi.org/10.1016/j.learninstruc.2009.08.008>
- van Ophuysen, S. (2010). Professionelle pädagogisch-diagnostische Kompetenz – eine theoretische und empirische Annäherung. In N. Berkemeyer, W. Bos, H. G. Holtappeis, N. McElvany & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Bd. 16, S. 203–234). Juventa.
- Walker, M. (2015). The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback. *Assessment & Evaluation in Higher Education, 40*(2), 232–247. <https://doi.org/10.1080/02602938.2014.898737>
- Wisniewski, B., Zierer, K. & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*(1), 1–14. <https://doi.org/10.3389/fpsyg.2019.03087>
- Zeeb, H., Bürgermeister, A., Saalbach, H., Renkl, A. & Glogger-Frey, I. (in Begutachtung). Effects of a digital support tool on student teachers' skills as self-regulated learners and as facilitators of self-regulated learning. *Unterrichtswissenschaft*.

Rekonstruktion mentaler Modelle Lehramtsstudierender zur pädagogischen Diagnostik

Claudia Schreiner | Fiona MacKay-Falls | Christian Kraler
Universität Innsbruck

Abstract: Pädagogisch-diagnostische Kompetenzen von Lehrpersonen sind sowohl für die Beurteilung von Schüler*innenleistungen (summativ) als auch im Rahmen der Lernbegleitung (formatives Assessment) erforderlich. In die verschiedenen Aspekte pädagogischer Diagnostik einzuführen, stellt eine spezifische Herausforderung für die Lehrerinnen- und Lehrerausbildung dar. Anhand eines konkreten Beispiels wird im ersten Teil dieses Beitrags die Verankerung der pädagogischen Diagnostik als wichtigem Bestandteil einer professionsorientierten, professionalisierenden Lehrerinnen- und Lehrerausbildung prototypisch-illustrativ dargestellt. Im zweiten Teil des Beitrags werden aus diesem Beispiel heraus Ergebnisse einer empirischen Untersuchung im Feld berichtet, die mithilfe von Freewriting-Mikroartikeln die Rekonstruktion mentaler Modelle Lehramtsstudierender zur pädagogischen Diagnostik zum Ziel hat. Die damit verbundenen Einblicke in die subjektiven Konzepte Studierender zum Thema ermöglichen Rückschlüsse sowohl auf deren Kenntnisstand als auch auf allfällige Adaptionen bzw. notwendige Weiterentwicklungen des Ausbildungscurriculums.

1 Pädagogische Diagnostik und Professionalisierung

Lehrpersonen agieren im formalen Bildungssystem (Rohlf's, 2011, S. 37 ff.) in der Regel in einer Doppelrolle, im Spannungsfeld zwischen unterrichtlicher, gruppenbezogener Lernbegleitung und individuumsbezogener Beurteilung der Lernergebnisse (Kraler, 2007). Die (Mit-)Gestaltung von Lernprozessen und die Evaluierung derselben bedürfen entsprechender Instrumente und Expertisen. Lehrende begleiten und leiten eine Gruppe Lernender beim Erarbeiten curricular vorgegebener (fachlicher) Inhalte. Der Unterricht entwickelt sich entsprechend dem didaktischen Dreieck (Gruschka, 2002) dialogisch zwischen lehrseitigem didaktischem Arrangement der Lehrpersonen und lernseitigem Aneignungsprozess der Lernenden. Diese Prozesse sind über Zwischenziele (Erarbeitung spezifischer Lerninhalte) auf Jahresziele sowie schulartspezifische fachbezogene und überfachliche Gesamtlernziele ausgelegt. Das erfolgreiche Erreichen von Lernzielen wird im formalen Bildungssystem in der Regel über in (Ziffern-)Noten gefasste Beurteilungen spezifiziert: Deren Definition ist vom Gesetzgeber vorgegeben (vgl. z. B. die in Österreich relevante Leistungsbeurteilungsverordnung¹).

1 <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009375>

Noten und Zeugnisse (Abgangszertifikate) stehen stets am Ende komplexer unterrichtlicher Lern- und Lehrprozesse. Sie stellen letzten Endes eine summative, standardisierte Vereinfachung komplexer Lernleistungen dar (Bölling, 2010).

Wesentliche gesellschaftliche Ansprüche an Schule werden etwa von Fend (2006), grundlegende Kritikpunkte von Holzkamp (1995) und eine Konkretisierung auf unterrichtliche Qualität von Helmke (2022) formuliert. Fend (spezifizierend) und Holzkamp (kritisierend) sprechen insbesondere die Selektions- und Allokationsfunktion von Schule an. Helmke betont das prozessuale Moment der innerunterrichtlichen (Qualitäts-)Entwicklung von Lernprozessen sowie der Schulqualität. Es handelt sich hierbei um eine Vielfalt summativ-zielbezogener (passende Bildungsentscheidung, Zeugnisse, ...) wie auch prozessorientierter Aspekte (zeitliche Taktung, Lernprozessbegleitung, ...). Was diese konzeptionell unterschiedlichen Handlungsfelder konstitutiv verbindet, ist die für deren Umsetzung grundlegende Notwendigkeit, diagnostische Verfahren in die Prozesse einzubinden. Mit der zweiten empirischen Wende gewannen zusätzlich quantitative psychometrische Verfahren auf Ebene der System- wie Individualdiagnostik an Bedeutung.

Insbesondere in den deutschsprachigen Ländern erfuhren damit Aspekte der Pädagogischen Diagnostik wieder erhöhte Aufmerksamkeit (Hesse & Latzko, 2017, S. 15 ff.; van Ophuysen, 2010); das nicht nur auf Ebene der (internationalen) Schulleistungsdiagnostik, sondern vor allem auch im Rahmen der Lehrerinnen- und Lehrerbildung (Hesse & Latzko, 2017, S. 20 ff.). Parallel zu den großen internationalen Schulleistungstudien (v.a. TIMSS, PIRLS und besonders PISA) zeichnete sich ein Entwicklungsschub in der inhaltlichen und strukturellen Gestaltung von Lehramtsausbildungen und deren Curricula ab. Ab etwa 2000 (Terhart, 2013) wurde basierend auf Befunden die Lehrerinnen- und Lehrerbildung in vielen Ländern der EU (EU, 2015; Zgaga, 2015) reformiert. Für Österreich liefert der Sammelband herausgegeben von Braunsteiner und Spiel (2019) einen umfassenden Überblick. Pädagogische Diagnostik stand nun auch in der Lehramtsausbildung neu zur Diskussion.

Ausgehend vom ersten, ursprünglich 1985 erschienenen klassischen Lehrbuch zur Pädagogischen Diagnostik von Ingenkamp (später mit Lissmann, 2008) erschienen ab Ende der 2010er Jahre parallel zu grundlegenden Reformen der Lehramtscurricula im Rahmen von Bologna (vgl. Kraler & Worek, 2021) weitere Werke zur pädagogischen Diagnostik (z. B. Jürgens & Lissmann, 2015; Jansen & Meyer, 2016; Hesse & Latzko, 2017). Vor allem in bildungswissenschaftlichen Teilen von Lehramtsausbildungen wird zudem der pädagogischen Diagnostik als eigenständigem Bereich ein zunehmendes Gewicht gegeben, vgl. z. B. die Curricula der Ausbildungsverbände in Österreich². Damit etabliert sich die pädagogische Diagnostik als eigenständiger (Kompetenz-)Bereich

2 Verbund Ost: <https://senat.univie.ac.at/curricularcommission/curricula/#c124174>; Verbund Mitte: <https://lehrerin-werden.at/lehramtsstudien/sekundarstufe-allgemeinbildung/curricula>; Verbund Süd-Ost: <https://www.lehramt-so.at/lehramtsstudien/>; Verbund West: <https://www.uibk.ac.at/de/studien/uf-sekundaerstufe/>

und hat sich letzten Endes von der psychologisch orientierten Diagnostik endgültig emanzipiert und abgegrenzt. Im gegenwärtigen Forschungsstand werden diagnostische Kompetenzen von Lehrkräften in der Regel im Spektrum deren professioneller Kompetenzen verortet (vgl. Heinrichs, 2015, S. 37 f.).

2 Lehrerinnen- und Lehrerbildung, pädagogische Diagnostik und formatives Assessment

Eine geschlossene Definition der pädagogischen Diagnostik stellt jedoch aktuell aufgrund ihrer dynamischen Entwicklung noch bzw. wieder ein Desiderat dar. Lange Zeit konnte primär auf Ingenkamp (vgl. Ingenkamp & Lissmann, 2008, S. 13) Bezug genommen werden. Ein Blick in die zitierte Literatur zeigt jedoch, dass ein primär individualdiagnostisch fokussiertes Verständnis heute zu kurz greift. Es lässt etwa Skalierungsaspekte bis hin zur Systemdiagnose unberücksichtigt. Eine Konvergenz der genannten Autorinnen und Autoren ergibt sich jedoch hinsichtlich der Verortung. Pädagogische Diagnostik findet in der Schule statt, hat einen primären Schulbezug und fokussiert auf individuelle, gruppenbezogene, bildungsinstitutions- und bildungssystemspezifische Lern- und Entwicklungsprozesse. Formatives Assessment nach Black und Wiliam (2009) kann demnach als Teil pädagogischer Diagnostik angesehen werden, in dem es vorrangig darum geht, das Lernen von Schülerinnen und Schülern innerhalb einer Lerngruppe durch die Verwendung diagnostischer Informationen zu verbessern. Es handelt sich um einen Ausschnitt pädagogischer Diagnostik, der als integraler Bestandteil des Lern-Lehr-Geschehens auf formative Zwecke und das Lernen der Lernenden selbst fokussiert.

Im Schulalltag bezieht sich pädagogische Diagnostik meist auf ein lehrpersonenberufsspezifisch konstitutives Handlungsfeld. Dieses ist seit Einführung der Schulpflicht mit summativen und formativen Beurteilungsfunktionen verbunden. Im didaktischen Dreieck stellen über das (gesetzlich vorgegebene) Curriculum Unterrichtsinhalte normative, von Lehrenden zu unterrichtende und den Lernenden zu lernende Bezugsgrößen dar (z. B. Oelkers, 2009). Der unterrichtliche Lehr-Lern-Prozess entwickelt sich dreifach, über eine initiale Status-Erhebung der Ausgangsbedingungen sowie eine kontinuierliche und eine – zu definierten Zeiten – abschließende Kontrolle des Lernfortschritts, wobei alle drei Aspekte gegenwärtig im Kontext professionsorientierter Lehrerbildungskonzepte pädagogisch-diagnostisch formuliert werden (Buch & Sparfeldt, 2020, S. 39; Buholzer et al., 2020, S. 630).

Pädagogische Diagnostik und diagnostisches Handeln umfassen im Berufsalltag von Lehrpersonen ein breites inhaltliches Spektrum. Formales Lernen, Lehren und Unterrichten unterliegen heute in einem von sozialen, technologischen und wirtschaftlichen Entwicklungen und steter Beschleunigung geprägten Umfeld sowie nach dem paradigmatischen Wandel vom Lehren zum Lernen (Barr & Tagg, 1995) und der zweiten empirischen Wende wesentlich komplexeren Anforderungen im Unterschied zu traditionellen, tendenziell unidirektional didaktisierten Lehr-Lern-Arrangements (Gräsel &

Mandl, 1999). Entsprechend differenziert und vielfältig stellen sich Fragen und Aspekte der pädagogischen Diagnostik in Forschung, Ausbildung (Lehre) und im aktuell vom über Bildungsstandards und Kompetenzen konzeptualisierten Berufsalltag dar. Die spezifischen Gegenstände pädagogischer Diagnosen, Funktionen und Ziele pädagogischer Diagnostik sowie Methoden zur Gewinnung diagnostischer Informationen sind im Kontext allgemeindidaktisch-bildungswissenschaftlicher wie fachdidaktisch-fachlicher Zusammenhänge verschränkt.

Die vielfältigen Verflechtungen im Berufsalltag von Lehrpersonen bedingen, dass – wie oben erwähnt – Grundlagen der pädagogischen Diagnostik in den zentralen Aspekten inzwischen Bestandteil von Ausbildungscurricula im Lehramt sind (vgl. Kraler et al., 2017, S. 203; Schreiner & Kraler, 2019). In diesen wird in der Regel besonderes Augenmerk auf die Differenzierung hinsichtlich der funktionalen Prozess- bzw. Produktorientierung (z. B. Heinrichs, 2015) gelegt, da sich hier insbesondere das professionsspezifische Spannungsfeld zwischen Lernbegleitung und Beurteilung der Lernleistung fruchtbar thematisieren und bearbeiten lässt.

In diese vielfältigen komplexen Zusammenhänge der pädagogischen Diagnostik im Rahmen der Lehrerinnen- und Lehrerausbildung einzuführen, stellt eine eigene Herausforderung dar. Die Diskussion in den folgenden Abschnitten wird hierzu mit einem doppelten Ziel geführt. Erstens wird an einem konkreten Beispiel aufgezeigt, wie Fragen der pädagogischen Diagnostik in der Lehrerinnen- und Lehrerausbildung umgesetzt werden können (siehe Abschnitt 3). Hierbei steht der prototypisch-illustrative Charakter im Vordergrund. Das konkret dargestellte Beispiel soll als ein Ausdruck des Ist-Stands im System Lehrerinnen- und Lehrerbildung im deutschen Sprachraum gelesen werden, wie pädagogische Diagnostik inzwischen als wichtiger Bereich in einer professionsorientierten, professionalisierenden Ausbildung verankert ist. Zweitens werden aus diesem Beispiel heraus Ergebnisse einer empirischen Untersuchung im Feld dargestellt, die die Rekonstruktion mentaler Modelle Studierender zur pädagogischen Diagnostik zum Ziel hat (Abschnitt 4). Sie illustrieren damit ohne Anspruch auf grundlegende Verallgemeinerbarkeit, wo die Lehramtsausbildung hinsichtlich der Vermittlung pädagogisch-diagnostischer Inhalte aktuell steht.

3 Pädagogische Diagnostik im Lehramtsstudium

Die Verankerung der pädagogischen Diagnostik im Lehramts-Curriculum sowie die konkrete Ausformung des entsprechenden Moduls sollen nun anhand des Beispiels der Universität Innsbruck dargestellt werden. Damit wird erstens eine konkrete Umsetzungsform exemplarisch beschrieben, welche zweitens den Kontext der in Abschnitt 4 präsentierten empirischen Untersuchung bildet.

Für die Erarbeitung des Curriculums waren folgende didaktische Prinzipien sowohl hinsichtlich der Gesamtarchitektur wie auch spezifischer Gebiete (auch der pädagogischen Diagnostik) leitend:

- das Backward Design (Wiggins & McTighe, 2005): die Planung der Lehr-Lern-Prozesse ausgehend von den zu erreichenden Lernzielen
- das Spiralprinzip nach Bruner (1960): das wiederholte Bearbeiten zentraler Ausbildungsinhalte über das Studium auf jeweils unterschiedlichen Kompetenzniveaus
- ein exemplarisch-prototypisch-genetischer Zugang (Wagenschein, 1968): herausfordernde Probleme als Ausgangspunkt des Lernens, Bearbeitung dieser als Prototypen
- das Bildungsgangprinzip (Hericks, 2006; Kraler & Schratz, 2012): Berücksichtigung des objektiven und je subjektiven Bildungsgangs mit der Perspektive auf eine berufslebenslange Professionalisierung

Dabei bilden die fünf Bereiche Professionalisierung, Lernen, Diversität, Praxis und pädagogische Diagnostik als inhaltliche Schwerpunktsetzung der pädagogisch-psychologischen Ausbildung den roten Faden durch das Studium (Kraler et al., 2017). Diese Bereiche werden zu Studienbeginn inhaltlich umrissen und über den Studienverlauf hinweg vertieft und entfaltet, damit sie sich bei den „Studienende[n] zu einem ersten kohärent-integrativen Wissensbestand“ (Kraler et al., 2017, S. 202) entwickeln.

Die vertiefte Auseinandersetzung mit pädagogischer Diagnostik ist in einem eigenen Modul der bildungswissenschaftlichen Grundlagen ab dem 4. Studiensemester vorgesehen. Es besteht aus drei Lehrveranstaltungen mit insgesamt 7,5 ECTS-Anrechnungspunkten. Zentrale Themen des Moduls sind die pädagogische Diagnostik (formativ), Kommunikation und Beratung sowie Classroom-Management und Leistungsbeurteilung (als summative Form der Diagnostik).

Im vorliegenden Beitrag konzentrieren wir uns auf eine spezifische Lehrveranstaltung, das Proseminar ‚Diagnostik & Beratung‘. Dieses hat zum Ziel, bezogen auf formative pädagogische Diagnostik zwei zentrale Aspekte aufzugreifen: (1) eine systematische problemfokussierte Diagnose von Lernschwierigkeiten, die anlassbezogen bei der Wahrnehmung von Umständen, die das Lernen von Schülerinnen und Schülern behindern könnten, initiiert wird, sowie (2) formatives Assessment als integraler Bestandteil des Unterrichts im Kontext von Förderung und Lernbegleitung.

Diese beiden Techniken unterscheiden sich insbesondere durch den Ausgangspunkt des Diagnoseprozesses. Während beim formativen Assessment die grundsätzlich in den Lehr-Lern-Prozess integrierte Diagnose zur kontinuierlichen Versicherung darüber, wo Lernende gerade im Hinblick auf das (nächste) Lernziel stehen, gemeint ist, ist der Ausgangspunkt der problemfokussierten Diagnose die Wahrnehmung eines (potenziellen) Problems. Zum Üben problemfokussierter Diagnosen bearbeiten die Studierenden in Kleingruppen je ein Fallbeispiel, in dem Wahrnehmungen zu einem spezifischen Schüler oder einer Schülerin, die auf potenziell das Lernen behindernde Umstände hindeuten, den Ausgangspunkt für diagnostisches Handeln bilden. Dabei erarbeiten die Studierenden ausgehend von so einem diagnostischen Szenario im Sinne eines prototypisch-genetischen Zugangs einen Diagnoseplan. Die diagnostischen

Szenarien sind bewusst realitätsnah angelegt und daher vielschichtig. Entsprechend kann es nicht darum gehen, eine Musterlösung für einen Fall zu finden, sondern vielmehr prototypisch-diagnostische Erfahrung über die Bearbeitung des Falls zu erlangen. Aufbauend auf einem fiktiven Diagnoseergebnis, wird dann ein Interventionsplan erarbeitet. Ziel ist eine vertiefte Auseinandersetzung mit einem konkreten Schüler oder einer konkreten Schülerin, ihrem Umfeld und möglichen Problemfeldern sowie Handlungsmöglichkeiten.

Als Grundlage für formative Diagnostik im Kontext der Lernbegleitung bezieht sich die LV auf das Modell des formativen Assessments nach Black und Wiliam (2009). „Embedded Formative Assessment“ (Wiliam, 2018) bildet die zentrale Pflichtliteratur für diesen Bereich, welche durch zwei Broschüren zu pädagogischer Diagnostik aus dem deutschsprachigen Raum (SQA, 2015; ISB, 2008) ergänzt wird. Das Modell des formativen Assessments mit den fünf Strategien – Lernziele klären und kommunizieren, Evidenzen über das Lernen generieren, lernförderliches Feedback geben, Peers als Ressource nutzen sowie die Lernenden zu Verantwortlichen ihres eigenen Lernens machen – bildet einen integralen Bestandteil des gesamten Diagnose-Moduls. Erstens stellt es einen zu erarbeitenden Inhalt dar. Die einzelnen Strategien werden in Form von Expertengruppen von den Studierenden erarbeitet und an die Mitstudierenden weitervermittelt. Hierbei ist die Perspektive der Studierenden die der angehenden Lehrpersonen; es geht um das Kennenlernen der Strategien als Werkzeug zur Gestaltung des eigenen (zukünftigen) Unterrichts. Zweitens bilden die fünf Strategien aus dem Modell Anwendung in der didaktischen Gestaltung des Kurses. Dadurch erleben die Studierenden die Strategien als Lernende und reflektieren in einer abschließenden Selbstanalyse, wie die verschiedenen Strategien zu ihrem persönlichen Lernprozess in der LV beigetragen haben. So können die Studierenden Aspekte formativen Assessments sowohl aus der Lehrenden- als auch aus der Lernenden-Perspektive bearbeiten.

Hierbei ist das Ziel, problemfokussierter, systematischer Diagnose als etwas Besonderem und formativem Assessment als integralem Bestandteil von zielgerichteten Lern-Lehr-Prozessen gleichermaßen entsprechendes Gewicht zu geben. Eine zentrale Herausforderung besteht in der Ausbildung mentaler Modelle zur pädagogischen Diagnostik, die eine wirkungsvolle Integration in die eigene (zukünftige) Lehrpraxis unterstützen.

4 Rekonstruktion mentaler Modelle zur pädagogischen Diagnostik

Welche subjektiven mentalen Modelle Studierende ausbilden, ist ein wesentliches Ergebnis des Lehr-Lern-Prozesses. Dies ist im Kontext einer Lehrerinnen- und Lehrerbildung zu sehen, die davon ausgeht, dass Lernen primär als Vollzug von Selbstlernen konzipiert werden kann. Was Studierende lernen, welche Kompetenzen sie aufbauen können, ist von den Lehrenden nur indirekt beeinflussbar, insbesondere über die Bereitstellung adäquater Lernräume, die verständnisvolles, sinnstiftendes Lernen er-

möglichen. Ein zentrales Moment beim Lernen ist der Aufbau mentaler Modelle über mentale Repräsentationsmuster (Kraler et al., 2006). In der Literatur werden hierfür unterschiedliche Terminologien verwendet, z. B. kognitive Landkarten (Tolman, 1948; Jensen, 2006) oder mentale Repräsentationen (Dann, 1989). Gemeint sind hierbei stets kognitive Modelle spezifischer (meist) außenweltlicher Phänomene bzw. Zusammenhänge. Seit den 2000er Jahren werden diese Konzepte unter Bezug auf Charles Sanders Peirce (1896/1931, *Lessons from the History of Science*), Ludwig Wittgenstein (1922, *Tractatus logico-philosophicus*) und Kenneth Craik (1943, *The Nature of Explanation*) zunehmend unter dem Terminus „Mentale Modelle“ subsumiert. Jones et al. (2011, S. 45) definieren mentale Modelle wie folgt:

„Mental models are personal, internal representations of external reality that people use to interact with the world around them. They are constructed by individuals based on their unique life experiences, perceptions, and understandings of the world. Mental models are used to reason and make decisions and can be the basis of individual behaviors. They provide the mechanism through which new information is filtered and stored.“

4.1 Zur empirischen Untersuchung

4.1.1 Methodisches Vorgehen

Das Verständnis von pädagogischer Diagnostik und ihre konkrete Ausgestaltung sind noch dynamisch im Fluss. Daher wird ein rekonstruktiver Ansatz zur Untersuchung gewählt. Zur Datenerhebung wird in Anlehnung an Kraler et al. (2006) die Methode des Freewriting (Elbow, 1973) mit der Struktur der Mikroartikel (Willke, 2009, 2018) kombiniert.

Mikroartikel wurden von Willke (2018) im Kontext des organisationalen Wissensmanagements entwickelt. Sie sind ein Instrument, um eine „(normalerweise) individuelle Lernerfahrung, Erkenntnis, Idee, Expertise, Überlegung, Reflexion etc. in eine bestimmte Form zu bringen“ (ebd., S. 94). Dabei ist Voraussetzung, dass der Autor oder die Autorin eine Lernerfahrung macht, also Wissen über das zu Verschriftlichende generiert bzw. explizit macht. Damit wird das mentale Muster einer spezifischen kognitiven Landkarte, d. h. die subjektive Geographie zu einem inhaltlichen Feld, beschrieben. Wir orientieren uns in unserer Datenerhebung an der Grundidee und Grundstruktur der Mikroartikel und kombinieren diese mit einer spontanen, intuitiven Schreibweise, der Technik des Freewriting.

Freewriting wurde in den 1970er Jahren als Methode kreativen Schreibens entwickelt (Elbow, 1973). Kernidee ist, regelmäßig für eine definierte Zeitspanne (zwischen 10 und 20 Minuten) den laufenden gedanklichen Prozess (die innere Stimme) zensurfrei niederzuschreiben. Zentral ist hierbei, den Schreibfluss aufrecht zu erhalten, ohne zu korrigieren oder zu überarbeiten. Wenn der Schreibprozess ins Stocken gerät, wird dennoch weitergeschrieben, etwa indem einfach festgehalten wird, dass man im Moment keine Idee hat, was man schreiben könnte (ebd.). Freewriting kann dabei unterstützen, einen inneren Monolog zu einem gewissen Grad explizit zu machen. Es wird ein in-

tuitives, vergleichsweise unzensuriertes Schreiben initiiert, das das Potenzial enthält, mentale Landkarten abzubilden.

Für die Rekonstruktion mentaler Modelle der Studierenden wurde nun eine Struktur ähnlich eines Mikroartikels nach Willke (2009, 2018) vorgegeben, welche in Form von Freewriting bearbeitet werden sollte. Folgende drei Impulsfragen fokussierten und strukturierten den Prozess:

1. Was ist für Sie pädagogische Diagnostik?
2. In welchen Kontexten bzw. Situationen ist für Sie pädagogische Diagnostik von ganz besonderer Bedeutung?
3. Was ist für Sie als Lehrperson im Unterricht im Hinblick auf pädagogische Diagnostik am wichtigsten?

Die Studierenden hatten insgesamt 15 Minuten Zeit, frei zu notieren, was ihnen in den Sinn kam. Dabei wurden die Impulsfragen als strukturierendes Element im Abstand von etwa jeweils fünf Minuten vorgelesen.

Ziel war, das persönliche semantische Feld explizit aufzuspannen. Dies gelingt über einen kurzen Prosatext deshalb leichter als etwa über Mindmaps, weil das Schreiben auch emotionale sowie kommunikative Aspekte aufweist. „Sich erzählen heißt sich (er) finden, heißt sich erleben, sich mit allen Sinnen begreifen“ (Kraler et al., 2006, S. 34). Die Reihenfolge der Fragen ist dabei von grundlegender Bedeutung. Sie führen von einer begrifflichen Annäherung über eine Kontextualisierung zur Beschäftigung mit für einen persönlich besonders bedeutsamen Bedingungen. Durch die Kombination dieser beiden Zugänge soll ein Zugang zu mentalen Modellen geschaffen werden, indem zum einen im Zuge des Freewriting den Gedanken freier Lauf gelassen und zum anderen durch die Impulsfragen in der Struktur des Mikroartikels eine inhaltliche Fokussierung angeregt wird.

Die Mikroartikel der Studierenden wurden mithilfe von MaxQDA analysiert. Hierfür wurde induktiv ein Kategoriensystem entwickelt. Die Texte wurden je Impulsfrage verarbeitet. Damit sollen die studierendenseitigen mentalen Modelle zur pädagogischen Diagnostik (basierend auf der Ausbildungsintervention im Modul) illustriert werden. Wir beschränken uns in der Darstellung auf zentrale Befunde. Für die Ergebnisdarstellung wurden die zentralen induktiv entwickelten Kategorien zu Frage 1 ausgewählt und anschließend durch über Frage 1 hinausgehende für das Verständnis relevante Aspekte der Studierendentexte zu den Fragen 2 und 3 ergänzt.

4.1.2 Forschungskontext und Stichprobe

Zielgruppe der aktuellen Studie sind Lehramtsstudierende ab dem 4. Semester der bildungswissenschaftlichen Grundlagen, wo pädagogische Diagnostik schwerpunktmäßig verortet ist. Ihre mentalen Modelle zur pädagogischen Diagnostik stehen im Mittelpunkt des Interesses. Die Teilnehmenden von fünf Parallelgruppen der LV ‚Diagnostik & Beratung‘ des Sommersemesters 2022 (zwischen 20 und 24 Teilnehmende pro Gruppe) wurden im vorletzten LV-Termin eingeladen, mithilfe einer strukturierten

Freewriting-Übung über ihre Vorstellungen zur pädagogischen Diagnostik zu reflektieren. Insgesamt haben 90 Studierende ihre Mikroartikel für die wissenschaftliche Analyse zur Verfügung gestellt, welche die Analysebasis für die im Folgenden vorgestellten Ergebnisse bilden.

4.2 Ergebnisse

Die Analyse und Kodierung des Datenmaterials liefern eine umfassende Bestandsaufnahme von Bausteinen der mentalen Modelle der Studierenden zur pädagogischen Diagnostik. Im Kontext der ersten und teilweise zweiten Impulsfrage nimmt die erworbene Wissensbasis der Studierenden eine dominante Rolle ein. Hier finden sich in den Texten vor allem Bezugnahmen auf Konzepte, Elemente und Instrumente der pädagogischen Diagnostik, die explizit Bestandteil der LV sind. Die Texte zu Frage 3 sind deutlich persönlicher geprägt und enthalten Hinweise auf Einstellungen und Werte der Studierenden. Im Folgenden werden relevante Aspekte der Studierendentexte entlang der drei Impulsfragen dargestellt. Dabei bedingt das Format dieses Beitrags die Notwendigkeit einer entsprechenden Zusammenfassung und Verdichtung auf relevante Kernthemen. Ausschnitte aus den Studierendentexten sind in Kursivsetzung gekennzeichnet³.

4.2.1 Was ist für die Studierenden pädagogische Diagnostik?

Die erste Impulsfrage spricht das subjektive Konzept von pädagogischer Diagnostik an. Folgende zentrale Themenfelder, die im Wesentlichen den entwickelten Kategorien entsprechen, ergaben sich aus dem Datenmaterial.

Pädagogische Diagnostik als Kernkompetenz von Lehrpersonen

Am Ende eines Semesters intensiver Beschäftigung mit Aspekten der pädagogischen Diagnostik nennen viele Studierende diese als ein Kernelement professionellen Handelns von Lehrpersonen, so z. B. im folgenden Textausschnitt:

Pädagogische Diagnostik ist für mich ein Vorgang, mit dem ich als künftige Lehrerin viel zu tun haben werde und den ich gut professionalisieren möchte, da es zu einer der Kernkompetenzen des Lehrer:innenberufes gehört. (Frage 1, Pos. 19)

Die Studierenden äußern, dass pädagogische Diagnostik eine wissenschaftliche Fundierung braucht. Dies ist im Kontext der Lehrerinnen- und Lehrerbildung als erster Schritt professioneller Entwicklung von besonderem Interesse.

³ Die Zitate aus den Mikroartikeln werden grundsätzlich im Original wiedergegeben; die Texte wurden jedoch bzgl. Tipp- und Rechtschreibfehlern, die natürlicher Bestandteil von Freewriting sind, sowie Zeichensetzung für den Zweck der Publikation bereinigt, um die Lesbarkeit zu verbessern.

Um Diagnostik-Instrumente erfolgreich einzusetzen, braucht es einen theoretischen Rahmen, der dann sozusagen als Anleitung für die praktische Umsetzung dient. (Frage 1, Pos. 20)

Dabei wird pädagogische Diagnostik von etlichen Studierenden explizit als Teil des Berufsalltags beschrieben.

Pädagogische Diagnostik findet sich laufend im Schulalltag wieder und trägt grundlegend zu Problemlösungen bei. (Frage 1, Pos. 10)

Erheben von Lernständen und darauf aufbauende Förderung

Als konsistent und durchgehend identifiziertes Thema zeigt sich pädagogische Diagnostik als das Nutzen von Daten, um den aktuellen Lernstand zu identifizieren und die Lehr-Lern-Prozesse entsprechend anzupassen. Ein Beispiel hierfür ist etwa Folgendes:

*Hier wird versucht, Lernstände von Schüler*innen einzuschätzen und passende Fördermaßnahmen in Form von Förderplänen vorzubereiten und umzusetzen. Sie umfasst alle Tätigkeiten, die die Voraussetzungen und Bedingungen für erfolgreiche Lehr- und Lernprozesse eines Lernenden ermitteln. (Frage 1, Pos. 45)*

Dabei waren das Verbessern des Lernprozesses der Lernenden bzw. ‚Lernen zu optimieren‘ wiederkehrende Themen. Bei einer Reihe von Antworten spielte darüber hinaus das Element der Analyse eine wichtige Rolle. Hierbei nehmen viele Studierende Bezug auf die individuelle Förderung, wie in den folgenden Beispielen ersichtlich ist.

Das Ziel ist es, den individuellen Lernprozess zu verbessern. (Frage 1, Pos. 4)

Als Lehrperson beobachtet man also die Lernprozesse aller einzelnen Schüler und analysiert sie, gibt Feedback und versucht so, das Lernen der einzelnen Schüler/innen zu optimieren. (Frage 1, Pos. 47)

Pädagogische Diagnostik kann sich aber auch auf positive Aspekte beziehen, beispielsweise wenn man als Lehrperson bemerkt, dass einzelne SchülerInnen besondere Fähigkeiten zeigen, diese zu diagnostizieren und dann darauf reagierend eine Möglichkeit zu geben, sich weiter in diese Richtung zu entwickeln. (Frage 1, Pos. 77)

Einen stärkeren Bezug auf den Lehr-Lern-Prozess insgesamt findet man in einzelnen Texten, wie zum Beispiel im folgenden Textausschnitt, wo explizit darauf Bezug genommen wird, die Daten über das Lernen der Schülerinnen und Schüler zu nutzen,

um die Lehr- und Lernbedingungen in meiner Klasse zu ermitteln, zu reflektieren und weiterzuentwickeln. (Frage 1, Pos. 20)

Problemidentifikation als primäres Ziel pädagogischer Diagnostik

Studierende nahmen umfangreich Bezug auf pädagogische Diagnostik als Ausgangspunkt zur Identifikation von Problemen. Dabei geht es um das Erkennen von Problemen und Herausforderungen und die genauere Eingrenzung des Problems.

Was ich bis jetzt weiß, ist, dass pädagogische Diagnostik das Erkennen und Definieren von Problemen im schulischen Kontext ist. (Frage 1, Pos. 106)

Vor allem im Kontext der Problemidentifikation betonten viele Studierende die Relevanz eines strukturierten Prozesses, eines systematischen Vorgehens im Rahmen der Diagnose und klarer Ziele, zum Beispiel:

Außerdem empfinde ich es als strukturierte Diagnostik mit klaren Zielen und einem lösungsorientierten Ansatz. (Frage 1, Pos. 9)

Methoden

Die Studierenden nehmen in diesem ersten Textteil vermehrt Bezug auf verschiedene Methoden der pädagogischen Diagnostik, was die intensive Auseinandersetzung mit verschiedenen Instrumentarien in der LV widerzuspiegeln scheint. Dabei ist das Beobachten die am häufigsten referenzierte Methode, welche insbesondere als erster Teil oder Ausgangspunkt des Diagnoseprozesses erwähnt wird.

Kooperativer Aspekt diagnostischer Tätigkeiten

Das Thema Zusammenarbeit wird von den Studierenden sowohl bei Frage 1 als auch unter Frage 3 hervorgehoben. Es wird als Kernelement pädagogischer Diagnostik beschrieben, sowohl für die Identifikation eines Problems als auch zum Finden potenzieller Lösungsmöglichkeiten. Die Zusammenarbeit kann sich dabei auf kollegiale Zusammenarbeit unter Lehrpersonen, mit den Eltern oder außerschulischen Ansprechpartnerinnen und -partnern beziehen. Darüber hinaus wird die aktive Einbeziehung der Lernenden betont.

*Natürlich werden dabei auch andere Personen involviert, z. B. Schulpsycholog*innen, Sozialberater*innen usw. Die Lehrperson fungiert hier aber als erster Ansprechpartner für Schüler*innen und Eltern und schafft daraufhin, evt. auch gemeinsam mit Kolleg*innen, ein Netzwerk aus allen Personen, die in den Fall involviert werden, um die Schüler*innen bestmöglich zu unterstützen. (Frage 1, Pos. 117)*

Wichtig ist es nicht nur allein zu handeln, sondern auch den Schüler oder die Schülerin in die Diagnose mit einbeziehen zu lassen. Der Schüler oder auch die Schülerin weiß oft selbst, an was es liegt, und kann einer Lehrperson somit auch weiterhelfen Schlüsse zu ziehen und vor allem Lösungen zu finden [...]. (Frage 1, Pos. 119)

4.2.2 Kontexte und Situationen, in denen pädagogische Diagnostik als relevant angesehen wird

Mit Impulsfrage 2 wurden die Studierenden explizit nach Kontexten und Situationen gefragt, in denen pädagogische Diagnostik für Sie besondere Relevanz hat. Durchgehend wird von den Studierenden darauf Bezug genommen, dass pädagogische Diagnostik in einer Reihe von Kontexten (individuell, auf der Ebene der Klasse, außerschulisch) und in Beziehung zu verschiedenen Arten von Schwierigkeiten (Lernschwierigkeiten, Verhalten, in der sozialen Konstellation etc.) relevant ist.

Bei Klassen mit Problemen in verschiedenen Gegenständen; Lernschwierigkeiten fachbezogen, aber auch verhaltensbedingt. (Frage 2, Pos. 11)

Lässt sich ein Problem im Elternhaus verorten? Ist die Klassensituation problematisch? Liegt womöglich im Kind selber die Ursache, sprich: Lässt sich eine Störung erkennen, oder handelt es sich um anderwärtig ausgeprägte Persönlichkeitsmerkmale, die zu Verhaltensauffälligkeiten führen? (Frage 2, Pos. 74)

Das mögliche Umfeld für Probleme wird bei Frage 2 deutlich weiter aufgespannt als bei der Bearbeitung von Frage 1. So spielt die Klassenebene inkl. Klassenklima und dem spezifischen Thema Mobbing (was vermutlich auf ein im relevanten Semester eingesetztes Fallbeispiel zu diesem Thema zurückzuführen ist) eine prominente Rolle in der Kontextualisierung pädagogischer Diagnostik.

In Problemsituationen, speziellen puhhh schwierige Frage, in Situationen, wo ein Konflikt entsteht, der sich auch vertiefen kann. (Frage 2, Pos. 68)

4.2.3 Was für die Studierenden im Unterricht im Hinblick auf pädagogische Diagnostik am wichtigsten ist

Zwei Aspekte scheinen dazu zu führen, dass die Texte zu Impulsfrage 3 persönlicher und stärker Repräsentationen auch mentaler Modelle zeigen. Zum einen wird mit fortschreitender Dauer der Übung des freien Schreibens ein stärker assoziatives Schreiben angeregt. Zum anderen spricht die dritte Frage explizit an, was als persönlich bedeutsam angesehen wird. Während viele Studierende bei Frage 1 vor allem auf das erworbene deklarative Wissen zurückgreifen, beziehen sich die Texte bei Frage 3 vermehrt auf persönliche Überzeugungen, Haltungen und Werte. Insofern sind die Themen, die sich aus dem Datenmaterial zu Frage 3 ergeben, von besonderem Interesse in Bezug auf die Rekonstruktion mentaler Modelle der Studierenden.

Während auch bei Frage 3 einige konkrete auf Diagnostik bezogene Themenbereiche angesprochen werden, treten insbesondere Ansprüche an sie selbst als Lehrpersonen ins Zentrum der Texte. Konkreter bezogen auf pädagogische Diagnostik und ihre Aufgabe in der Schule ist Diagnostik als integraler Bestandteil des Unterrichts ein prominentes Thema. Wie die folgenden Textausschnitte zeigen, sind die Antworten bei Frage 3

stärker von persönlichen Überzeugungen geprägt als noch am Beginn des freien Schreibens. Dabei treten die Schülerinnen und Schüler in den Mittelpunkt, sollen als ‚Menschen‘ wahrgenommen werden und aktive Teilnehmende sein, wie die folgenden zwei Ausschnitte illustrieren.

Es ist für mich wichtig, dass ich mich nicht nur mit Lerninhalten beschäftige, sondern auch mit den Menschen, denen ich sie vermitteln möchte. Ich möchte, dass jeder/jede bestmöglich vom Unterricht profitieren kann. Wenn ich merke, dass ein Schüler/eine Schülerin nicht mitkommt, möchte ich die Situation so lange optimieren, bis es passt. Jeder Mensch lernt anders gut und ich sehe es als meine Aufgabe an, die Voraussetzungen dafür zu schaffen. (Frage 3, Pos. 55)

Weiters halte ich es für wichtig, besonders in der individuellen Diagnostik die Perspektive des Kindes selbst zu beleuchten, da eine außenstehende Person innere Prozesse schwer sehen oder beschreiben kann, während das ein Kind oft gut kann. (Frage 3, Pos. 12)

Eine Reflexion der eigenen Rolle als Lehrperson ist Teil vieler Texte zu Frage 3. Das betrifft die Fragen, wie sich die Studierenden zukünftig selbst sehen (möchten), wie sie ihren Schülerinnen und Schülern gegenüber treten und von ihnen gesehen werden möchten. Dabei ist die Perspektive von Lehrpersonen als Fürsprecherinnen bzw. Fürsprecher der Schülerinnen und Schüler ein prominentes Thema sowie Vertrauen und für die Lernenden da sein.

*Dass meine Schüler*innen wissen, dass sie sich an mich wenden können. Dass ich eine Ansprechperson für meine Schüler*innen bin. Dass ich im Notfall in der Lage bin, ein Netzwerk zu schaffen, das meinen Schüler*innen weiterhilft. Dass ich die Kompetenz entwickle, ihnen weiterzuhelfen. Dass ich mir die Zeit nehme, meinen Schüler*innen zu helfen. [...] Dass sich die Eltern bei Problemen an mich wenden können. (Frage 3, Pos. 112)*

*Es ist wichtig, sich nicht allein gelassen zu fühlen, denn dann können Situationen schnell überfordernd wirken und nicht mehr in Ruhe bearbeitet werden. Ich persönlich möchte mich auf die Schüler*innen einlassen, ihnen die Möglichkeit geben, mit Fragen, Wünschen und Problemen zu mir zu kommen. (Frage 3, Pos. 18)*

*Deshalb sollte ich als Lehrperson immer aufmerksam sein und sensibel mit den Problemen der Schüler*innen im Unterricht umgehen. (Frage 3, Pos. 99)*

In Bezug auf die Identifikation von Problemen werden insbesondere Aspekte professionellen Handelns angesprochen, welches sich laut den Studierenden in Objektivität, einem systematischen Vorgehen sowie der Wichtigkeit von Selbstreflexion zeigt.

Objektivität sollte hier im Fokus stehen, auch wenn dies natürlich in gewissen Situationen sehr schwer sein kann. (Frage 3, Pos. 33)

Objektivität, Einfühlungsvermögen, Verständnis und Respekt. Ohne diese Aspekte wird pädagogische Diagnostik nicht zielbasiert funktionieren. (Frage 3, Pos. 51)

Ein gezielter Umgang mit Problemen. Das bedeutet eine strukturiert und gut vorbereitete Herangehensweise, damit ein Problemfall gezielt und gemeinsam gelöst werden kann. [...] (Frage 3, Pos. 9)

Für mich als Lehrperson ist es am wichtigsten, nie das Vermögen zu verlieren, mich selbst zu hinterfragen und auch mein eigenes Handeln und Sein zu beobachten. Ich denke, es ist ein großes Problem im Schulalltag, dass viele Lehrer denken, sie seien den Schülern überlegen. (Frage 3, Pos. 103)

5 Resümee

Pädagogische Diagnostik wird gegenwärtig wieder verstärkt als eigenständiger Bereich und zentrale Aufgabe von Lehrpersonen angesehen. Inzwischen gibt es verschiedene Perspektiven auf das, was heute pädagogische Diagnostik genannt wird. Der empirisch-rekonstruierende Zugang im vorliegenden Beitrag hat zweierlei zum Ziel. Er soll das Verständnis pädagogischer Diagnostik als Basis für einen Diskurs erweitern, der erforderlich ist, um den Prozess der Professionalisierung voranzutreiben. Zweitens sollte idealerweise eine Beforschung mentaler Modelle Studierender Bestandteil der laufenden Begleitforschung zur kontinuierlichen Weiterentwicklung des Ausbildungsmoduls zur pädagogischen Diagnostik sein, um evidenzorientiert die Studienqualität weiterentwickeln zu können.

5.1 Die Rekonstruktion mentaler Modelle als Instrument der Begleitforschung zum Lehramtsstudium

Resümierend kann festgehalten werden, dass die Rekonstruktion mentaler Modelle Studierender – im vorliegenden Fall zur pädagogischen Diagnostik – interessante Einblicke hinsichtlich der Entwicklung Studierender im Rahmen der Lehrerinnen- und Lehrerausbildung bietet. Diese stellen eine substanzielle Basis für die Diskussion des derzeitigen Ausbildungsmoduls, seine Stärken und Schwächen und mögliche Weiterentwicklungen dar. Sie ergänzen Erkenntnisse aus traditionellen Methoden der LV-Evaluierung etwa in Form standardisierter Fragebögen in fruchtbarer Weise.

5.2 Zentrale Erkenntnisse in Bezug auf das beforschte Ausbildungsmodul

Konkret auf das an der Universität Innsbruck implementierte und laufend weiterentwickelte Ausbildungsmodul zur pädagogischen Diagnostik wird aus den Mikroartikeln sowohl der Stand des Aufbaus eines Wissensbestands (v.a. an deklarativem Wissen zum Bereich der pädagogischen Diagnostik) bei den Studierenden als auch die Wahrnehmung komplexer Zusammenhänge ersichtlich.

Vor allem in den Texten zu Impulsfrage 1 sind viele Elemente zu finden, die als explizit zu vermittelnde Inhalte für das Ausbildungsmodul festgelegt sind. So findet sich in vielen Texten pädagogische Diagnostik als Instrument der Lernbegleitung als auch als Instrument, Problemen auf den Grund zu gehen. Der Bezug zur (individuellen und gruppenbezogenen) Lernförderung wird hergestellt, die Wichtigkeit eines systematischen und theoriebasierten Vorgehens findet Ausdruck in den Texten.

In den Texten zu Frage 3, wo die persönliche Sicht der Studierenden stärker einzufließen scheint, ist im Kontext der konkreten LV ein interessantes Phänomen zu beobachten. Inhaltlich nehmen die Studierenden auch hier in vielen Fällen Bezug auf pädagogische Diagnostik im Sinne des formativen Assessments in der Lernbegleitung. Ohne direkte Bezugnahme auf die Fallbeispiele spielt hier jedoch pädagogische Diagnostik als systematischer Prozess zum Umgang mit Lernschwierigkeiten eine dominante Rolle. Wichtige Themen sind Objektivität, Kollaboration, systematisches Vorgehen.

5.3 Das Modell des formativen Assessments in den mentalen Modellen zur pädagogischen Diagnostik

Bezüge zum Modell des formativen Assessments nach Black und Wiliam (2009) finden sich vermehrt bei den Textteilen zu Impulsfrage 1, also jenem Teil, der vor allem vom erworbenen Wissen gespeist wird. Diese Bezüge sind eher auf der Ebene des Modells an sich anzusiedeln. Vor allem die Grundidee des formativen Assessments, Evidenzen über das Lernen zu generieren und diese zu verwenden, um Entscheidungen über das weitere Vorgehen im Lernen zu treffen, wird in vielen Studierendentexten angesprochen. Studierende beziehen sich auch explizit darauf, Diagnoseergebnisse zu verwenden, um Feedback zu geben und individuell abgestimmte Förderung anzubieten.

Bezogen auf die fünf Strategien des formativen Assessments (Wiliam, 2018) sind in den Texten zumindest indirekte Bezüge zu den Strategien 2 (Evidenzen über das Lernen generieren) und 3 (Feedback geben, das das Lernen voranbringt) zu finden. Strategie 5, Schüler*innen zu Verantwortlichen ihres Lernens zu machen, findet sich am ehesten in dem mehrfach geäußerten Gedanken, Schüler*innen eine aktive Rolle im Diagnoseprozess geben zu wollen, wieder. Interessanterweise wird letzteres nicht im Rahmen der deklarativen Wissensbestände angeführt, sondern ist ein relevantes Element in den persönlicheren Textteilen zu Frage 3. Die Strategien 1 (Lernziele und

Erfolgskriterien klären) und 4 (Peers als Ressource nutzen) werden von den Studierenden nicht angesprochen.

5.4 Kritische Betrachtung und Ausblick

Die Methode der Freewriting-Mikroartikel zeigt Potenzial für die Rekonstruktion mentaler Modelle. Über die stark verdichtete, beschreibende Darstellung in diesem Beitrag hinaus bedarf es jedoch noch weiterer Analysezugänge, mit denen die Komplexität der mentalen Landkarten abgebildet werden kann.

Das für die LV relevante Modell des formativen Assessments zeigt sich nur indirekt in den Themen, die aus den Daten extrahiert werden konnten. Das deutet darauf hin, dass es erforderlich ist, die Rolle des formativen Assessments stärker zu entwickeln und dies noch stärker im Curriculum und dessen Implementation zu verankern. Gleichzeitig ist als eine Limitation der verwendeten Vorgehensweise zu beachten, dass die Rekonstruktion mentaler Modelle zu einem spezifischen Zeitpunkt vorgenommen wurde. Der Vorteil der Erhebung gegen Ende der genannten Lehrveranstaltung liegt in der gerade erfolgten intensiven Beschäftigung mit Aufgaben und Herausforderungen der pädagogischen Diagnostik.

Jedenfalls bedarf es weiterer Forschung, um die Rolle der pädagogischen Diagnostik fassen zu können. Die vorliegenden Daten sollen zukünftig durch die Rekonstruktion mentaler Modelle anderer Personengruppen, etwa Studierender zu Beginn sowie am Ende ihres Lehramtsstudiums, aber auch (erfahrenen) Lehrpersonen ergänzt werden. Weitere Untersuchungen unter Einbeziehung des Feldes sind wichtig auf dem Weg zur Kanonisierung.

Literatur

- Barr, R.B. & Tagg, J. (1995). From Teaching to Learning. *Change*, 13–23.
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Bölling, R. (2010). *Kleine Geschichte des Abiturs*. Ferdinand Schöningh. <https://doi.org/10.30965/9783657769049>
- Braunsteiner, M.-L. & Spiel, C. (Hrsg.). (2019). *PädagogInnenbildung*. Be&Be.
- Bruner, J. (1960). *The process of education*. HUP.
- Buch, S. & Sparfeldt, J. (2020). Diagnostik, Beurteilung und Förderung als Gegenstand der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 39–46). Klinkhart. <https://doi.org/10.35468/hblb2020-003>
- Buholzer, A., Baer, M., Zulliger, S., Torchetti, L., Ruelmann, M., Häfliger, A. & Lötscher, H. (2020). Formatives Assessment im alltäglichen Mathematikunterricht von Primarlehrpersonen: Häufigkeit, Dauer und Qualität. *Unterrichtswissenschaft*, 48(4), 629–661. <https://doi.org/10.1007/s42010-020-00083-7>

- Craik, K. J. W. (1943). *The Nature of Explanation*. Cambridge University Press.
- Dann, H.-D. (1989). Subjektive Theorien als Basis erfolgreichen Handelns von Lehrkräften. *Beiträge zur Lehrerbildung*, 7(2), 247–254.
- Elbow, P. (1973). *Writing Without Teachers*. Oxford University Press.
- EU – European Commission/EACEA/Eurydice (2015). *The Teaching Profession in Europe: Practices, Perceptions, and Policies. Eurydice Report*. Publications Office of the European Union. <https://doi.org/10.2797/63946>
- Fend, H. (2006). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen*. VS.
- Gräsel, C. & Mandl, H. (1999). Problemorientiertes Lernen. *Empirische Pädagogik*, 13(4), 371–391.
- Gruschka, A. (2002). *Didaktik. Das Kreuz mit der Vermittlung*. Büchse der Pandora.
- Heinrichs, H. (2015). *Diagnostische Kompetenz von Mathematik-Lehramtsstudierenden*. Springer. <https://doi.org/10.1007/978-3-658-09890-2>
- Helmke, A. (2022). *Unterrichtsqualität und Professionalisierung. Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung*. Kallmeyer.
- Hericks, U. (2006). *Professionalisierung als Entwicklungsaufgabe. Rekonstruktionen zur Berufseingangsphase von Lehrerinnen und Lehrern*. VS.
- Hesse, I. & Latzko, B. (2017). *Diagnostik für Lehrkräfte*. Barbara Budrich/utb. <https://doi.org/10.36198/9783838547510>
- Holzkamp, K. (1995). *Lernen: Subjektwissenschaftliche Grundlegung*. Campus.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik*. Beltz.
- ISB – Staatsinstitut für Schulqualität und Bildungsforschung (Hrsg.) (2008). *Pädagogisch diagnostizieren im Schulalltag. Grundlageninformationen mit Anregungen für die Praxis*. München. http://www.isb.bayern.de/download/7409/paedagogisch_diagnostizieren.pdf
- Jansen, C. & Meyer, M. (2016). *Diagnostizieren im Dialog: Ein Leitfaden für den individualisierten Unterricht*. Beltz.
- Jensen R. (2006). Behaviorism, latent learning, and cognitive maps: needed revisions in introductory psychology textbooks. *The Behavior analyst*, 29(2), 187–209. <https://doi.org/10.1007/BF03392130>
- Jones, N. A., Ross, H., Lynam, T., Perez, P. & Leitch, A. (2011). Mental models: an interdisciplinary synthesis of theory and methods. *Ecology and Society*, 16(1), 46. <http://www.ecologyandsociety.org/vol16/iss1/art46/>
- Jürgens, E. & Lissmann, U. (2015). *Pädagogische Diagnostik: Grundlagen und Methoden der Leistungsbeurteilung in der Schule*. Beltz.
- Kraler, C. (2007). Beurteilen und-oder begleiten? Bedingungen, Perspektiven und Grenzen einer portfoliogestützten LehrerInnenausbildung. *Seminar – Lehrerbildung und Schule*, 1, 75–102.
- Kraler, C., Dittrich, A. K. & MacKay-Falls, F. (2017). Ausbildungsreform als Chance für einen Musterwechsel im Verständnis pädagogisch-psychologischen Professionswissens. *Zeitschrift für Bildungsforschung*, 7, 193–207. <https://doi.org/10.1007/s35834-017-0198-z>
- Kraler, C., Mayr, K. & Schratz, M. (2006). *MikroArtikel in der LehrerInnenaus- und Weiterbildung*. *Journal für LehrerInnenbildung*, 3, 31–39.
- Kraler, C. & Schratz, M. (2012). From best practice to next practice: a shift through research-based teacher education. *Reflecting Education*, 8(2), 88–125.
- Kraler, C. & Worek, D. (Hrsg.). (2021). *Teacher Education. The Bologna Process and the Future of Teaching*. Waxmann.
- Oelkers, J. (2009). Fächerkanon und Fachunterricht. In S. Blömeke, T. Bohl, L. Haag, G. Lang-Wojtasik & W. Sacher (Hrsg.), *Handbuch Schule* (S. 305–313). Klinkhardt.

- Peirce, C. S. (1931). *The Collected Papers of Charles Sanders Peirce, Vol. I: The Principles of Philosophy*. CP 1. Edited by C. Hartshorne & P. Weiss. Harvard University Press.
- Rohlf, C. (2011). *Bildungseinstellungen, Schule und formale Bildung aus der Perspektive von Schülerinnen und Schülern*. VS Verlag für Sozialwissenschaften.
- Schreiner, C. & Kraler, C. (2019). Pädagogische Diagnostik und LehrerInnenbildung nach PISA. *Erziehung und Unterricht, 9/10*, 861–871.
- SQA – Schulqualität Allgemeinbildung (2015). *Pädagogische Diagnostik*. BMBF (Bundesministerium für Bildung und Frauen).
- Terhart, E. (2013). *Erziehungswissenschaft und Lehrerbildung*. Waxmann.
- Tolman, E.C. (1948). Cognitive maps in animals and man. *Psychological Review, 55*, 189–208. <https://doi.org/10.1037/h0061626>
- van Ophuysen, S. (2010). Professionelle pädagogisch-diagnostische Kompetenz – eine theoretische und empirische Annäherung. In N. Berkemeyer, W. Bos, H.G. Holtappels, N. McElvany & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Bd. 16, S. 203–234). Juventa.
- Wagenschein, M. (1968). *Verstehen lehren. Genetisch – Sokratisch – Exemplarisch*. Beltz.
- Wiggins, G. & McTighe, J. (2005). *Understanding by design*. ASCD.
- William, D. (2018). *Embedded formative assessment*. Solution Tree.
- Willke, H. (2009). Der MikroArtikel als Instrument des Wissensmanagements. In S. Rietmann & G. Hensen (Hrsg.), *Werkstattbuch Familienzentrum* (S. 97–108). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91640-8_8
- Willke, H. (2018). *Einführung in das systemische Wissensmanagement* (4. Aufl.). Carl-Auer.
- Wittgenstein, L. (2019/1922). *Tractatus logico-philosophicus*. stw.
- Zgaga, P. (2015). Rethinking Teacher Education in Europe and beyond. Editorial introduction. *Education Inquiry, 6*(3). <https://doi.org/10.3402/edui.v6.29354>

Einstellungen pädagogischer Fachkräfte in der KiTa zur Bedeutung von formativem Assessment im Kontext der Naturwissenschaften

Laura Venitz¹ | Ilonca Hardy² | Mirjam Steffensky³ | Anika Bürgermeister⁴ | Katharina Junge³ | Henrik Saalbach⁴ | Miriam Leuchter¹

¹ Rheinland-Pfälzische Technische Universität Kaiserslautern Landau |

² Goethe-Universität Frankfurt a.M. | ³Universität Hamburg | ⁴Universität Leipzig

Abstract: Pädagogische Fachkräfte (PFK) benötigen hohe professionelle Kompetenzen, um kindliche Lern- und Entwicklungsprozesse zu diagnostizieren und auf dieser Basis Entscheidungen über passende Schritte im Bildungsprozess zu treffen. Im Beitrag werden Daten einer Teilstichprobe des Projekts FinK¹ berichtet, bei der anhand von Online-Befragungen und Interviews untersucht wurde, a) welche Überzeugungen PFK gegenüber der förderbezogenen Diagnostik aufweisen, b) inwiefern diese mit ihren Selbstwirksamkeitsüberzeugungen zu Diagnostik und Förderung zusammenhängen und sich c) in Praktiken der frühen naturwissenschaftlichen Bildung widerspiegeln. Die Ergebnisse zeigen, dass diagnostische Maßnahmen im KiTa-Alltag nur wenig genutzt werden, dass PFK aber hohe Selbstwirksamkeitsüberzeugungen sowie positive Einstellungen zu Naturwissenschaften in der KiTa aufweisen. Die eingeschränkte Bedeutung der prozessbegleitenden Diagnostik zeigt sich auch in Aussagen zu Selbstbildungsprozessen mit einer eher passiven Rolle der PFK.

1 Einführung

Formatives Assessment wird als die wiederholte Diagnose von kindlichen Voraussetzungen und Verläufen von Lern- und Entwicklungsprozessen beschrieben, auf deren Basis Entscheidungen über passende (Förder-)Schritte im Bildungsprozess getroffen werden (Schütze et al., 2018). Diesem Prozess wird eine zentrale Bedeutung in Lehr-Lern-Situationen zugeschrieben (Kingston & Nash, 2011). Formatives Assessment spielt auch im Elementarbereich eine wichtige Rolle, obwohl in Kindertagesstätten (KiTas) die Bedingungen der Organisation und Teilnahme an Bildungsgelegenheiten andere sind als in der Schule. So sind zwar domänenspezifische Bildungsbereiche wie Naturwissenschaften, die in diesem Beitrag im Fokus stehen, in den Orientierungsplänen der deutschen Bundesländer verankert, konkrete Zielsetzungen in Bezug auf angestrebte kindliche Kompetenzen fehlen jedoch häufig oder sind unspezifisch formuliert (Steffensky, 2017). Auch die professionellen Kompetenzen pädagogischer Fachkräfte (PFK)

1 Projekt FinK: Formatives Assessment in der frühen inklusiven Bildung in der KiTa, <https://www.ew.uni-hamburg.de/einrichtungen/ew5/didaktik-der-chemie/projekte/fink.html>

sind je nach Ausbildungshintergrund unterschiedlich und insbesondere fachspezifische Kompetenzen, welche für diagnostische Prozesse grundlegend sind, finden sich in eher geringer Ausprägung (Dunekacke & Barenthien, 2021). Vor diesem Hintergrund stellen die wiederholte Diagnostik und die darauf abgestimmten Entscheidungen im Sinne des formativen Assessments oft eine Herausforderung für PFK dar (Beher & Walter, 2012).

Im Mittelpunkt dieses Beitrags stehen Einstellungen von PFK zur Diagnostik und Förderung im Kontext der Naturwissenschaften. Es wird zunächst kurz auf Ansätze des formativen Assessments sowie auf frühe naturwissenschaftliche Bildungsprozesse eingegangen, bevor Einstellungen von PFK skizziert werden. Nach der Einleitung werden Daten einer Interview- und Fragebogenstudie mit PFK dargestellt und diskutiert, in der u.a. Einstellungen und Selbstwirksamkeitserwartungen zur Diagnostik und Förderung in der inklusiven naturwissenschaftlichen Bildung in der KiTa erhoben wurden.

2 Überzeugungen zu naturwissenschaftlicher Bildung von pädagogischen Fachkräften

2.1 Formatives Assessment

Die unterschiedlichen Voraussetzungen von Kindern zeigen sich z. B. in Lern- und Entwicklungsständen der Sprachkompetenz, Selbstregulation oder sozialen Kompetenz (Skopek & Passaretta, 2021). PFK in der KiTa haben die anspruchsvolle Aufgabe, diese unterschiedlichen Voraussetzungen und spezifischen Förderbedarfe von Kindern angemessen zu berücksichtigen und entsprechend angepasste Bildungsangebote bereitzustellen (Bartusch et al., 2019). So werden auch im Kontext inklusiver Bildung insbesondere die prozessbegleitende Diagnostik und abgestimmte Förderung hervorgehoben (Schmidt & Liebers, 2017). Im Ansatz des formativen Assessment finden sich Diagnostik und individuelle Förderung in den Merkmalen der Klärung von Lernzielen, der wiederholten Einschätzung von Lernausgangslagen und Lernständen, der individuellen Rückmeldungen zu Lernständen und der entsprechenden Anpassung von Bildungsangeboten wieder (Bürgermeister & Saalbach, 2018; Schildkamp et al., 2020; Schütze et al., 2018). Darüber hinaus wird die Einbindung von Lernenden in die Einschätzung von Lernständen, z. B. über Selbst- und Peer-Assessments, fokussiert (Sanchez et al., 2017). Grundsätzlich kann zwischen einer prozessbezogenen Diagnostik, die häufig als Beobachten und Erfassen von Lernständen in relevanten Interaktionen („on the fly“) bezeichnet wird (Shavelson et al., 2008) und einer stärker formalisierten Diagnostik, die beispielsweise durch den Einsatz von Einschätzskalen umgesetzt wird, unterschieden werden. Jenseits dieser Ausprägungen einer prozessbezogenen Diagnostik findet sich die Diagnostik von spezifischen kindlichen Förderbedarfen, welche – bis zur Feststellung eines Inklusions- oder Integrationsstatus – i.d.R. über standardisierte Verfahren unter

Beachtung gesetzlicher Vorgaben zum Zusammenwirken unterschiedlicher professioneller Berufsgruppen umgesetzt wird.

Prinzipiell bieten sich im KiTa-Alltag zur Umsetzung einer prozessbegleitenden Diagnostik eine Vielzahl von Anlässen, z. B. die Beobachtung von Kindern in Spiel- und Routinesituationen oder das gezielte Fragen, z. B. nach kindlichen Vorstellungen. In deutschen KiTas gibt es keine einheitliche Praxis der Bildungsdokumentation. So werden verschiedene Instrumente zur Beobachtung und Dokumentation kindlicher Grundfähigkeiten, z. B. Grob- und Feinmotorik, Kognition, Sprache oder sozial-emotionale Entwicklung, eingesetzt, die sich im Grad der Strukturierung unterscheiden und verschiedene Funktionen haben: Dokumentation, Erkennen und Screening von Entwicklungsrisiken und Ableitung von Fördermaßnahmen. Beispiele sind KOMPIK (Bauer et al., 2012) oder Kuno Bellers Entwicklungstabelle (Beller, 2016). Das Instrumentarium ILEA-T ermöglicht die Diagnose von individuellen Lernvoraussetzungen und Entwicklungsständen von Kindern am Übergang zur Grundschule (Geiling & Berger, 2015) mit der Intention der Ableitung von adaptiven Lernarrangements unter Nutzung von empirisch überprüften domänenspezifischen Stufenmodellen. Auch der Portfolio-Ansatz (vgl. Hooker, 2019) lässt sich mit den Zielen der prozessorientierten Diagnostik verbinden, sofern Entwicklungsstände dokumentiert und als Grundlage für pädagogische Entscheidungen genutzt werden.

Gemäß dem Ansatz des formativen Assessments sollen aus der Diagnostik gewonnene Einschätzungen die Grundlage für eine adaptive Gestaltung von Bildungsgelegenheiten bieten. Diese können auf der Makro- und Mikroebene umgesetzt werden (Corno, 2008). Auf der Makroebene sind langfristige und übergeordnete Anpassungen von Bildungs- und Entwicklungsangeboten gemeint, wie die Bereitstellung bestimmter Materialien. Auf der Mikroebene finden kurzfristige Anpassungen, oft im Rahmen von Fachkraft-Kind-Interaktionen statt, die Ansätzen der kognitiven Unterstützung (Kleickmann et al., 2020) oder dem Contingent Support bzw. Scaffolding nach van de Pol et al. (2015) nahestehen.

Die Wirksamkeit von formativem Assessment konnte in verschiedenen Studien gezeigt werden (Hattie, 2009; Lee et al., 2020). Studien für den vorschulischen Bereich, insbesondere für die Naturwissenschaften, finden sich allerdings deutlich seltener (Bruns, 2014). Auch im Zusammenhang mit den diagnostischen Instrumenten für den Elementarbereich lässt sich feststellen, dass Beobachtungssitems vorhandener Instrumente in der Regel auf einer sehr allgemeinen Ebene angesiedelt sind, z. B. „Das Kind überlegt bei Aufgaben“, „Das Kind hat unterschiedliche Lösungsideen“ (aus KOMPIK). Dies ist sinnvoll, um generelle kindliche Entwicklungsstände zu erfassen, aber nicht ausreichend, um in einem Bildungsangebot, das auf einen bestimmten Lerngegenstand ausgerichtet ist, den kindlichen Lernstand einzuschätzen und Fördermaßnahmen abzuleiten. Hier fehlen bisher empirisch validierte Instrumente bzw. Aufgaben- und Fragepools, die curricular eingebettet werden können.

2.2 Naturwissenschaftliche Bildung in der Kita

Es gibt einen breiten Konsens darüber, dass die Naturwissenschaften ein wichtiger Bildungsbereich für die KiTa sind, der in erster Linie auf die Entwicklung von grundlegenden Erfahrungen und Wissen sowie Motivation, Neugierde und Interesse abzielt (Anders et al., 2013). Gleichzeitig fehlt ein Konsens zu spezifischen Inhalten, die adressiert werden sollten, sowie zur Tiefe, in welcher (grundlegendes) Wissen angebahnt werden sollte. So werden in einigen Orientierungsplänen lediglich Erfahrungen (z. B. mit Matsch spielen) und in anderen zusätzlich spezifischere und umfassendere Inhalte benannt (z. B. Stoffumwandlungen kennenlernen) (Steffensky, 2017). Aus einer fachdidaktischen Perspektive sind beide Erfahrungsräume sinnvoll, da Kinder anschlussfähiges Wissen aus basalen Erfahrungen aufbauen, die nach und nach in tragfähige Konzepte ausdifferenziert werden. Allerdings ist es bei der Auswahl von Inhalten und Aktivitäten wichtig, den genauen Lerngegenstand zu spezifizieren und dabei zu beachten, dass das anzubahnde Wissen anschlussfähig sowohl an bereits vorhandene Erfahrungen ist als auch an Konzepte, die im weiteren (schulischen) Lernen bedeutsam sind.

Neben den inhaltlichen Schwerpunkten ist zudem unklar, inwieweit PFK geplante Bildungsangebote umsetzen, da es in Deutschland eine weit verbreitete Orientierung an Selbstbildungsansätzen gibt (Schmidt & Smidt, 2021). Diese pädagogischen Ansätze lehnen die gezielte Umsetzung von Bildungsangeboten mit der Intention, spezifische Kompetenzen zu fördern, ab. Bildungs- oder Weltaneignungsprozesse sind stattdessen vor allem durch Eigenaktivität des Kindes zu erreichen. Dass Bildungsprozesse aktive Vorgänge sind, die Selbsttätigkeit erfordern, wird kaum bestritten. Problematisch ist die angenommene Rolle der PFK, die stark auf das Beobachten der Kinder, aber nicht auf ein Unterstützen ausgerichtet ist. Ein weiterer pädagogischer Ansatz ist die Ko-Konstruktion. Bildung wird hier als sozialer Prozess angenommen, in welchem in gemeinsamen Interaktionen von PFK und Kind Basiskompetenzen von Kindern angebahnt werden (Fthenakis, 2009). Die Orientierung an reiner Instruktion ist im KiTa-Bereich eher selten und findet sich vor allem in gezielten Förderprogrammen (Krajewski et al., 2007). Welche Bildungsgelegenheiten Kinder in ihrer KiTa-Zeit erfahren, hängt entsprechend nicht nur vom Fachwissen und fachdidaktischen Wissen der PFK ab, sondern auch von ihren Einstellungen im Zusammenhang mit früher naturwissenschaftlicher Bildung.

2.3 Einstellungen und Selbstwirksamkeitserwartungen

Unter pädagogischen Einstellungen werden eine Bandbreite verschiedener subjektiver Annahmen, Vorstellungen, Orientierungen und Überzeugungen über bildungsbezogene Sachverhalte und Prozesse, z. B. Werte und Ziele der frühpädagogischen Arbeit oder Einschätzungen zur Bedeutung verschiedener Förderbereiche verstanden (Tietze, 2004). Sie enthalten in der Regel bewertende Anteile und können explizit oder implizit sein. Vor allem gelten sie als wichtige Facette professioneller Kompetenz (Fröhlich-

Gildhoff et al., 2011). Sie werden als WahrnehmungsfILTER beschrieben (Fives & Buehl, 2012) und gelten auf der Basis der Theorie des geplanten Verhaltens (Ajzen, 1985) als leitend für das Handeln. Ob und wie also (bereichsspezifische) Bildungsgelegenheiten von PFK umgesetzt werden, hängt demnach wesentlich von ihren Einstellungen ab. Entsprechende Zusammenhänge zwischen Einstellungen und Handeln konnten zum Beispiel im Kontext der frühen mathematischen Bildung gezeigt werden (Anders & Rossbach, 2015; Benz, 2012).

Tendenziell schätzen PFK naturwissenschaftliche Bildung als wichtigen Bildungsbereich ein (Drexl et al., 2019) und aktuelle Befragungen deuten zudem auf positiv ausgeprägte Selbstwirksamkeitserwartungen in Bezug auf die Umsetzung naturwissenschaftlicher Bildung hin (Steffensky et al., 2018). Im Hinblick auf die Diagnostik in den Bereichen Sprache, Mathematik und sozial-emotionale Entwicklung schätzen PFK ihre Kompetenzen eher hoch ein (Höltge et al., 2019). Selbstwirksamkeitsüberzeugungen speziell in Bezug auf die förderbezogene Diagnostik im Kontext naturwissenschaftlicher Bildung sind allerdings bislang nicht untersucht.

Vor diesem Hintergrund gehen wir in der nachfolgenden explorativen Studie der Frage nach, welche Einstellungen zur frühen naturwissenschaftlichen Bildung, zur eigenen Rolle bei der Lernbegleitung und zur förderbezogenen Diagnostik PFK aufweisen. Zudem untersuchen wir, wie hilfreich sie die Nutzung eines begleitenden Portfolios, welches die Umsetzung einer prozessbezogenen Diagnostik und die Planung von individueller Förderung unterstützen soll, einschätzen (vgl. Hardy et al., im Druck). Ausgehend davon, dass Einstellungen handlungsleitend sind, kann man annehmen, dass dies in besonderem Maß im KiTa-Kontext gilt, da PFK (im Vergleich zur Schule) deutlich größere Entscheidungsfreiräume haben, bestimmte Bildungsbereiche auf spezifische Weise zu fördern. Zudem ist davon auszugehen, dass Einstellungen zur förderbezogenen Diagnostik eine besondere Bedeutung haben, diese möglicherweise wenig kompatibel mit Ansätzen der Selbstbildung sind.

3 Studie

3.1 Design und Stichprobe der Studie

An der Professionalisierungsstudie FinK (Formatives Assessment in der inklusiven naturwissenschaftlichen Bildung in der KiTa) nahmen insgesamt 54 PFK aus 40 verschiedenen Kindertageseinrichtungen aus dem Raum Landau und Leipzig teil. Die teilnehmenden Fachkräfte wurden in drei quasi-experimentelle Gruppen (zwei Interventionsgruppen und eine Vergleichsgruppe) eingeteilt. Beide Interventionsgruppen erhielten eine Weiterbildung zur Umsetzung eines naturwissenschaftlichen Bildungsangebotes zum Thema Schwimmen und Sinken sowie begleitende Materialien. Eine der Weiterbildungsgruppen erhielt zusätzlich eine Weiterbildung zur Prozessdiagnostik und adaptiven Unterstützung von Kindern mit unterschiedlichen Voraussetzungen

sowie ein inhaltlich daran ausgerichtetes Portfolio. Ein Überblick zu den verwendeten Materialien findet sich bei Hardy et al. (im Druck) und Hardy et al. (2021). Die Wirksamkeit der Professionalisierungsmaßnahmen wurde auf Ebene der PFK (professionelle Kompetenzen), auf Ebene der teilnehmenden Kinder (Kompetenzentwicklung in den Bereichen Sprache, Selbstregulation, naturwissenschaftliches Wissen und Interesse an Naturwissenschaften) sowie auf Ebene der PFK-Kind-Interaktion (Interaktionsqualität) im Prä-Post-Vergleich überprüft. Zur Untersuchung von Effekten der Professionalisierungsmaßnahmen in Bezug auf die professionellen Kompetenzen der teilnehmenden PFK fanden zu zwei Messzeitpunkten Online-Befragungen statt. In diesem Rahmen wurden die PFK zu grundlegenden Überzeugungen zur inklusiven naturwissenschaftlichen Bildung in der KiTa, zu motivationalen Orientierungen (z. B. Selbstwirksamkeit) sowie zu ihrem fachdidaktischen Wissen und ihrem Fachwissen befragt. 20 PFK der Gesamtstichprobe erklärten sich außerdem dazu bereit, nach der zweiten Befragung zusätzlich vertiefende Leitfadeninterviews mit einer der Projektleitungen zu führen. Die Interviews fanden telefonisch statt. Die Audioaufnahmen wurden transkribiert. Die Interviews dauerten in der Regel 25 Minuten. Die PFK der Interview-Teilstichprobe waren durchschnittlich 38,95 Jahre ($SD = 10.53$) alt, 16 der 20 PFK waren weiblich. Sie wiesen durchschnittlich eine 13-jährige Tätigkeit in Kindertagesstätten auf.

Um Informationen zur Repräsentativität der Teilstichprobe in Bezug auf die Gesamtstichprobe zu erhalten, wurden die interviewten PFK im Hinblick auf zentrale Merkmale der professionellen Kompetenz mittels Mann-Whitney-U-Tests (ein nicht-parametrisches Verfahren wurde aufgrund der geringen Stichprobengrößen gewählt) mit der Gesamtstichprobe verglichen. Dabei ließen sich keine signifikanten Unterschiede in Bezug auf die Einstellungen zum Lehren und Lernen (*Selbstbildung*: $z = -1.44$, $p = .15$; *Instruktion*: $z = -2.21$, $p = .07$; *Ko-Konstruktion*: $z = -.88$, $p = .41$), den Enthusiasmus für Naturwissenschaften ($z = -.24$, $p = .81$), den Selbstwirksamkeitsüberzeugungen zu Naturwissenschaften in der KiTa ($z = -.53$, $p = .60$) und die Selbstwirksamkeit zum adaptiven professionellen Handeln in der KiTa (*Gesamt*: $z = -.37$, $p = .73$) zum zweiten Messzeitpunkt nach Beendigung der Professionalisierungsmaßnahmen feststellen. Im Detail ergab sich, dass die Teilstichprobe ähnlich hohe Werte in Bezug auf die Einstellungen *Selbstbildung* und *Ko-Konstruktion* und niedrige Werte in Bezug auf die *Instruktion* aufwies wie die Gesamtstichprobe (siehe Tabelle 1). Auch in Bezug auf den *Enthusiasmus* und die *Selbstwirksamkeit* zur Vermittlung von Naturwissenschaften in der KiTa sowie hinsichtlich der *Selbstwirksamkeit zum adaptiven Handeln* schätzten sich die PFK hoch ein. Ausprägungen innerhalb der Teilstichprobe hinsichtlich der naturwissenschaftlichen Bildungsarbeit sowie der Bedeutung der prozessbezogenen Diagnostik nachdem das naturwissenschaftliche Bildungsangebot umgesetzt wurde, werden im Ergebnisteil berichtet.

Tabelle 1: Mittelwert (M), Standardabweichung (SD) und Reliabilität (Cronbach's α) für Einstellungen zum Lehren und Lernen, Enthusiasmus und Selbstwirksamkeit in Bezug auf die Vermittlung von Naturwissenschaften in der KiTa (Steffensky, 2018) und Selbstwirksamkeit in Bezug auf adaptives Handeln (Meschede & Hardy, 2020) der Interview-Teilstichprobe zum zweiten Messzeitpunkt.

Skala	Beispielitem	M (SD)	Cronbach's α
Selbstbildung	Kinder lernen am besten, wenn sie sich ganz selbstbestimmt mit naturwissenschaftlichen Inhalten beschäftigen.	3.25 (0.33)	.69
Instruktion	Am besten verstehen Kinder naturwissenschaftliche Inhalte aus den Erklärungen und Darstellungen der Pädagogischen Fachkraft.	1.90 (0.43)	.82
Ko-Konstruktion	Es hilft Kindern, wenn sie im Alltag immer wieder zum Nachdenken über naturwissenschaftliche Inhalte angeregt werden.	3.26 (0.51)	.60
Enthusiasmus Naturwissenschaften	Es macht mir Spaß, Kindern Naturphänomene im Alltag nahezubringen.	3.50 (0.41)	.71
Selbstwirksamkeit Naturwissenschaften	Ich traue mir zu, naturwissenschaftsbezogene Themen so zu vermitteln, dass die Kinder die Inhalte verstehen können.	2.98 (0.54)	.76
Selbstwirksamkeit adaptives Handeln	Ich bin in der Lage, Beobachtungsverfahren einzusetzen, um Hinweise auf individuelle Lernentwicklungen zu bekommen.	3.11 (0.33)	.92

4-stufige Skala: 1= stimme gar nicht zu, 2= stimme eher nicht zu, 3= stimme eher zu, 4= stimme voll zu

3.2 Methodisches Vorgehen

Die Transkripte wurden anhand der Leitfragen zu den Themenblöcken a) Naturwissenschaftliche Bildungsarbeit, b) Diagnostik und Förderung in der KiTa und c) Umgang mit einem prozessbezogenen Portfolio in Einzelaussagen eingeteilt. Es wurde ein inhaltsanalytisches Vorgehen gewählt, in welchem zunächst theoriebezogen Kategorien

gebildet wurden und in einem iterativen Verfahren im Abgleich mit den Aussagen der PFK weitere Kategorien pro Leitfrage generiert wurden. Die Einzelaussagen jeder PFK wurden im letzten Schritt den gebildeten Kategorien zugeordnet, wobei pro Leitfrage und Person mindestens eine Kategorie, in Fällen mit unterschiedlichen Aussagen zum Themenblock jeweils mehrere Kategorien, kodiert wurden. Dies hat den Vorteil, die Gesamtheit der getroffenen Aussagen abzubilden und Personen mit unterschiedlichen eingenommenen Perspektiven auf den Gegenstand in den entsprechenden Häufigkeiten abzubilden. Die kodierten Aussagen wurden von zwei Raterinnen eingeschätzt und kommunikativ validiert.

3.3 Ergebnisse

3.3.1 Welches Verständnis von gelungener naturwissenschaftlicher Bildung zeigen die PFK?

Das naturwissenschaftliche Bildungsverständnis wurde in die Kategorien *Selbstbildung*, *Ko-Konstruktion* und *Instruktion* unterschieden. Es zeigte sich, dass die Mehrheit der PFK naturwissenschaftliche Bildungsprozesse in der Kindheit als Selbstbildung verstehen, welche kindinitiiert ist und insbesondere auf dem kindlichen Interesse beruht, sich mit spezifischen Phänomenen näher auseinanderzusetzen ($N = 18$). Dies zeigt sich beispielsweise in Äußerungen wie [Gelungene naturwissenschaftliche Bildung für mich ist,] „wenn's ohne Druck und aus dem Interesse der Kinder passiert“ (Transkript 1121), oder „wenn wirklich die Kinder das gerne mitmachen und viel nachfragen“ (Transkript 1111). Ein Zusammenhang zwischen naturwissenschaftlicher Bildung und dem professionellen Handeln von PFK, dem Aufrechterhalten von kindlichem Interesse, dem Aufgreifen von kindlichen Vorstellungen und Konzepten und der Rolle der PFK bei der Aufbereitung von Bildungsdomänen wurde in den wenigsten Fällen konstatiert. Ein ko-konstruktivistisches Verständnis konnte nur in einer Aussage (Transkript 2161) festgestellt werden. Trotz dieser untergeordneten Rolle, die sich die PFK in den Bildungsprozessen zuschreiben, berichten die meisten PFK von Bildungsaktivitäten im Bereich der Naturwissenschaften. Diese finden häufig anlassbezogen (z. B. durch Jahreszeiten, Waldtage) statt und werden mit unterschiedlicher Intensität und Dauer umgesetzt. Während wenige PFK von Ressourcen wie Experimentierrechen berichten ($N = 3$), rangieren der Hauptteil der Aktivitäten zwischen gelegentlichen Ausflügen in die Natur oder Aktivitäten auf dem Außengelände (z. B. Spielen mit Wasser oder Grünprojekte im Garten). Somit wird deutlich, dass naturwissenschaftliche Bildung insgesamt als kindlich gesteuert und anlassbezogen konzipiert und umgesetzt wird. Eine Referenz zu Orientierungsplänen, Kompetenzentwicklungen und spezifischen kindlichen Voraussetzungen findet sich in den Interviews nicht. Somit ist fraglich, ob eine Grundlage für den Einsatz von diagnostischen Verfahren und informellen In-situ-Beobachtungen im Sinne des Formativen Assessments als Grundlage für individuelle Förderentscheidungen gegeben ist.

3.3.2 Wie schätzen sich PFK hinsichtlich der förderbezogenen Diagnostik ein und welche Bedeutung hat diese in der Kita-Arbeit?

Die PFK schätzten sich in Bezug auf ihre Selbstwirksamkeit bei der Umsetzung von alltagsbezogener Diagnostik und Förderung in einem adaptierten schriftlichen Fragebogen auf der Grundlage von Meschede & Hardy (2020) im Mittel hoch ein. Ähnlich wie in den vorausgehenden Analysen ergaben sich zum zweiten Messzeitpunkt keine signifikanten Unterschiede zwischen der Teilstichprobe und der Gesamtstichprobe (Selbstwirksamkeit Diagnostik: $z = -.52$, $p = .60$; Selbstwirksamkeit Förderung: $z = -.47$, $p = .64$). So gaben sie an, sich zuzutrauen, individuelle Lernentwicklungen mit unterschiedlichen Methoden festzustellen, Lernentwicklungen prozessbezogen zu erfassen, Gelegenheiten zur Feststellung von Förderbedarfen zu schaffen und Kinder in heterogenen Lerngruppen individuell zu fördern (siehe Tabelle 2).

Tabelle 2: Mittelwert und Standardabweichung für die Items der Förderorientierungen im Vergleich zwischen Interviewstichprobe und Gesamtstichprobe

Skala	Item	<i>M (SD)</i> Interview- stichprobe	<i>M (SD)</i> Gesamt- stichprobe
Selbstwirksamkeit Diagnostik	D_1 Ich traue mir zu, individuelle Lernentwicklungen von Kindern mit unterschiedlichen Methoden festzustellen.	3.00 (0.60)	3.06 (.68)
	D_2 Ich traue mir zu, Angebote auf Grundlage von den festgestellten Förderbedarfen individuell auf das Kind anzupassen.	2.92 (0.67)	3.19 (0.75)
	D_3 Ich fühle mich in der Lage, passende Verfahren für die Feststellung von Förderbedarfen in unterschiedlichen Bereichen auszuwählen.	2.97 (0.63)	3.00 (0.82)
	D_4 Ich traue mir zu, individuelle Lernentwicklungen von Kindern prozessbegleitend zu erfassen.	3.25 (0.45)	3.44 (0.63)
	D_5 Ich bin in der Lage, Schwierigkeiten der Kinder als Ansatzpunkt für die Feststellung von Förderbedarfen zu nutzen.	3.25 (0.48)	2.94 (0.68)

	D_6 Ich traue mir nicht zu, Kinder in die Feststellung von Förderbedarfen einzubeziehen.	2.91 (0.83)	2.94 (0.98)
	D_7 Ich fühle mich nicht dazu in der Lage, auch während eines Angebots immer wieder Gelegenheiten zur Feststellung von Förderbedarfen zu schaffen.	3.00 (0.61)	3.06 (0.93)
	D_8 Ich bin in der Lage, Beobachtungsverfahren einzusetzen, um Hinweise auf individuelle Lernentwicklungen zu bekommen.	3.35 (0.62)	3.31 (0.70)
Selbstwirksamkeit Förderung	F_1 Ich fühle mich in der Lage, Angebote zu gestalten, mit denen selbst unterschiedlichste Ausgangslagen der Kinder berücksichtigt werden.	2.92 (0.67)	3.06 (0.68)
	F_2 Ich bin in der Lage, auch in einer Kindergruppe mit sehr unterschiedlichen Voraussetzungen die Kinder individuell zu unterstützen.	3.08 (0.67)	3.19 (0.54)
	F_3 Ich bin zuversichtlich, dass ich Kindern mit unterschiedlichen Förderbedarfen selbstständiges Lernen auf ihrem Niveau ermöglichen kann.	3.08 (0.52)	3.06 (0.68)
	F_4 Ich traue mir zu, Angebote so umzusetzen, dass sie Erfolgserlebnisse für alle Kinder ermöglichen.	3.17 (0.58)	3.31 (0.60)
	F_5 Ich fühle mich im Stande, während der Umsetzung von Bildungsangeboten begründete Entscheidungen zur Unterstützung des Lernens einzelner Kinder zu treffen.	3.16 (0.56)	3.19 (0.75)
	F_6 Ich bin sicher, die unterschiedlichen Voraussetzungen der Kinder während eines Angebots produktiv nutzen zu können.	3.25 (0.45)	3.25 (0.58)

F_7 Ich fühle mich im Stande, unterschiedliche Angebote für Kinder mit wenig und mit viel Vorwissen zu erstellen.	3.33 (0.49)	3.31 (0.60)
F_8 Ich traue mir zu, Kinder mit sehr unterschiedlichen Voraussetzungen während eines gemeinsamen Angebots individuell zu fördern.	2.75 (0.62)	3.00 (0.82)

4-stufige Skala: 1= stimme gar nicht zu, 2= stimme eher nicht zu, 3= stimme eher zu, 4= stimme voll zu

Werden die PFK jedoch im Detail hinsichtlich der Umsetzung von diagnostischen Aktivitäten in der KiTa befragt, zeigt sich ein abweichendes Bild. Zwar wird die Diagnostik häufig als ein Beobachten von Kindern im KiTa-Alltag beschrieben (13 Nennungen), dieses findet jedoch meist anlassbezogen, und insbesondere bei spezifischen Auffälligkeiten wie Entwicklungsverzögerungen oder zu festgesetzten Zeitpunkten wie Elterngesprächen statt. Zudem verweisen die PFK auf den Einsatz von KiTa-spezifischen teil-standardisierten Verfahren (z. B. Entwicklungsschnecke (Schlaaf-Kirschner, 2014); Entwicklungstabelle (Beller, 2016)), welche als Grundlage für die Einordnung von eigenen Beobachtungen bzw. die Fokussierung auf bestimmte Entwicklungsfelder (sprachlich, kognitiv, motorisch etc.) herangezogen werden (14 Nennungen). Keine der interviewten Personen brachte die Diagnostik im formalen Sinne oder das Beobachten im Sinne einer prozessbezogenen Diagnostik mit dem Kontext der naturwissenschaftlichen Bildung in Verbindung. In Bezug auf die Frage, inwieweit sich die PFK zuständig für die diagnostischen Aktivitäten in der KiTa fühlen, zeigt sich ein differenziertes Bild. Ein Teil der PFK sehen diagnostische Tätigkeiten als wichtige Aufgabe ihres Professionsfeldes an (12 Nennungen). Die PFK verweisen zudem auf Zuständigkeiten außerhalb der KiTa wie Ärzte und Ärztinnen, Förderkräfte oder Institutionen für den Bereich der Diagnostik (15 Nennungen), welche unterschiedlich konnotiert sind. Während einige PFK hier einen Mehrwert in Bezug auf das Aufgreifen unterschiedlicher Expertisen u.a. für Austausch und Kooperation bei spezifischen Förderbedarfen sehen (8 Nennungen), verweisen andere PFK auf die eigene zeitliche Überlastung und die wünschenswerte Abgabe von Verantwortung an Außenstehende (7 Nennungen). Sie indizieren damit, dass die Diagnostik eher als belastender und wenig ertragreicher Bereich des professionellen Handelns erlebt wird und in keinerlei Zusammenhang mit der Gestaltung von Bildungsgelegenheiten steht. Insgesamt wird somit deutlich, dass die formative Diagnostik als Teil einer (naturwissenschaftlichen) Bildungsarbeit bisher wenig bis keine Bedeutung hat, sondern eine punktuelle Diagnostik in Einzelfällen bzw. zu spezifischen Anlässen in unterschiedlichem Ausmaß der professionellen Zusammenarbeit und der Nutzung diagnostischer Verfahren umgesetzt wird.

Beispielhaft zeigt sich in den folgenden Äußerungen die Bandbreite der umgesetzten Praxis in Bezug auf (prozessbezogene) Diagnostik. Im Zusammenspiel mit den hohen Selbstwirksamkeitserwartungen der PFK ist zu vermuten, dass einige PFK die von ihnen beschriebenen Beobachtungen hinsichtlich ihrer Zuverlässigkeit und Validität überschätzen.

„Ja, wir haben auch freie Beobachtungen, wo wir einfach dann an die Kinder, so zehn Minuten viertel Stunde einfach nur beobachten und aufschreiben, was passiert.“ (Transkript 2011)

„Ist natürlich nichts im Kern, was wir als unsere Aufgabe verstehen, konkrete Diagnostik zu machen, aber, um uns zu sortieren und, um zu schauen, genau, was liegt denn da jetzt eigentlich vor, verwenden wir schon das ein oder andere.“ (Transkript 2161)

„Ja, das machen wir einmal die Woche beobachten wir und verschriftlichen das, sprechen das dann mit der Kollegin in der Gruppe nochmal durch, wenn wir ein Kleinteam machen. Und wenn es jetzt irgendwie was ganz Besonderes ist geht es dann auch noch ins Großteam zum kollegialen Austausch (...) Aus den Beobachtungen werden wir dann... schreiben wir dann eine Geschichte für das Portfolio, wie entwickelt sich das Kind weiter, welche Interessen hat es gerade, wie versuchen wir das zu unterstützen und gehen so dann eben auch in ein Elterngespräch.“ (Transkript 2171)

Insbesondere die letzte Aussage zeigt, wie auf der Grundlage von Beobachtungen und deren Dokumentation eine Praxis der kollegialen Zusammenarbeit mit Blick auf die Beschreibung und Förderung kindlicher Entwicklung umgesetzt werden kann.

3.3.3 Welche Potenziale eröffnen sich durch die Nutzung eines prozessbegleitenden Portfolios für die frühe Bildung?

Im Rahmen der Studie der Autorengruppe erhielten die PFK ein prozessbegleitendes Portfolio für den Bildungsinhalt „Schwimmen und Sinken“, in welchem die kindlichen Lernvoraussetzungen zu spezifischen Zeitpunkten des Bildungsangebots eingeschätzt wurden, Entwicklungen dokumentiert und Förderentscheidungen abgeleitet wurden (vgl. Hardy et al., im Druck). Im Interview sollte anhand der Leitfrage „Könnten Sie sich eine Nutzung eines ähnlichen Portfolios im KiTa-Alltag vorstellen?“ erkundet werden, welche Einstellungen von den PFK zur prozessbezogenen Diagnostik anhand der unterstützenden Materialien im Verlauf des Projekts bzw. nach der Umsetzung des Bildungsangebots entwickelt wurden. Die Antworten der PFK lassen sich in Aussagen einordnen, welche keinen weiteren Nutzen eines prozessbegleitenden Portfolios für die Arbeit in der KiTa sehen (5 Nennungen) und Aussagen, welche Begründungen auf einer stärker produktorientierten oder stärker prozessorientierten Ebene anführen. Eine Produktorientierung zielt auf den schnellen Überblick zum Entwicklungsstand und den getroffenen Fördermaßnahmen ab (9 Nennungen), beispielsweise *„Und ansonsten hat*

mir dieses Portfolio einfach nochmal geholfen, den Überblick zu behalten, an welcher Stelle ich da den Kindern eben vielleicht noch eine Förderung zugutekommen lassen kann“ (Transkript 2161). Die Prozessorientierung hebt die Sichtbarkeit der Entwicklung eines Kindes in einem begrenzten Zeitraum sowie Messung und Beobachtung dieser Entwicklung aus unterschiedlichen Perspektiven hervor (9 Nennungen). Sie zeigt sich in Aussagen wie *„weil in der Situation an sich, wenn man das halt durchführt, dann ist man auch eher noch einmal darauf bedacht, dass man das, was man da vorhatte, auch macht und man hat da schon den Blick auf die Kinder, aber man ist sich in dem Augenblick ja nicht so bewusst (...) und erst bei dem Fragebogen, also, wenn man dann vor dem Bogen gesessen hat und dann das einzelne Kind sich noch einmal hergeholt hat vor Augen, war das viel einfacher das dann noch einmal zu sehen.“* (Transkript 2162) und *„und das war auch schon interessant, in welchen Sachen wir uns einig waren oder auch gesagt haben, och ne, das sehe ich jetzt ganz anders oder halt auch eben der Beobachter, der objektive Beobachter zu sein von außen.“* (Transkript 2211). Schließlich lässt sich die Kindorientierung als weitere Kernaussage herausarbeiten (3 Nennungen), welche insbesondere das Potenzial eines Portfolios für die Dokumentation und Reflexion von Bildungsprozessen mit dem jeweiligen Kind hervorhebt, so beispielsweise *„aber das jetzt richtig mit Bildern zu unterlegen, das glaube ich, ist ja richtig und den Kindern das nochmal, na, und dann wenn man die dann fragt, kannst du dich da noch erinnern, was haben wir denn da gemacht, und was, da kommt ja dann auch nochmal viel, das ist nochmal immer, immer beim Portfolio glaube ich sehr wichtig“* (Transkript 1061). Insgesamt zeigt sich in den Aussagen zu den Potenzialen eines prozessbegleitenden Portfolios, dass die PFK ihren Blick auf naturwissenschaftliche Bildungsprozesse erweitern. Anders als in ihren Aussagen zum Stellenwert der naturwissenschaftlichen Bildung sowie der Diagnostik in der KiTa werden hier Beobachtungen stärker mit Lernprozessen verknüpft, eine Verknüpfung mit individueller Förderung hergestellt und der Abgleich von Beobachtungen sowie Einschätzungen aus unterschiedlicher Perspektive hervorgehoben. Trotz der genannten neuen Perspektiven, welche z.T. auch die Bedeutung von Portfolios für die naturwissenschaftliche Domäne aufnehmen, wird von PFK jedoch der Zeitaufwand bei einer solchen prozessbegleitenden Diagnostik als zu hoch für die Integration in die tägliche Praxis eingeschätzt.

4 Diskussion und Ausblick

Ziel der vorliegenden Studie war es, Einblicke in die Einstellungen von PFK zur prozessbegleitenden Diagnostik im Kontext naturwissenschaftlicher Bildung zu erhalten. Die Ergebnisse zeigen, dass die PFK naturwissenschaftliche Bildung als wichtig erachten und sich als selbstwirksam hinsichtlich der Diagnostik und adaptiven Begleitung erleben. Dies sind zunächst wichtige Voraussetzungen für die Umsetzung einer gelingenden naturwissenschaftlichen Bildung. Gleichzeitig deuten die Ergebnisse der Interviews darauf, dass prozessbegleitende Diagnostik im Bil-

dungsbereich Naturwissenschaften kaum eine Rolle spielt. So werden zwar Beispiele zur Umsetzung naturwissenschaftlicher Bildung genannt, es wird aber vor allem auf Oberflächenmerkmale verwiesen (z. B. Waldtag), sodass unklar bleibt, inwiefern naturwissenschaftliche Bildungsprozesse stattfinden. Die Fachkräfte berichten zudem über den Einsatz KiTa-spezifischer Dokumentationsinstrumente, die eher generelle Entwicklungen abbilden und andere Funktionen haben als eine prozessbegleitende Diagnostik in einem spezifischen Bildungsangebot.

In den Interviews zeigt sich eine starke Orientierung an Selbstbildungsansätzen, die mit einer eher passiven Rolle der PFK einhergehen (Schmidt & Smidt, 2021). Umgekehrt werden bei einer ko-konstruktivistischen Einstellung, die eine aktivere Rolle der PFK annimmt, auch die Diagnostik und die Anpassung von Bildungsangeboten bedeutsam. Möglicherweise stehen also die selbstbildungsorientierten Einstellungen in unserer Stichprobe dem Einsatz von formativem Assessment im Wege. Eine weitere denkbare Ursache, die hier nicht im Mittelpunkt stand, ist ein niedriges fachspezifisches Professionswissen. Dieses ist notwendig, um z. B. zu erkennen, welches naturwissenschaftliche Lernpotential eine Situation hat und welche (lerngegenstandsspezifischen) Fragen oder Impulse zur Diagnostik und Förderung geeignet wären. In den Bewertungen des Portfolios (Hardy et al., im Druck), das Elemente der Diagnostik und Förderung in der naturwissenschaftlichen Domäne koppelt, zeigten sich durchaus (domänenspezifische und domänenübergreifende) diagnostische Aussagen der PFK. Dies deutet auf die Bedeutung unterstützender Maßnahmen für Fachkräfte, z. B. geeignete Materialien und Fortbildungen, für eine gelungene Umsetzung von formativem Assessment hin. Eine Herausforderung ist dabei sicherlich die wahrgenommene Belastung vieler PFK. So werden z. B. die Auswahl und Umsetzung der Entwicklungsdokumentationen oft als zeitlich überfordernd eingeschätzt (Ruppin et al., 2015; Viernickel et al., 2013). Für die Implementierung von entsprechenden Instrumenten zur Unterstützung einer prozessbezogenen Diagnostik und Förderung muss also insbesondere ihre Praxistauglichkeit in den Blick genommen werden.

Literatur

- Ajzen, I. (1985). From intentions to actions. A theory of planned behavior. In J. Kuhl & J. Beckmann (Hrsg.), *Action control: From cognition to behavior* (S. 11–39). Springer. https://doi.org/10.1007/978-3-642-69746-3_2
- Anders, Y., Hardy, I., Pauen, S. & Steffensky, M. (2013). Zieldimensionen naturwissenschaftlicher Bildung im Kita-Alter und ihre Messung. In Stiftung Haus der kleiner Forscher (Hrsg.), *Wissenschaftliche Untersuchungen zur Arbeit der Stiftung „Haus der kleinen Forscher“* (Bd. 5, S. 19–82). Schubi Lernmedien AG.
- Anders, Y. & Rossbach, H. G. (2015). Preschool teachers' sensitivity to mathematics in children's play: The influence of math-related school experiences, emotional attitudes, and pedagogical beliefs. *Journal of Research in Childhood Education*, 29(3), 305–322. <http://dx.doi.org/10.1080/02568543.2015.1040564>

- Bartusch, S., Klektau, C., Simon, T., Teumer, S. & Weidermann, A. (2019). *Lernprozesse begleiten: Anforderungen an pädagogische Institutionen und ihre Akteur*innen*. Springer VS. <https://doi.org/10.1007/978-3-658-21924-6>
- Bauer, C., Krause, M. & Mayr, T. (2012). *Kompik. Kompetenzen und Interessen von Kindern. Beobachtungs- und Entwicklungsbogen für Kinder von 3,5 bis 6 Jahre*. Bertelsmann Stiftung.
- Behr, K. & Walter, M. (2012). *Qualifikation und Weiterbildung frühpädagogischer Fachkräfte* (WiFF Studien, Bd. 15). DJI.
- Beller, S. (2016). *Kuno Bellers Entwicklungstabelle 0–9*. Eigenverlag.
- Benz, C. (2012). Attitudes of kindergarten educators about math. *Journal für Mathematik-Didaktik*, 33(2), 203–232. <https://doi.org/10.1007/s13138-012-0037-7>
- Bruns, J. (2014). *Adaptive Förderung in der elementarpädagogischen Praxis. Eine empirische Studie zum didaktischen Handeln von Erzieherinnen und Erziehern im Bereich Mathematik*. Waxmann.
- Bürgermeister, A. & Saalbach, H. (2018). Formatives Assessment: Ein Ansatz zur Förderung individueller Lernprozesse. *Psychologie in Erziehung und Unterricht*, 65(3), 194–205. <https://doi.org/10.2378/peu2018.art11d>
- Corno, L. Y. N. (2008). On teaching adaptively. *Educational Psychologist*, 43(3), 161–173. <https://doi.org/10.1080/00461520802178466>
- Drexl, D., Born-Rauchenecker, E. & Kalicki, B. (2019). Naturwissenschaftliche Einstellungen angehender pädagogischer Fachkräfte. *Frühe Bildung*, 8(1), 30–36. <https://doi.org/10.1026/2191-9186/a000408>
- Dunekacke, S. & Barenthien, J. (2021). Research in early childhood teacher domain-specific professional knowledge – a systematic review. *European Early Childhood Education Research Journal*, 29(4), 633–648. <https://doi.org/10.1080/1350293X.2021.1941166>
- Fives, H. & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers’ beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer & M. Zeidner (Hrsg.), *APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (S. 471–499). American Psychological Association. <https://doi.org/10.1037/13274-019>
- Fröhlich-Gildhoff, K., Nentwig-Gesemann, I. & Pietsch, S. (2011). *Kompetenzorientierung in der Qualifizierung frühpädagogischer Fachkräfte. Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF)* (Bd. 19). DJI.
- Fthenakis, W. E. (2009). Ko-Konstruktion: Lernen durch Zusammenarbeit. *Kinderzeit*, (3), 9–13.
- Geiling, U. & Berger, M. (2015). Diagnostischer Informationstransfer von der Kita in die Grundschule aus Sicht der Eltern, der Erzieherinnen und Lehrerinnen – ausgewählte Ergebnisse aus dem Forschungsprojekt ILEA T. In M. Urban, M. Schulz, K. Meser & S. Thoms (Hrsg.), *Inklusion und Übergang. Perspektiven der Vernetzung von Kindertageseinrichtungen und Grundschulen* (S. 217–231). Klinkhardt.
- Hardy, I., Leuchter, M., Saalbach, H., Steffensky, M., Bürgermeister, A., Junge, K. & Venitz, L. (2021). *Handreichung – Überblick und Umsetzungsbeispiele aus dem Projekt FinK. Begleitende Materialien zu einem Bildungsangebot „Schwimmen und Sinken“ für inklusiv arbeitende Kitas*. Online verfügbar unter: <https://www.ew.uni-hamburg.de/einrichtungen/ew5/didaktik-der-chemie/files/handreichtung-fink.pdf>
- Hardy, I., Bürgermeister, A. & Leuchter, M. (im Druck). Portfolios in der frühen naturwissenschaftlichen Bildung: Bedeutung für die Sprachförderung in inklusiven Kontexten. In E. Blumberg, C. Niederhaus & A. Mischendahl (Hrsg.), *Mehrsprachigkeit in der Schule. Sprachbildung im und durch Sachunterricht*. Kohlhammer.

- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Höltge, L., Ehm, J.-H., Hartmann, U. & Hasselhorn, M. (2019). Teachers' self-efficacy beliefs regarding assessment and promotion of school-relevant skills of preschool children. *Early Child Development and Care*, 189(2), 339–351. <https://doi.org/10.1080/03004430.2017.1323888>
- Hooker, T. (2019). Using ePortfolios in early childhood education: Recalling, reconnecting, restarting and learning. *Journal of Early Childhood Research*, 17(4), 376–391. <https://journals.sagepub.com/doi/10.1177/1476718X19875778>
- Kingston, N. & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Kleickmann, T., Steffensky, M. & Praetorius, A.-K. (2020). Quality of teaching in elementary science education: More than three basic dimensions? In A.-K. Praetorius & E. Klie-me (Hrsg.), *Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen* (S. 37–55). Beltz Juventa.
- Krajewski, K., Nieding, G. & Schneider, W. (2007). *Mengen, zählen, Zahlen. Die Welt der Mathematik verstehen. Koffer mit Fördermaterialien*. Cornelsen Verlag.
- Lee, H., Chung, H. Q., Zhang, Y., Abedi, J. & Warschauer, M. (2020). The Effectiveness and Features of Formative Assessment in US K-12 Education: A Systematic Review. *Applied Measurement in Education*, 33(2), 124–140. <https://doi.org/10.1080/08957347.2020.1732383>
- Meschede, N. & Hardy, I. (2020). Selbstwirksamkeitserwartungen von Lehramtsstudierenden zum adaptiven Unterrichten in heterogenen Lerngruppen. *Zeitschrift für Erziehungswissenschaft*, 23, 565–589. <https://doi.org/10.1007/s11618-020-00949-7>
- Ruppig, I., Prigge, J., Pages, S. & Adam, A. (2015). Professionalisierung der pädagogischen Fachkräfte in Kindertagesstätten durch Beobachtung und Dokumentation. Möglichkeiten und Grenzen der Fort- und Weiterbildung. In I. Ruppig (Hrsg.), *Professionalisierung in Kindertagesstätten* (S. 140–162). Beltz Juventa.
- Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H. & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, 109(8), 1049–1066. <https://doi.org/10.1037/edu0000190>
- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B. & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 101602. <https://doi.org/10.1016/j.ijer.2020.101602>
- Schlaaf-Kirschner, K. (2014). *Der Beobachtungsbogen für Kinder von 3–6. Mit Tipps und Materialien für die Kita-Praxis*. Verl. an der Ruhr.
- Schmidt, T. & Smidt, W. (2021). Selbstbildung, Ko-Konstruktion oder Instruktion? *Zeitschrift für Pädagogik*, 67(2), 251–274. <https://doi.org/10.3262/ZP2102251>
- Schmidt, C. & Liebers, K. (2017). Formatives Assessment im inklusiven Unterricht—Forschungsstand und erste Befunde. In F. Hellmich & E. Blumberg (Hrsg.), *Inklusiver Unterricht in der Grundschule* (S. 50–65). Kohlhammer.
- Schütze, B., Souvignier, E. & Hasselhorn, M. (2018). *Stichwort – formatives Assessment*. <https://doi.org/10.25656/01:16754>
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M. K. & Yin, Y. (2008). On the impact of curriculum-embedded formative assess-

- ment on learning: a collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295–314. <https://doi.org/10.1080/08957340802347647>
- Skopek, J. & Passaretta, G. (2021). Socioeconomic Inequality in Children's Achievement from Infancy to Adolescence: The Case of Germany. *Social Forces*, 100(1), 86–112. <https://doi.org/10.1093/sf/soaa093>
- Steffensky, M. (2017). *Naturwissenschaftliche Bildung in Kindertageseinrichtungen. Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF)* (Band 48). Deutsches Jugendinstitut e.V.
- Steffensky, M., Anders, Y., Barenthien, J., Hardy, I., Leuchter, M. & Oppermann, E. (2018). Early Steps into Science – EASI Science Wirkungen früher naturwissenschaftlicher Bildungsangebote auf die naturwissenschaftlichen Kompetenzen von Fachkräften und Kindern. In Stiftung Haus der kleinen Forscher (Hrsg.), *Wirkungen naturwissenschaftlicher Bildungsangebote auf pädagogische Fachkräfte und Kinder* (S. 51–136). Verlag Barbara Budrich. <https://doi.org/10.2307/j.ctvmx3jn8.8>
- Tietze, W. (2004). Pädagogische Qualität in Familie, Kindergarten und Grundschule und ihre Bedeutung für die kindliche Entwicklung. In G. Faust-Siehl, M. Götz, H. Hacke & H.-G. Roßbach (Hrsg.), *Anschlussfähige Bildungsprozesse im Elementar- und Primarbereich* (S. 139–153). Klinkhardt Verlag.
- Van de Pol, J., Volman, M., Oort, F. & Beishuizen, J. (2015). The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support. *Instructional Science*, 43(5), 615–641. <https://doi.org/10.1007/s11251-015-9351-z>
- Viernickel, S., Nentwig-Gesemann, I., Nicolai, K., Schwarz, S. & Zenker, L. (2013). *Schlüssel zu guter Bildung, Erziehung und Betreuung: Bildungsaufgaben, Zeitkontingente und strukturelle Rahmenbedingungen in Kindertageseinrichtungen*. GEW.

Formatives Assessment im naturwissenschaftlichen Grundschulunterricht

Was wissen Studierende über Möglichkeiten der Diagnostik und der Rückmeldung und wie hängt dieses Wissen mit ihren Lehr-Lern-Überzeugungen zusammen?

Verena Zucker¹ | Nicola Meschede¹ | Miriam Leuchter²

¹ Universität Münster | ²Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau

Abstract: Formatives Assessment stellt eine wichtige Voraussetzung für einen wirksamen naturwissenschaftlichen Grundschulunterricht dar. Eine entsprechende Umsetzung ist für Lehrpersonen allerdings mit verschiedenen Herausforderungen verbunden. Eine frühzeitige Anbahnung professioneller Kompetenz bereits in der universitären Ausbildung scheint daher bedeutsam. Um davon ausgehend Empfehlungen für die Konstruktion von universitären Lehrveranstaltungen abzuleiten, wurden 148 Bachelorstudierende des Grundschullehramts hinsichtlich ihres Wissensstands zu Möglichkeiten der Diagnostik und der Rückmeldung befragt. Zusätzlich wurde der Zusammenhang dieses Wissens mit Lehr-Lern-Überzeugungen untersucht. Die Ergebnisse zeigen, dass nach dem Besuch einer Vorlesung bereits Wissen zu Möglichkeiten der Diagnostik und der Rückmeldung vorhanden zu sein scheint. Dieses ist allerdings häufig noch nicht ausdifferenziert sowie nicht immer formativ ausgerichtet. Ein Zusammenhang zeigt sich zudem zwischen dem Wissen zur Diagnostik und den konstruktivistischen Lehr-Lern-Überzeugungen. Aus den Ergebnissen werden Implikationen für die Entwicklung wirksamer Lehrkonzepte sowie Desiderate für die Forschung abgeleitet.

1 Einleitung

In der Lehr-Lern-Forschung wird das formative Assessment als ein wichtiges Merkmal von Unterrichtsqualität diskutiert (Praetorius et al., 2020). Es beschreibt den Prozess, individuelle Lernstände während des Lernens zu diagnostizieren und entsprechende Informationen so zu nutzen, dass Lehr-Lern-Prozesse optimiert werden können (Black & Wiliam, 2009; Cowie & Bell, 1999). Dem Geben von Rückmeldungen wird dabei eine Kernaufgabe zugesprochen (Heitink et al., 2016).

Die Wirksamkeit formativen Assessments zeigt sich u.a. in Bezug auf die Leistung (Klute et al., 2017) sowie auf die Motivation von Lernenden (Weurlander et al., 2012), dies gilt auch für den naturwissenschaftlichen Grundschulunterricht (Decristan et al., 2015; Hondrich et al., 2018). Die Umsetzung von formativem Assessment ist für Lehrpersonen allerdings mit verschiedenen Anforderungen verbunden (Hondrich et al., 2016), ein Umgang damit setzt professionelle Kompetenz voraus. Dies umfasst in An-

lehnung an Blömeke et al. (2015) u.a. professionelles Wissen. Zusätzlich scheinen aber auch ausgeprägte konstruktivistische Lehr-Lern-Überzeugungen bedeutsam (Heitink et al., 2016). Eine frühzeitige Anbahnung des formativen Assessments als Teil professioneller Kompetenz sollte daher bereits in der universitären Ausbildung von Lehrkräften erfolgen. Für die Konstruktion passender Lehrveranstaltungen können dabei Informationen über bestehende Kompetenzen von Studierenden hilfreich sein. So zeigt z.B. eine Studie von Maclellan (2004), dass angehende Lehrpersonen lediglich über begrenztes Wissen hinsichtlich Prinzipien sowie konkreter Methoden des formativen Assessments verfügen. Daran setzt auch die vorliegende Studie an und untersucht den Wissensstand zu Möglichkeiten der Diagnostik und der Rückmeldung bei Bachelorstudierenden sowie den Zusammenhang dieser Aspekte des formativen Assessments mit Lehr-Lern-Überzeugungen. Der Fokus liegt auf dem naturwissenschaftlichen Grundschulunterricht.

2 Formatives Assessment im naturwissenschaftlichen Grundschulunterricht

2.1 Diagnostik und Rückmeldung als zwei Kernmerkmale

In Studien zeigen sich vielfältige Konzeptualisierungen des formativen Assessments, die zum Teil unterschiedliche Ausrichtungen und Kernmerkmale aufweisen (Schildkamp et al., 2020). Als übereinstimmende Merkmale können jedoch (1) das Diagnostizieren individueller Lernstände sowie (2) das Geben von Rückmeldungen identifiziert werden (Black & Wiliam, 2009; Bürgermeister & Saalbach, 2018). Beides ist auch für den naturwissenschaftlichen Grundschulunterricht bedeutsam (Zucker & Leuchter, 2018):

- (1) Lernen wird im naturwissenschaftlichen Grundschulunterricht als konstruktivistischer Prozess verstanden, bei dem ausgehend von den bestehenden Konzepten der Lernenden Lernprozesse angeregt werden (Möller, 2018). Allerdings sind die Konzepte der Lernenden häufig individuell verschieden und können sich im Unterrichtsverlauf auch auf unterschiedliche Art und Weise verändern (Jonen et al., 2003). Um vor diesem Hintergrund eine dauerhafte Passung angebotener Lerngelegenheiten gewährleisten zu können, bedarf es stetiger Diagnostik.
- (2) Die individuellen Konzepte von Lernenden können sich zu unterschiedlichen Zeitpunkten im Unterricht hinsichtlich des Grads ihrer Wissenschaftlichkeit unterscheiden (Jonen et al., 2003). Eine Umstrukturierung bzw. Ausdifferenzierung bestehender Konzepte kann durch entsprechende Rückmeldungen unterstützt werden. Bestätigende Rückmeldungen könnten z. B. eine Festigung bereits annähernd wissenschaftlicher Konzepte bzw. Anteile ermöglichen, korrigierende Rückmeldungen könnten dagegen eine Umstrukturierung wissenschaftlich nicht haltbarer Konzepte anregen.

2.2 Anforderungen an die Lehrperson

Die Umsetzung von formativem Assessment ist ein komplexer Prozess mit verschiedenen Anforderungen (Hondrich et al., 2016; Yan et al., 2022). Im Kontext des Diagnostizierens und Rückmeldens sind es u.a. Folgende:

- (1) Um individuelle Konzepte ermitteln zu können, bedarf es der *Gestaltung diagnostischer Methoden* bzw. der Evaluation bestehender Methoden (Heitink et al., 2016). Solche Methoden können z. B. schriftlich, mündlich oder auch in Form einer Beobachtung umgesetzt werden. Ihre Gestaltung ist dabei u.a. abhängig von der Art der Information, die eine Lehrperson erhalten möchte (z. B. zu Konzepten vor dem Unterricht oder persönlichen Interessen; Cowie & Bell, 1999), und von den Voraussetzungen der Lernenden wie z. B. ihren Schreib- und Lesekompetenzen.
- (2) Um aufbauend auf den Konzepten der Lernenden passende Rückmeldungen generieren zu können, ist die *Interpretation ermittelter Informationen* von großer Relevanz. Diese Interpretation setzt u.a. voraus, dass Lehrpersonen geäußerte Konzepte der Lernenden hinsichtlich des Grads der Wissenschaftlichkeit einschätzen können (Cowie & Bell, 1999).
- (3) Aufgrund heterogener Konzepte (Jonen et al., 2003) bedarf es auch des *Gebens individueller Rückmeldungen* an die Lernenden. Diese sollten nicht nur informierend und wertschätzend formuliert, sondern auch an die jeweiligen Voraussetzungen der Lernenden angepasst sein (Bürgermeister & Saalbach, 2018; Lipowsky, 2015). So zeigen z. B. Studien, dass die Komplexität einer Rückmeldung abhängig von u.a. der Leistungsstärke der Lernenden gestaltet werden sollte (Mory, 2004).
- (4) Zusätzlich können neben der Lehrperson auch die Lernenden für sich (*Self-Assessment*) und füreinander (*Peer-Assessment*) als Ressourcen des formativen Assessments aktiviert werden (Black & Wiliam, 2009). Dies kann als anspruchsvoller Prozess für Lernende angesehen werden, der ebenfalls Unterstützung durch die Lehrperson bedarf (Bürgermeister & Saalbach, 2018).

2.3 Professionelle Kompetenz von Lehrpersonen

Der Umgang mit den aufgeführten Anforderungen setzt professionelle Kompetenz seitens der Lehrperson voraus. Dazu gehören u.a. kognitive Dispositionen wie das professionelle Wissen und affektiv-motivationale Dispositionen wie Lehr-Lern-Überzeugungen (Blömeke et al., 2015).

Professionelles Wissen zur Diagnostik und Rückmeldung. Um formatives Assessment erfolgreich umsetzen zu können, benötigt eine Lehrperson eine differenzierte Wissensbasis (Cowie & Bell, 1999; Heitink et al., 2016). Dazu gehört u.a. pädagogisches Wissen, Fachwissen und fachdidaktisches Wissen (Shulman, 1987). Pädagogisches Wissen umfasst das fachunabhängige Wissen über die Gestaltung von Lehr-Lern-Prozessen wie z. B. fachübergreifende Prinzipien der Prozessdiagnostik (Baumert &

Kunter, 2011). Fachwissen beinhaltet das konzeptuelle Verständnis zu einem fachlichen Inhalt (Baumert & Kunter, 2011) und ist insbesondere für die wissenschaftliche Einschätzung bestehender Konzepte bedeutsam. Fachdidaktisches Wissen bezeichnet das unterrichts- und schüler*innenbezogene fachliche Wissen (Baumert & Kunter, 2011) wie z. B. Wissen zu typischen Konzepten von Lernenden und dazu passenden lernförderlichen Rückmeldungen.

Studien zum professionellen Wissen zeigen, dass eine differenzierte Wissensbasis bei Grundschullehrpersonen häufig nicht in ausreichendem Maße vorhanden ist (Kleickmann, 2015; Zulliger & Buholzer, 2020). So zeigen z. B. Zulliger und Buholzer (2020) im Kontext des pädagogischen Wissens, dass Grundschullehrpersonen zwar die Bedeutung und Funktion des Vorwissens kennen, jedoch über ein geringeres Wissen zur individuellen Unterstützung von Lernprozessen sowie zur formativen Nutzung von Lernkontrollen verfügen. Für den naturwissenschaftlichen Grundschulunterricht gilt als zusätzliche Herausforderung zu berücksichtigen, dass aufgrund der Vielperspektivität des Faches gleich in verschiedenen Bezugsdisziplinen (Physik, Biologie, Chemie) sowohl Fach- als auch fachdidaktisches Wissen erworben werden muss (Kleickmann, 2015).

Lehr-Lern-Überzeugungen. Erste empirische Befunde deuten darauf hin, dass die Umsetzung formativen Assessments positiv mit konstruktivistischen Lehr-Lern-Überzeugungen zusammen zu hängen scheint (Birenbaum et al., 2011; Schmidt, 2018). Dabei wird unter konstruktivistischen Lehr-Lern-Überzeugungen die Auffassung verstanden, dass Lernen das aktive Konstruieren von Wissen durch die Lernenden selbst umfasst (Kleickmann, 2008). Solch eine Sichtweise entspricht auch der auf das Lernen im naturwissenschaftlichen Grundschulunterricht (Möller, 2018). In der Literatur werden konstruktivistische Lehr-Lern-Überzeugungen häufig transmissiven Lehr-Lern-Überzeugungen gegenübergestellt. Darunter werden die Auffassungen subsumiert, die vom Lernen als stark gesteuerten Prozess und einer Weitergabe von Wissen ausgehen (Kleickmann, 2008). Obwohl es sich bei den konstruktivistischen und den transmissiven Überzeugungen um zwei verschiedene Dimensionen handelt, können beide auch simultan vorliegen (Voss et al., 2011). In Bezug auf Grundschullehrpersonen zeigen Studien, dass konstruktivistische Überzeugungen stärker ausgeprägt zu sein scheinen als transmissive (Kleickmann, 2015).

Zusammenhang zwischen den Konstrukten. Studien, bspw. aus dem naturwissenschaftlichen Grundschulunterricht, zeigen zwischen Fachwissen bzw. fachdidaktischem Wissen und (a) konstruktivistischen Lehr-Lern-Überzeugungen positive Zusammenhänge sowie (b) transmissiven Lehr-Lern-Überzeugungen negative Zusammenhänge (Brunner et al., 2006; Meschede et al., 2017).

Aus der Literatur können unterschiedliche Erklärungen herangezogen werden, z. B.: (1) Professionelles Wissen dient als Quelle für die Entwicklung von Überzeugungen (Buehl & Fives, 2009) sowie (2) Überzeugungen sind „Filter“, vor deren Hintergrund u.a. neues Wissen interpretiert und hinsichtlich der Vereinbarkeit mit bestehendem Wissen eingeschätzt wird (Levin, 2015).

3 Fragestellungen

Folgende Forschungsfragen stehen im Fokus dieser Studie:

- (1) Über welches Wissen zu Möglichkeiten der Diagnostik und der Rückmeldung verfügen Studierende im Kontext des naturwissenschaftlichen Grundschulunterrichts?
- (2) Wie hängt dieses Wissen mit den Lehr-Lern-Überzeugungen der Studierenden zusammen?

4 Methode

4.1 Stichprobe & Erhebung

Die Stichprobe umfasste 148 Grundschullehrerstudierende in ihrem Bachelorstudium (\bar{X} -Semester: 5,72; \bar{X} -Alter: 23,04 Jahre) an der Universität Koblenz-Landau (Standort Landau). 84,5 % der Teilnehmenden waren weiblich. Die Erhebung wurde im Wintersemester 2016/2017 bzw. im Sommersemester 2017 am Ende einer Vorlesung zum Thema „Grundlagen des Sachunterrichts“ umgesetzt, die neben allgemeinen lehr-lerntheoretischen Grundlagen (u.a. Conceptual Change, Scaffolding) in einer Sitzung auch das formative Assessment fokussierte. Die Studierenden erhielten eine kurze Einführung in das formative Assessment (Kernmerkmale, Zeitpunkt), darauf aufbauend wurden das Diagnostizieren und Rückmelden konkretisiert. Bei der Diagnostik standen insgesamt zwölf Methoden im Fokus (u.a. Gespräche, Zeichnungen). Tests wurden im Kontext der formativen Nutzung diskutiert. Für die Rückmeldung wurden neun Methoden im Sinne des Teacher- und Peer-Assessments (u.a. Kommentare als schriftliche und mündliche Variante) und vier Methoden im Sinne des Self-Assessments (u.a. Selbsteinschätzungsbögen) präsentiert. Vereinzelt wurden Methoden durch Bilder bzw. transkribierte Unterrichtssituationen veranschaulicht, einmal erfolgte eine Veranschaulichung anhand des naturwissenschaftlichen Grundschulunterrichts.

Die Erhebung des Wissens zu Möglichkeiten der Diagnostik und der Rückmeldung sowie der Lehr-Lern-Überzeugungen war in eine größer angelegte Online-Befragung zur Erfassung professioneller Kompetenz zum formativen Assessment im naturwissenschaftlichen Grundschulunterricht eingebettet (Dauer: ca. sechzig Minuten). Diese bearbeiteten die Studierenden in der letzten Vorlesungssitzung und ca. vier Wochen nach der Sitzung zum formativen Assessment zuhause.

4.2 Instrumente & statistische Analysen

Wissen zur Diagnostik und Rückmeldung. Das Wissen wurde anhand von zwei Items mit offenem Antwortformat erfasst, die sich auf Möglichkeiten der Diagnostik und

der Rückmeldung im naturwissenschaftlichen Grundschulunterricht beziehen („Welche diagnostischen Möglichkeiten/Möglichkeiten der Rückmeldung im naturwissenschaftlichen Sachunterricht sind Ihnen bekannt?“). Die Antworten der Studierenden wurden anschließend durch ein deduktiv-induktives Verfahren qualitativ inhaltsanalytisch ausgewertet (31 % Doppelkodierung; Cohen's κ Min = .77 und Max = 1.0). In Bezug auf die Diagnostik lag der Fokus auf konkreten Methoden der Diagnostik sowie deren zeitlicher Einsatz im Lehr-Lern-Prozess. Da im Kontext der Diagnostik vielfältige Methoden genannt wurden, wurden die Kategorien, die von weniger als sechs Personen genannt wurden, in der Kategorie „Sonstiges“ zusammengefasst. In dieser Kategorie wurden bei Nennung mehrerer Methoden mehrere Punkte vergeben. Ansonsten erhielten die Teilnehmenden pro passende Analyseeinheit einen Punkt. In Bezug auf die Rückmeldung lag der Fokus auf Methoden der Rückmeldung und Gestaltungsmöglichkeiten.

Lehr-Lern-Überzeugungen. Die Erfassung der Lehr-Lern-Überzeugungen erfolgte in Anlehnung an die Skalen von Kleickmann (2008) auf einer fünfstufigen Ratingskala (1 = „stimmt gar nicht“ – 5 = „stimmt völlig“). Die konstruktivistischen Überzeugungen umfassen insgesamt neun Items zu Schüler*innenvorstellungen sowie zum Conceptual Change ($M = 3.51$, $SD = 0.62$). Die transmissiven Lehr-Lern-Überzeugungen umfassen sieben Items ($M = 2.88$, $SD = 0.59$). Die interne Konsistenz der Skalen beträgt Cronbachs $\alpha = .80$ (konstruktivistisch) bzw. Cronbachs $\alpha = .73$ (transmissiv).

Die Auswertung der Daten erfolgte mittels deskriptiver Verfahren. Zur Beantwortung der ersten Fragestellung wurden die quantifizierten Daten prozentual berechnet. Zur Beantwortung der zweiten Fragestellung wurden die einzelnen Kategorien der Wissensbereiche addiert und die Korrelationskoeffizienten nach Pearson zwischen diesen Variablen und den Lehr-Lern-Überzeugungen bestimmt.

5 Ergebnisse

Wissen zur Diagnostik. Die Studierenden nannten im Durchschnitt 3.4 Möglichkeiten ($SD = 1.58$; Range = 0-8), mit denen im naturwissenschaftlichen Grundschulunterricht Lernstände diagnostiziert werden können. Die genannten Möglichkeiten unterschieden sich dabei (1) in ihrem fachspezifischen Bezug zum naturwissenschaftlichen Grundschulunterricht (s. Tabelle 1). So beschrieben die Studierenden sowohl fachübergreifende methodische Möglichkeiten (z. B. „Fragen stellen“ als Methode der mündlichen Diagnostik), die auch im naturwissenschaftlichen Grundschulunterricht eingesetzt werden können, als auch naturwissenschaftsspezifische methodische Möglichkeiten (z. B. „Vermutungen erfassen“). Die genannten Möglichkeiten unterschieden sich (2) in ihrer Konkretheit. So wurde teilweise ausschließlich der grobe Kontext genannt (z. B. „beim Experimentieren“), z.T. wurden diese Möglichkeiten detaillierter beschrieben (z. B. „Experiment planen und notieren lassen“).

Bezüglich des Zeitpunkts der Diagnostik wurde von vielen Studierenden häufig eine Erfassung des Vorwissens benannt (37,2 %). Die Erfassung von Postkonzepten wurde lediglich von wenigen Studierenden beschrieben (5,4 %), die Erfassung von Zwischenkonzepten sogar nur von einer Person.

Tabelle 1: Kategorien und Häufigkeit der beschriebenen diagnostischen Möglichkeiten in Bezug auf Methoden und Zeitpunkt

Diagnostische Möglichkeiten	% (Anzahl Personen)
Fachübergreifende Möglichkeiten	
Schriftliche Diagnostik (allgemein)	5,4 % (8)
Lerntagebücher/Portfolios	9,5 % (14)
Aufgaben	11,5 % (17)
Quiz	5,4 % (8)
Tests/Klassenarbeiten	51,4 % (76)
Hausaufgaben	4,7 % (7)
Mündliche Diagnostik (allgemein)	4,7 % (7)
Unterrichtsgespräch (Diskussionen/Fragen)	65,5 % (97)
Präsentationen	5,4 % (8)
Beobachten	18,9 % (28)
Naturwissenschaftsspezifische Möglichkeiten	
Experimentierprozesse	27,7 % (41)
Spezifische Denk-, Arbeits- und Handlungsweisen	17,6 % (26)
Sonstige fachübergreifende und naturwissenschaftsspezifische Möglichkeiten (Zeichnungen, Mappingverfahren, ...)	31,8 % (47)
Zeitpunkt	
Präkonzepte erfassen	37,2 % (55)
Zwischenkonzepte erfassen	0,7 % (1)
Postkonzepte erfassen	5,4 % (8)

Anmerkung. Mehrfachnennungen waren möglich. Kategorien mit weniger als sechs Nennungen wurden unter „Sonstiges“ zusammengefasst.

Wissen zu Rückmeldungen. Die Studierenden nannten im Durchschnitt 1.55 Möglichkeiten ($SD = 1.09$; $Range = 0-4$), um Lernenden im naturwissenschaftlichen Grundschulunterricht eine Rückmeldung geben zu können. Meistens wurde sich dabei allgemein auf schriftliche und verbale Möglichkeiten bezogen, wenige Studierende nannten hier bereits konkrete Methoden wie z. B. das Verfassen eines Kommentars. Dagegen beschrieben jedoch ein Drittel aller Studierenden die Möglichkeit von Noten als eine Form der Rückmeldung (27 %, s. Tabelle 2). Naturwissenschaftsspezifische Möglichkeiten wurden von den Studierenden so gut wie gar nicht genannt, lediglich 4,7 % bezogen sich auf das Experiment als Rückmeldemöglichkeit. Diesbezüglich wurde von einer Studierenden z. B. beschrieben, dass mithilfe von Experimenten (erneut) verdeutlicht werden könnte, dass eine Äußerung von Lernenden wissenschaftlich nicht angemessen ist.

Die Studierenden brachten zudem selbst verschiedene Möglichkeiten der inhaltlichen Gestaltung von Rückmeldungen an. Am häufigsten wurden korrigierende Rückmeldungen beschrieben (21,6 %), am zweithäufigsten das Loben von Lernenden (17,6%).

Tabelle 2: Kategorien und Häufigkeit der beschriebenen Möglichkeiten der Rückmeldung in Bezug auf Methoden und Gestaltung

Möglichkeiten der Rückmeldung	% (Anzahl Personen)
Fachübergreifende Möglichkeiten	
Schriftliche Rückmeldung (z. B. kurze Texte)	47,3 % (70)
Verbale Rückmeldung (z. B. in einem Gespräch)	62,8 % (93)
Non-verbale Rückmeldung (z. B. Mimik und Gestik)	8,8 % (13)
Selbsteinschätzung im Sinne des Self-Assessments	5,4 % (8)
Einschätzung durch Mitschüler*innen im Sinne des Peer-Assessments	4,1 % (6)
Noten	27 % (40)
Naturwissenschaftsspezifische Möglichkeit	
Experimente (z. B. als Reaktion auf wissenschaftlich nicht angemessene Vorstellungen)	4,7 % (7)
Gestaltungsmöglichkeiten	
Sachliche Rückmeldung	12,2 % (18)
Implizite Rückmeldung (z. B. Gegenfrage stellen)	7,4 % (11)
Explizite Rückmeldung (z. B. gelungene Aspekte hervorheben)	3,4 % (5)
Bestätigende Rückmeldung (z. B. Lösung zustimmen)	16,2 % (24)
Korrigierende Rückmeldung (z. B. Fehler benennen)	21,6 % (32)
Lob	17,6 % (26)
Zeitlicher Einsatz (z. B. direkt nach einer Aussage, nach einer Stunde)	16,2 % (24)

Anmerkung: Mehrfachnennungen waren möglich.

Zusammenhang zwischen Wissen und Lehr-Lern-Überzeugungen. Zwischen dem Wissen zur Diagnostik und den konstruktivistischen Lehr-Lern-Überzeugungen zeigt sich ein geringer, aber signifikanter Zusammenhang ($r = .239, p = .003$). Kein Zusammenhang zeigt sich dagegen mit den transmissiven Überzeugungen ($r = -.037, p = .652$). Das Wissen zur Rückmeldung hängt weder mit konstruktivistischen ($r = .117, p = .158$) noch mit transmissiven Überzeugungen ($r = -.138, p = .095$) zusammen.

6 Diskussion

In der vorliegenden Studie wurde untersucht, über welches Wissen 148 Bachelorstudierende der Universität Koblenz-Landau (Standort Landau) zu Möglichkeiten der Diagnostik und der Rückmeldung im Kontext des naturwissenschaftlichen Grundschulunterrichts verfügen. Es wird davon ausgegangen, dass dieses Wissen einen bedeutsamen, wenn auch nicht hinreichenden Teil des professionellen Wissens zur

Umsetzung formativen Assessments darstellt (vgl. Kapitel 2.3). So wird z. B. in Anlehnung an Baumert und Kunter (2011) ergänzend u.a. fachdidaktisches Wissen zur Gestaltung von lernförderlichen Rückmeldungen sowie Fachwissen zur Interpretation bestehender Vorstellungen als bedeutsam erachtet.

Die Ergebnisse der vorliegenden Studie zeigen, dass nach dem Besuch einer Vorlesung bereits erstes Wissen zu Möglichkeiten der Diagnostik und der Rückmeldung vorhanden zu sein scheint, dieses allerdings häufig noch nicht ausdifferenziert sowie nicht immer formativ ausgerichtet ist. In Bezug auf die Diagnostik und die Rückmeldung beschreiben die Studierenden z. B. lediglich einen Bruchteil der Methoden, die in der Vorlesung genannt und z.T. veranschaulicht wurden. Eine Ursache dafür könnte sein, dass die Studierenden das diagnostische Potenzial verschiedener Methoden noch nicht vollends für sich erschlossen haben (Zucker & Leuchter, 2018). Die geringe Nennung des Self- und Peer-Assessments könnte zudem darauf zurückgeführt werden, dass ein Bewusstsein für die Möglichkeiten von deren Einsatz bereits in der Grundschule noch nicht ausreichend ausgeprägt ist und/oder diesbezüglich Erfahrungen aus der eigenen Schulzeit fehlen.

Ggf. fällt in beiden Kontexten zusätzlich auch die Übertragung auf den naturwissenschaftlichen Grundschulunterricht schwer, da aufgrund der Interdisziplinarität des Faches Sachunterricht (Kleickmann, 2015) der Fokus in der Vorlesung lediglich einmal auf die naturwissenschaftliche Perspektive gelegt wurde. So benennen zwar ein Drittel der Studierenden das Experimentieren als Möglichkeit zur Diagnostik, aber lediglich 17,6 % konkretisieren dies durch die Beachtung spezifischer Denk-, Arbeits- und Handlungsweisen. In Bezug auf die Rückmeldung findet ein spezifischer Übertrag auf den naturwissenschaftlichen Grundschulunterricht so gut wie gar nicht statt.

Die beschriebenen Möglichkeiten der Studierenden zur Diagnostik und Rückmeldung deuten an einigen Stellen bereits auf eine formative Ausrichtung hin (z. B. Diagnostik: Lernportfolios, Unterrichtsgespräche; Rückmeldung: direkter zeitlicher Einsatz, mündliche Rückmeldung). Allerdings ist nicht bei allen beschriebenen Möglichkeiten identifizierbar, inwiefern eine formative Ausrichtung tatsächlich mitgedacht wurde. Im Kontext der Diagnostik wird z. B. der Test als zweithäufigste Methode beschrieben. Tests waren in der Vorlesung auch besonders fokussiert worden, um die formative Nutzung zu veranschaulichen. Inwiefern diese allerdings in der vierwöchigen späteren Befragung noch mitgedacht wurde, ist unklar. Studien haben gezeigt, dass sich (angehende) Lehrpersonen ohne entsprechende Einordnung bei Tests primär an einer summativen Überprüfung orientieren (Zulliger & Buholzer, 2020). Im Kontext der Rückmeldungen fällt der verhältnismäßig hohe Anteil an Nennungen von Noten auf, obwohl diese in der Vorlesung nicht thematisiert worden waren. Noten können in Anlehnung an Bürgermeister und Saalbach (2018) als rein summativer Ansatz gewertet werden, weil durch ihren Einsatz nicht von einer Optimierung der Lernprozesse ausgegangen werden kann. Gleiches gilt für Bekräftigungen ohne Bezug zur erbrachten Leistung im Sinne eines Lobs (Lipowsky, 2015). Ggf. hätte der Auf-

bau entsprechenden „Abgrenzungswissens“ (Oser et al., 1999) ein Verständnis des formativen Assessments für die Studierenden ausbauen können.

In der vorliegenden Studie wurde zusätzlich untersucht, wie das Wissen der Studierenden mit ihren Lehr-Lern-Überzeugungen zusammenhängt. Die Ergebnisse weisen diesbezüglich auf einen positiven Zusammenhang zwischen dem Wissen zur Diagnostik und den konstruktivistischen Lehr-Lern-Überzeugungen hin. Ein Zusammenhang zum Wissen zur Rückmeldung besteht nicht. Aus theoretischer Sicht lässt sich dieses Ergebnis anhand des Grundgedankens des Konstruktivismus erklären. Unter diesem wird Lernen als aktives Umkonstruieren bereits vorhandener Konzepte zusammengefasst (Möller, 2018). Um dieses passend begleiten zu können, ist insbesondere das stetige Diagnostizieren von Konzepten eine notwendige Voraussetzung. Konstruktivistische Vorstellungen könnten daher ein nützlicher Filter insbesondere für den Aufbau von Wissen zur Diagnostik sein (Levin, 2015) bzw. ggf. auch umgekehrt (Buehl & Fives, 2009). Transmissive Lehr-Lern-Überzeugungen scheinen dagegen in diesem Fall keinen Filter darzustellen.

Die zuvor diskutierten Ergebnisse sind vor dem Hintergrund verschiedener Limitationen zu betrachten. Zwei Limitationen betreffen u.a. die Erfassung des Wissens. Erstens wurde das Wissen anhand von zwei offen gestellten Fragen erfasst, bei denen die Studierenden den Fokus auf verschiedene Möglichkeiten legen konnten. Fokussierte Fragen (z. B. in Bezug auf Methoden oder den Zeitpunkt, ggf. eingebettet in den Kontext des formativen Assessments) hätten das Wissen der Studierenden voraussichtlich umfassender abbilden können. Zweitens erfolgte die Erfassung des Wissens in zwei verschiedenen Semestern. Dabei wurde zwar die identische Vorlesung von der gleichen Dozierenden umgesetzt, eine explizite Kontrolle der Umsetzung wurde allerdings nicht durchgeführt. Auch die Berücksichtigung weiterer Kontroll- und Störvariablen erfolgte nicht. Ein Einsatz vor und nach der Vorlesung hätte zudem spezifischere Hinweise auf die Wirksamkeit der Vorlesung geben können. Eine zusätzliche Limitation betrifft die untersuchte Stichprobe. In der Studie wurden Studierende in ihrem Bachelor an der Universität Koblenz-Landau befragt. Inwiefern die vorliegenden Ergebnisse auch für andere Hochschulstandorte repräsentativ sind, müsste zukünftig geprüft werden.

7 Implikationen

Der Ausbau professionellen Wissens zum formativen Assessment wird in der Lehrkräfteausbildung als bedeutsam erachtet (Bürgermeister & Saalbach, 2018; Zulliger & Buholzer, 2020). Aus den Ergebnissen der vorliegenden Studie lassen sich folgende Empfehlungen ableiten:

- (1) Die Ergebnisse deuten darauf hin, dass eine Sitzung im Rahmen einer Vorlesung nicht für den Ausbau entsprechenden Wissens zum formativen Assessment ausreichend ist. Vielmehr wird wahrscheinlich die Vertiefung in weiteren Vorlesungssitzungen oder bestenfalls in einem Seminar benötigt. Dies könnte

die Möglichkeit schaffen, Wissen stärker auszudifferenzieren (z. B. Rückmeldungen in Abhängigkeit individueller Voraussetzungen generieren, formative Nutzung von eher summativ angelegten Assessments abgrenzen) sowie dieses in den Perspektiven des naturwissenschaftlichen und sozialwissenschaftlichen Grundschulunterrichts zu konkretisieren.

- (2) Der Aufbau von „Abgrenzungswissen“ (Oser et al., 1999) kann zusätzlich als förderlich angenommen werden. Insbesondere im Kontext der Rückmeldung scheint dies hinsichtlich des Einsatzes von Noten sowie von Lob bedeutsam. Im Kontext der Diagnostik könnten zudem verschiedene Potenziale (z. B. der Diagnostik, der Lernunterstützung etc.) von Äußerungen und Tätigkeiten abgegrenzt werden.
- (3) Theorie und Praxis sollten über einzelne Veranschaulichungen hinaus stärker miteinander verknüpft werden, damit Wissen nachhaltig in verschiedenen Situationen anwendbar wird (Blömeke et al., 2015). Dies könnte z. B. dadurch ermöglicht werden, dass die Studierenden Methoden eigenständig erproben (Bürgermeister & Saalbach, 2018) – bspw. durch den Einsatz von Self- und Peer-Assessments untereinander oder in eigenen Lehr-Lern-Situationen mit Kindern (u. a. im Rahmen von Kinder-Unis, Kurzpraktika).
- (4) Zusätzlich zum Wissen sollten auch weitere Kompetenzfacetten der Studierenden wie ihre Lehr-Lern-Überzeugungen berücksichtigt werden und der Aufbau konstruktivistischer Überzeugungen gestärkt werden. Ggf. könnten Studierende diesbezüglich beim Aufbau eines metakognitiven Bewusstseins unterstützt werden, um den Einfluss von Lehr-Lern-Überzeugungen auf den eigenen Kompetenzausbau zu reflektieren (Levin, 2015).

Ergänzend zu den genannten hochschuldidaktischen Empfehlungen ergeben sich verschiedene Desiderate für die Forschung. So erscheint es u. a. für die Entwicklung passender Lehrveranstaltungen zum formativen Assessment bedeutsam, neben dem Wissen zu Möglichkeiten der Diagnostik und der Rückmeldung auch das Vorliegen weiteren fachdidaktischen Wissens sowie Fachwissens bei Studierenden zu untersuchen. Gleiches gilt auch für eine entsprechende Anwendung des professionellen Wissens in eigenen Lehr-Lern-Situationen der Studierenden, bei dem zusätzlich der konkrete Einfluss konstruktivistischer Lehr-Lern-Überzeugungen betrachtet werden könnte. Zur entsprechenden Förderung des professionellen Wissens in der universitären Lehrkräfteausbildung wäre ausgehend von den genannten Empfehlungen abschließend interessant, wie eine Lehrveranstaltung konkret konzipiert werden sollte (u. a. Umfang, Anwendungssituationen), um professionelles Wissen gewinnbringend fördern zu können.

Förderhinweis: Die diesem Artikel zugrundeliegende Studie wurde im Rahmen des Projekts „Schülerlabore als Lehr-Lern-Labore – Forschungsorientierte Verknüpfung von Theorie und Praxis in der MINT-Lehramtsausbildung“ von der *Deutsche Telekom Stiftung* gefördert (Laufzeit: 2014–2018).

Literatur

- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Waxmann. https://doi.org/10.1007/978-3-658-00908-3_13
- Birenbaum, M., Kimron, H. & Shilton, H. (2011). Nested contexts that shape assessment “for” learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation*, 37, 35–48. <https://doi.org/10.1016/j.stueduc.2011.04.001>
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Brunner, M., Kunter, M., Krauss, S., Klusmann, U., Baumert, J., Blum, W., Neubrand, M., Dubberke, T., Jordan, A., Löwen, K. & Tsai, Y.-M. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht. Eine Zwischenbilanz des COACTIV-Projekts. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 54–82). Waxmann.
- Buehl, M. M. & Fives, H. (2009). Exploring teachers’ beliefs about teaching knowledge: Where does it come from? Does it change? *The Journal of Experimental Education*, 77(4), 367–407. <https://doi.org/10.3200/JEXE.77.4.367-408>
- Bürgermeister, A. & Saalbach, H. (2018). Formatives Assessment: Ein Ansatz zur Förderung individueller Lernprozesse. *Psychologie in Erziehung und Unterricht*, 65(3), 194–205. <https://doi.org/10.2378/peu2018.art11d>
- Cowie, B. & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101–116. <https://doi.org/10.1080/09695949993026>
- Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., Lühken, A., Adl-Amini, K., Djakovic, S.-K., Mannel, S., Naumann, A. & Hardy, I. (2015). Impact of additional guidance in science education on primary students’ conceptual understanding. *The Journal of Educational Research*, 108(5), 358–370. <https://doi.org/10.1080/00220671.2014.899957>
- Heitink, M. C., van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K. & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50–62. <https://doi.org/10.1016/j.edurev.2015.12.002>
- Hondrich, A. L., Decristan, J., Hertel, S. & Klieme, E. (2018). Formative assessment and intrinsic motivation: The mediating role of perceived competence. *Zeitschrift für Erziehungswissenschaft*, 21(4), 717–734. <https://doi.org/10.1007/s11618-018-0833-z>
- Hondrich, A. L., Hertel, S., Adl-Amini, K. & Klieme, E. (2016). Implementing curriculum-embedded formative assessment in primary school science classrooms. *Assessment in Education: Principles, Policy & Practice*, 23(3), 353–376. <https://doi.org/10.1080/0969594X.2015.1049113>
- Jonen, A., Hardy, I. & Möller, K. (2003). Schwimmt ein Holzbrett mit Löchern? – Erklärungen von Kindern zum Schwimmen und Sinken verschiedener Gegenstände vor und nach dem

- Unterricht. In A. Speck-Hamdan, H. Brügelmann, M. Fölling-Albers & S. Richter (Hrsg.), *Kulturelle Vielfalt. Religiöses Lernen* (S. 159–164). Kallmeyersche Verlagsbuchhandlung.
- Kleickmann, T. (2008). *Zusammenhänge fachspezifischer Vorstellungen von Grundschullehrkräften zum Lehren und Lernen mit Fortschritten von Schülerinnen und Schülern im konzeptuellen naturwissenschaftlichen Verständnis* (Dissertation, Westfälische Wilhelms-Universität). Publikationsserver der WWU Münster. https://repositorium.uni-muenster.de/document/miami/642aa4ce-7149-4cdb-a938-c37f3c64cbe2/diss_kleickmann.pdf
- Kleickmann, T. (2015). Professionelle Kompetenz von Primarschullehrkräften im Bereich des naturwissenschaftlichen Sachunterrichts. *Zeitschrift für Grundschulforschung*, 8(1), 7–22.
- Klute, M., Apthorp, H., Harlacher, J. & Reale, M. (2017). *Formative assessment and elementary school student academic achievement: A review of the evidence* (REL2017–259). U.S. Department of Education, Institute of Education Sciences.
- Levin, B. B. (2015). The development of teachers' beliefs. In H. Fives & G. M. Gill (Hrsg.), *International Handbook of Research on Teachers' Beliefs* (S. 48–65). Routledge.
- Lipowsky, F. (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), *Springer-Lehrbuch. Pädagogische Psychologie* (S. 69–105). Springer. https://doi.org/10.1007/978-3-642-41291-2_4
- MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, 20(5), 523–535. <https://doi.org/10.1016/j.tate.2004.04.008>
- Meschede, N., Fiebranz, A., Möller, K. & Steffensky, M. (2017). Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. *Teaching and Teacher Education*, 66, 158–170. <https://doi.org/10.1016/j.tate.2017.04.010>
- Möller, K. (2018). Die Bedeutung von Schülervorstellungen für das Lernen im Sachunterricht. In M. Adamina, M. Kübler, K. Kalcsics, S. Bietenhard & E. Engeli (Hrsg.), *„Wie ich mir das denke und vorstelle...“ – Vorstellungen von Schülerinnen und Schülern zu Lerngegenständen des Sachunterrichts und des Fachbereichs Natur, Mensch, Gesellschaft* (S. 35–50). Klinkhardt.
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Hrsg.), *Handbook of Research on Educational Communications and Technology* (S. 745–783). Lawrence Erlbaum.
- Oser, F., Hascher, T. & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie des „negativen“ Wissens. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 11–41). Leske+Budrich. https://doi.org/10.1007/978-3-663-07878-4_1
- Praetorius, A.-K., Rogh, W. & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft*, 48(3), 303–318. <https://doi.org/10.1007/s42010-020-00072-w>
- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B. & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 101602. <https://doi.org/10.1016/j.ijer.2020.101602>
- Schmidt, C. (2018). Formatives Assessment als Aspekt professioneller diagnostischer Kompetenz im Kontext des Grundschulunterrichts. In S. Miller, B. Holler-Nowitzki, B. Kottmann, S. Lesemann, B. Letmathe-Henkel, N. Meyer, R. Schroeder & K. Velten (Hrsg.), *Profession und Disziplin. Grundschulpädagogik im Diskurs* (S. 158–163). Springer VS. https://doi.org/10.1007/978-3-658-13502-7_16
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>

- Voss, T., Kleickmann, T., Kunter, M. & Hachfeld, A. (2011). Überzeugungen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 235–257). Waxmann.
- Weurlander, M., Söderberg, M., Scheja, M., Hult, H. & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: Students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education*, 37(6), 747–760. <https://doi.org/10.1080/02602938.2011.572153>
- Yan, Z., Chiu, M. M. & Keung Cheng, E. C. (2022). Predicting teachers' formative assessment practices: Teacher personal and contextual factors. *Teaching and Teacher Education*, 114, 103718. <https://doi.org/10.1016/j.tate.2022.103718>
- Zucker, V. & Leuchter, M. (2018). Die Fähigkeit von Studierenden im Grundschullehramt, Formative Assessment im naturwissenschaftlichen Sachunterricht zu erkennen. In U. Franz, H. Giest, A. Hartinger, A. Heinrich-Dönges & B. Reinhoffer (Hrsg.), *Handeln im Sachunterricht* (S. 175–182). Klinkhardt.
- Zulliger, S. & Buholzer, A. (2020). Pädagogisch-psychologisches Wissen von Lehrpersonen zu formativem Assessment. *Erziehung & Unterricht. Österreichische Pädagogische Zeitschrift*, 170(9–10), 753–764.

Autorinnen und Autoren

Arnold, Julia, Prof. Dr., seit 2023 Leiterin des Zentrums Naturwissenschafts- und Technikdidaktik (ZNTD) der Pädagogischen Hochschule FHNW. Forschungs- und Arbeitsschwerpunkte: Lernunterstützungen und Forschendes Lernen, Digitale Kompetenzen (angehender) Lehrpersonen sowie Themen an den Schnittstellen von Wissenschaft, Gesundheit und Nachhaltigkeit.

Anschrift: Pädagogische Hochschule, Fachhochschule Nordwestschweiz, Hofackerstrasse 30, 4132 Muttenz (Schweiz) / E-Mail: julia.arnold@fhnw.ch

Bachmann, Patricia, M.A., seit 2015 wissenschaftliche Mitarbeiterin und Lehrbeauftragte an der Pädagogischen Hochschule St. Gallen. Forschungs- und Arbeitsschwerpunkte: Formative Beurteilung, Lehr-/Lernforschung, Professionsforschung, Ausbildung von Lehrpersonen.

Anschrift: Pädagogische Hochschule St. Gallen, Institut Professionsforschung und Kompetenzentwicklung, Notkerstrasse 27, 9000 St. Gallen (Schweiz) / E-Mail: patricia.bachmann@phsg.ch

Barzel, Bärbel, Prof. Dr., seit 2013 Professorin für Mathematikdidaktik an der Universität Duisburg-Essen, seit 2021 Co-Vorsitz des Interdisziplinären Zentrums für Bildungsforschung (Izfb) an der Universität Duisburg-Essen und Netzwerkpartnerin des Deutschen Zentrums für Lehrkräftebildung Mathematik (DZLM). Forschungs- und Arbeitsschwerpunkte: Online-Diagnostik, Einsatz digitaler Medien im Mathematikunterricht, Lehrerprofessionalisierung.

Anschrift: Universität Duisburg-Essen, Thea-Leymann-Straße 9, 45127 Essen / E-Mail: baerbel.barzel@uni-due.de

Boegel, Svenja, M.Ed., seit 2022 Doktorandin in der Arbeitsgruppe von Prof. Dr. Mathias Ropohl an der Universität Duisburg-Essen in der Didaktik der Chemie. Forschungs- und Arbeitsschwerpunkte: Lernen in Experimentiersituationen, Einfluss affektiver Schüler:innenmerkmale beim Lernen

Anschrift: Universität Duisburg-Essen, Schützenbahn 70, 45127 Essen / E-Mail: svenja.boegel@uni-due.de

Brovelli, Dorothee, Prof. Dr., Prorektorin Forschung und Entwicklung und Professorin für Didaktik der Naturwissenschaften/Physik an der Pädagogischen Hochschule Luzern sowie Honorarprofessorin an der Pädagogischen Hochschule Heidelberg. Forschungs- und Arbeitsschwerpunkte: professionelle Kompetenzen von Lehramtsstudierenden, gendergerechter Physikunterricht, Kontextorientierung und Digitalisierung in den MINT-Fächern.

Anschrift: Pädagogischen Hochschule Luzern, Sentimatt 1, 6003 Luzern (Schweiz) / E-Mail: dorothee.brovelli@phlu.ch

Buholzer, Alois, Prof. Dr., seit 2006 Leiter des Instituts für Diversität und inklusive Bildung (IDB) im Leistungsbereich Forschung und Entwicklung und ab 2007 Professor für Bildungs- und Sozialwissenschaften an der Pädagogischen Hochschule Luzern. Forschungs- und Arbeitsschwerpunkte: Pädagogische Diagnostik, Diversität in Schule und Unterricht, Integration und Inklusion.

Anschrift: Pädagogische Hochschule Luzern, Sentimatt 1, 6003 Luzern (Schweiz) / E-Mail: alois.buholzer@phlu.ch

Bürgermeister, Anika, Dr., seit 2015 wissenschaftliche Mitarbeiterin für Pädagogische Psychologie mit dem Schwerpunkt Lehren, Lernen und Entwicklung an der Universität Leipzig. Forschungs- und Arbeitsschwerpunkte: Formatives Assessment, frühe Bildung, soziale Interaktionen in Lehr-Lernprozessen und Higher Education.

Anschrift: Universität Leipzig, Marschnerstraße 31, 04109 Leipzig / E-Mail: anika.buergermeister@uni-leipzig.de

Dober, Heidi, M.A., seit 2018 Wissenschaftliche Mitarbeiterin für SNF-Forschungsprojekt FEMAR an der Pädagogischen Hochschule Zug. Forschungs- und Arbeitsschwerpunkte: Ausbildung und Forschung in Fachdidaktik Mathematik, Mathematisches Argumentieren, Feedback.

Anschrift: Pädagogische Hochschule Zug, Zugerbergstrasse 3, 6300 Zug (Schweiz) / E-Mail: heidi.dober@phzg.ch

Dohmen, Yannick, seit 2019 Student der angewandten Informatik mit dem Schwerpunkt Systems Engineering an der Universität Duisburg-Essen.

Anschrift: E-Mail: yannick.dohmen@stud.uni-due.de

Glogger-Frey, Inga, Prof. Dr., seit 2021 Inhaberin der Professur für Bildungspsychologie (Fachgebiet Psychologie) an der Universität Erfurt. Forschungs- und Arbeitsschwerpunkte: Diagnosekompetenz von Lehrkräften, Selbstreguliertes Lernen, Beispielbasiertes Lernen, Lehr-Lernmethoden zum Aufgreifen von Vorwissen.

Anschrift: Universität Erfurt, Fachgebiet Psychologie, Nordhäuser Straße 63, 99089 Erfurt / E-Mail: inga.frey@uni-erfurt.de

Graewert, Laura, M.Ed., seit 2021 wissenschaftliche Mitarbeiterin in der Fachgruppe Didaktik der Mathematik an der Universität Siegen. Forschungs- und Arbeitsschwerpunkte: Digitales formatives (Selbst-)Assessment, Game-Based-Learning.

Anschrift: Universität Siegen, Adolf-Reichwein-Str. 2, 57076 Siegen / E-Mail: laura.graewert@uni-siegen.de

Grob, Urs, Dr., Dozent und Spezialist für quantitative Methoden am Institut für Erziehungswissenschaft der Universität Zürich. Aktuelle Forschungs- und Arbeitsschwerpunkte: Unterrichtsforschung, Kontexteffekte, Datenanalyseverfahren.

Anschrift: Universität Zürich, Institut für Erziehungswissenschaft, Freiestrasse 36, 8032 Zürich (Schweiz) / E-Mail: grob@ife.uzh.ch

Hardy, Ilonca, Prof. Dr., seit 2008 Professorin für Erziehungswissenschaft mit dem Schwerpunkt Grundschulpädagogik/Empirische Bildungsforschung am Institut für Pädagogik der Elementar- und Primarstufe, Goethe-Universität Frankfurt. Forschungs- und Arbeitsschwerpunkte: Frühe naturwissenschaftliche Bildung, Adaptivität von Unterricht, Sprachbildung und sprachsensibler Unterricht, Förderung.

Anschrift: Goethe-Universität Frankfurt a.M., Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main / E-Mail: Hardy@em.uni-frankfurt.de

Hess, Kurt, Prof. Dr., seit 2007 Professor für Mathematisches Denken und Lernen und ab 2021 Co-Leiter des gleichnamigen Kompetenzzentrums an der Pädagogischen Hochschule Zug. Forschungs- und Arbeitsschwerpunkte: Frühe mathematische Bildung, mathematisches Argumentieren, Schwierigkeiten beim Mathematiklernen, integrativer und inklusiver Mathematikunterricht, dialogisches und kooperatives Lernen.

Anschrift: Zugerbergstrasse 3, 6300 Zug (Schweiz) / E-Mail: kurt.hess@phzg.ch

Hofmann, Florian, Dr., Akademischer Oberrat am Lehrstuhl für Schulpädagogik mit dem Schwerpunkt empirische Unterrichtsforschung an der Friedrich-Alexander-Universität Erlangen-Nürnberg; Koordinator des Promotionsprogrammes Lehr-Lernforschung und Studienfachkoordinator Schulpädagogik. Forschungs- und Arbeitsschwerpunkte: Pädagogische Diagnostik mit dem Schwerpunkt zeitgemäße Förder- und Adaptionsdiagnostik, Professionalität und Professionalisierung von Lehrkräften, Emotionen und Wohlbefinden in Schule und Hochschule.

Anschrift: Friedrich-Alexander-Universität Erlangen-Nürnberg, Institut für Erziehungswissenschaft, Lehrstuhl für Schulpädagogik mit dem Schwerpunkt empirische Unterrichtsforschung, Regensburger Straße 160, 90478 Nürnberg / E-Mail: florian.hofmann@fau.de

Hußmann, Stephan, Prof. Dr., seit 2005 Professor für Didaktik der Mathematik der Sekundarstufen an der TU Dortmund und seit 2011 Leiter des Dortmunder Kompetenzzentrums für Lehr-/Lernforschung und Lehrkräftebildung (DoKoLL). Forschungs- und Arbeitsschwerpunkte: Gestaltungsorientierte Forschung zur Entwicklung von sinnstiftenden Lernumgebungen in Schule, Hochschule und Fortbildung, Inklusionsorientierung im Mathematikunterricht der Sekundarstufen, Diagnostik und Förderung.

Anschrift: Technische Universität Dortmund, Institut für Entwicklung und Erforschung des Mathematikunterrichts (IEEM), Vogelpothsweg 87, 44227 Dortmund / E-Mail: stephan.hussmann@tu-dortmund.de

Junge, Katharina, Dipl.-Päd., seit 2020 Wissenschaftliche Mitarbeiterin an der Universität Hamburg. Forschungs- und Arbeitsschwerpunkte: Frühe naturwissenschaftliche Bildung, naturwissenschaftliche Lernprozesse in der frühkindlichen häuslichen Lernumgebung.

Anschrift: Universität Hamburg, Von-Melle-Park 8, 20146 Hamburg / E-Mail: katharina.junge@uni-hamburg.de

Kampa, Nele, Prof. Dr., seit 2022 Professorin für Schulpädagogik unter besonderer Berücksichtigung der Sekundarstufe. Forschungs- und Arbeitsschwerpunkte: Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern, Heterogenität von Lernenden und Lehrkräften, Kompetenzen in der Sekundarstufe II.

Anschrift: Universität Wien, Zentrum für Lehrer*innenbildung & Fakultät für Bildungswissenschaft, Porzellangasse 4, 1090 Wien (Österreich) / E-Mail: nele.kampa@univie.ac.at

Klotz, Christian, Sekundarstufenlehrer, seit 2020 Abordnung an die Pädagogische Hochschule Schwäbisch Gmünd an das Institut für Erziehungswissenschaft. Forschungs- und Arbeitsschwerpunkte: Leistungsmessung, Evaluation digitaler Lernsysteme, Formatives Assessment.

Anschrift: Pädagogische Hochschule Schwäbisch Gmünd, Oberbettringerstraße 200, 73525 Schwäbisch Gmünd / E-Mail: christian.klotz@ph-gmuend.de

Kraler, Christian, Univ.-Prof. Dr., Professur für Lehrer/innenbildung und Lernen an der Universität Innsbruck. Arbeits- und Forschungsschwerpunkte: Aspekte (internationaler) Lehrer/innenbildung, Lernen im Kontext formaler Bildung, Grundlagen und Umsetzung formaler Bildungssysteme, Modellierung von Lernprozessen.

Anschrift: Universität Innsbruck, Institut für LehrerInnenbildung und Schulforschung, Innrain 52a, 6020 Innsbruck (Österreich) / E-Mail: christian.kraler@uibk.ac.at

Leuchter, Miriam, Prof. Dr., seit 2016 Professorin für Grundschulpädagogische Forschung mit dem Schwerpunkt Sachunterricht am Institut für Bildung im Kindes- und Jugendalter an der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau. Forschungs- und Arbeitsschwerpunkte: Frühes naturwissenschaftliches Lernen, digitale Bildung und professionelle Kompetenzen von Studierenden, pädagogischen Fachkräften und Lehrkräften.

Anschrift: Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Institut für Bildung im Kindes- und Jugendalter, August-Croissant-Str. 5, 76829 Landau / E-Mail: miriam.leuchter@rptu.de

Lötscher, Hanni, lic. phil., Studienbereichsleiterin Bildungs- und Sozialwissenschaften an der Pädagogischen Hochschule Luzern. Arbeitsschwerpunkte: Diagnose- und Reflexion von Lehr-Lernprozessen.

Anschrift: Pädagogische Hochschule Luzern, Sentimatt 1, 6003 Luzern (Schweiz) / E-Mail: hanni.loetscher@phlu.ch

MacKay-Falls, Fiona, B.A. (Hons), M.Sc., seit 2011 wissenschaftliche Mitarbeiterin, seit 2017 Senior Lecturer am Institut für LehrerInnenbildung & Schulforschung an der Universität Innsbruck. Arbeits- und Forschungsschwerpunkte: Pädagogische Diagnostik, Lernforschung.

Anschrift: Universität Innsbruck, Institut für LehrerInnenbildung und Schulforschung, Innrain 52a, 6020 Innsbruck (Österreich) / E-Mail: fiona.mac-kay@uibk.ac.at

Maier, Uwe, Prof. Dr., seit 2012 Professor für Erziehungswissenschaft mit Schwerpunkt empirische Schulforschung an der Pädagogischen Hochschule Schwäbisch Gmünd. Forschungs- und Arbeitsschwerpunkte: Unterrichtsplanung, Allgemeine Didaktik, Formatives Assessment, Datenbasierte Schulentwicklung, Adaptive Lernsysteme.

Anschrift: Pädagogische Hochschule Schwäbisch Gmünd, Oberbettringerstraße 200, 73525 Schwäbisch Gmünd / E-Mail: uwe.maier@ph-gmuend.de

Meschede, Nicola, Prof. Dr., seit 2018 Professorin für Naturwissenschaftliche Elementar- und Primärerziehung am Institut für Didaktik des Sachunterrichts der Universität Münster. Forschungs- und Arbeitsschwerpunkte: Naturwissenschaftliches Lehren und Lernen im Elementar- und Primarbereich, Professionelle Kompetenz von (angehenden) Sachunterrichtslehrpersonen, Einsatz von Unterrichtsvideos in der Lehrerbildung, Lernprozessunterstützung und Umgang mit Heterogenität.

Anschrift: Universität Münster, Institut für Didaktik des Sachunterrichts, Leonardo-Campus 11, 48149 Münster / E-Mail: nicola.meschede@uni-muenster.de

Neitemeier, Annika, M.Ed., von 2021 bis 2022 wissenschaftliche Mitarbeiterin am Institut für Entwicklung und Erforschung des Mathematikunterrichts (IEEM) der Technischen Universität Dortmund und seit 2022 im Vorbereitungsdienst für das Lehramt an Berufskollegs am Zentrum für schulpraktische Lehrerausbildung Leverkusen.

Anschrift: E-Mail: a.neitemeier@gsbk.schulen-lev.de

Opfermann, Maria, Dr., seit 2020 wissenschaftliche Mitarbeiterin in der AG Grundschulforschung an der Bergischen Universität Wuppertal. Forschungs- und Arbeitsschwerpunkte: Kognitive Belastung beim Lernen, Pädagogisch-psychologische Diagnostik, Minderung pandemiebedingter Lernrückstände bei Grundschulkindern.

Anschrift: Bergische Universität Wuppertal, School of Education, Gaußstraße 20, 42119 Wuppertal / E-Mail: opfermann@uni-wuppertal.de

Ropohl, Mathias, Prof. Dr., seit 2017 Professor für Didaktik der Chemie an der Universität Duisburg-Essen. Forschungs- und Arbeitsschwerpunkte: Formatives Assessment beim forschenden Lernen, Lehren und Lernen mit digitalen Medien, Förderung professionellen Wissens von angehenden Lehrkräften.

Anschrift: Universität Duisburg-Essen, Schützenbahn 70, 45127 Essen / E-Mail: mathias.ropohl@uni-due.de

Ruelmann, Merle, M.Ed., wissenschaftliche Mitarbeiterin im Leistungsbereich Forschung und Entwicklung der Pädagogischen Hochschule Luzern und Doktorandin am Institut für Erziehungswissenschaft der Universität Zürich. Aktuelle Forschungs- und Arbeitsschwerpunkte: Lernunterstützung, formatives Assessment und Feedback.

Anschrift: Pädagogische Hochschule Luzern, Sentimatt 1, 6003 Luzern (Schweiz) / E-Mail: merle.ruelmann@phlu.ch

Saalbach, Henrik, Prof. Dr., Professor für Pädagogische Psychologie mit dem Schwerpunkt Lehren, Lernen und Entwicklung an der Universität Leipzig. Forschungs- und Arbeitsschwerpunkte: Funktion von Sprache und sprachlichen Fähigkeiten in der kognitiven und sozioemotionalen Entwicklung sowie in Lehr-Lernprozessen, Lernen und Entwicklung im Kontext von Mehrsprachigkeit.

Anschrift: Universität Leipzig, Marschnerstraße 31, 04109 Leipzig / E-Mail: henrik.saalbach@uni-leipzig.de

Schreiner, Claudia, Ass.-Prof. Dr., Assistenzprofessorin am Institut für LehrerInnenbildung & Schulforschung der Universität Innsbruck. Forschungs- und Arbeitsschwerpunkte: Pädagogische Diagnostik, Kompetenzorientierung und Bildungsstandards, Chancengerechtigkeit und evidenzorientierte Qualitätsentwicklung.

Anschrift: Universität Innsbruck, Institut für LehrerInnenbildung & Schulforschung, Innrain 52a, 6020 Innsbruck (Österreich) / E-Mail: claudia.schreiner@uibk.ac.at

Smit, Robbert, Dr., seit 2008 Dozent für Erziehungswissenschaften und wissenschaftlicher Mitarbeiter an der Pädagogischen Hochschule St. Gallen, Forschungs- und Arbeitsschwerpunkte: Differenzierter Unterricht, Formative Beurteilung, Ausbildung von Lehrpersonen in den Mathematik- und Naturwissenschaften.

Anschrift: Notkerstr. 27, 9000 St. Gallen (Schweiz) / E-Mail: robbert.smit@phsg.ch

Steffensky, Mirjam, Prof. Dr., seit 2020 Professorin für Erziehungswissenschaften (Didaktik der Chemie) Universität Hamburg. Forschungs- und Arbeitsschwerpunkte: Frühe naturwissenschaftliche Bildung, Professionelle Kompetenzen pädagogischer Fach- und Lehrkräfte, Prozessqualität.

Anschrift: Universität Hamburg, Von-Melle-Park 8, 20146 Hamburg / E-Mail: mirjam.steffensky@uni-hamburg.de

Thurm, Daniel, Jun.-Prof. Dr., seit 2021 Juniorprofessor für Mathematikdidaktik an der Universität Siegen. Forschungs- und Arbeitsschwerpunkte: Digitale Medien im Mathematikunterricht, Lehrkräfteprofessionalisierung, Überzeugungen von Mathematiklehrkräften, Game-Based-Learning.

Anschrift: Universität Siegen, Adolf-Reichwein-Straße 2, 57076 Siegen / E-Mail: daniel.thurm@uni-siegen.de

Venitz, Laura, Dr., seit 2018 Wissenschaftliche Mitarbeiterin im Arbeitsbereich Grundschulpädagogische Forschung mit dem Schwerpunkt Sachunterricht am Institut für Bildung im Kindes- und Jugendalter an der Rheinland-Pfälzischen Technischen Universität Kaiserslautern. Forschungs- und Arbeitsschwerpunkte: Frühes naturwissenschaftliches Lernen, selbstreguliertes Lernen im Vorschulalter, Professionelle Kompetenzen von pädagogischen Fachkräften.

Anschrift: Rheinland-Pfälzische Technische Universität Kaiserslautern Landau, August-Croissant-Str. 5, 76829 Landau / E-Mail: venitz.laura@rptu.de

Weimar, Hilda (geb. Scheuermann), Dr., seit 2022 abgeordnete Lehrkraft am Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik in der Abteilung Didaktik der Chemie. Forschungs- und Arbeitsschwerpunkte: Formatives Assessment, forschendes Lernen, Experimentierkompetenz.

Anschrift: IPN Kiel, Olshausenstraße 62, 24118 Kiel / E-Mail: weimar@leibniz-ipn.de

Zucker, Verena, Dr., seit 2020 wissenschaftliche Mitarbeiterin am Institut für Didaktik des Sachunterrichts der Universität Münster. Forschungs- und Arbeitsschwerpunkte: Videobasierte Unterrichtsforschung in der Lehrkräftebildung, Professionelle Unterrichtswahrnehmung von angehenden Sachunterrichtslehrpersonen, Formatives Assessment im naturwissenschaftlichen Grundschulunterricht.

Anschrift: Universität Münster, Institut für Didaktik des Sachunterrichts, Leonardo-Campus 11, 48149 Münster / E-Mail: verena.zucker@uni-muenster.de

Zulliger, Sandra, lic. phil., wissenschaftliche Mitarbeiterin und Projektleiterin im Leistungsbereich Forschung und Entwicklung der Pädagogischen Hochschule Luzern und Doktorandin der Universität Freiburg. Forschungs- und Arbeitsschwerpunkte: Pädagogische Diagnostik, Unterrichtsqualität und Evaluationen.

Anschrift: Pädagogische Hochschule Luzern, Sentimatt 1, 6003 Luzern (Schweiz) / E-Mail: sandra.zulliger@phlu.ch