

REFERIERTE BEITRÄGE

ZEITSCHRIFT FÜR BERUFS- UND WIRTSCHAFTSPÄDAGOGIK 118, 2022/4, 594–620

DOI 10.25162/ZBW-2022-0022

WIEBKE VORPAHL / JULIA WARWAS

Interpretation von und Einstellungen zu lernverlaufsdiagnostischen Informationen in der beruflichen Bildung

Explorative Interviewstudie mit angehenden und praktizierenden Lehrkräften

Interpretation of and Attitudes towards Learning Progress Assessment Data in Vocational Education

Explorative Interview Study with Pre- and In-Service Teachers

KURZFASSUNG: Digitale Assessments in der beruflichen Bildung erleichtern die Sammlung, Verknüpfung und grafische Verdichtung lernverlaufsdiagnostischer Daten. Eine Laut-Denken-Studie eruiert, wie Lehrkräfte Lernverlaufskurven interpretieren und welche didaktischen Konsequenzen sie daraus ziehen. Ergänzend geben Leitfadeninterviews Einblicke in die Einstellungen der 63 Proband:innen gegenüber Lernverlaufsdaten. Analysen auf Basis eines Kategoriensystems funktionaler Lesefähigkeiten dokumentieren, dass der Informationsgehalt des Kurvendiagramms auf den Stufen der Datenbeschreibung und der Schlussfolgerungen nicht voll ausschöpft wird. Die Befragten erkennen große Potentiale der Lernverlaufsdiagnostik bzgl. verbesserter Unterrichtsgestaltung, begegnen aber deren praktischer Umsetzung skeptisch. **Schlagworte:** Lernverlaufsdiagnostik, Dateninterpretation, Assessment-bezogene Erfahrungen und Einstellungen, Vergleich angehender und berufserfahrener Lehrpersonen, Berufliche Bildung

ABSTRACT: Digital assessments in vocational education facilitate the collection, integration, and graphical aggregation of Learning Progress Assessment Data (LPAD). A think-aloud-study explores how teachers interpret learning progress curves and which didactic conclusions they draw. In addition, guided interviews elucidate the attitudes of the 63 study participants towards LPAD. Analyses based on a category system of functional reading skills document that diagnostic information inherent in the curve diagram was not fully extracted in the steps of 'reading the data' and 'reading beyond the data'. The interviewees see great potentials of LPAD for improved instructional design but are skeptical about their practical implementation.

Keywords: Learning Progress Assessment Data, Data Interpretation, Assessment-Related Experiences and Attitudes, Comparison of Pre- and In-Service Teachers, Vocational Education

1. Ausgangslage und Zielstellung

Diagnostische Fähigkeiten von Lehrpersonen stellen eine notwendige Bedingung für die Lern- und Leistungsförderung ihrer Schüler:innen dar, etwa für die flexible Anpassung des Unterrichts an akute Verständnisprobleme oder thematische Interessiertheit (SCHRADER, 2013; BEHRMANN/SOUVIGNIER, 2013; WARWAS/KÄRNER/GOLYSZNY, 2015). Dementsprechend bedeutend scheinen Untersuchungen zur Frage, wie Lehrkräfte insbesondere solche diagnostischen Daten lesen und verwerten, die Lernentwicklungsverläufe ihrer Schüler:innen abbilden sollen, um noch während einer Lernphase (z. B. größere Unterrichtseinheit) lernförderliche Maßnahmen (z. B. didaktische Variation) ergreifen zu können. Für den berufsbildenden Kontext gibt es zwar erste wertvolle Bestandsaufnahmen zu IT-basierten Lern- und Assessmentsystemen, die Lehrpersonen bei dieser anspruchsvollen Aufgabe informatorisch unterstützen könnten (z. B. SCHUMANN, 2021) sowie Entwicklungsarbeiten für lernbegleitende Diagnoseinstrumente (z. B. BERDING, 2020; KÄRNER/WARWAS/SCHUMANN, 2021; WARWAS/SCHIRMER/HELM, 2020), aber nur wenige Erkenntnisse zur lehrendenseitigen Interpretation und Nutzung von Lernverlaufsdaten.

Lernverlaufsdaten gehören zu den gängigen Darstellungsformen sog. formativer Assessmentsysteme (SCHÜTZE et al., 2018). Die für ihre Nutzung relevanten diagnostischen Fähigkeiten lassen sich unter das Dachkonzept der funktionalen Lesefähigkeit einordnen und weiter ausdifferenzieren (ZIEGLER/FREY/SEEGER/BALKENHOL/BERNHARDT, 2016; vertiefend Kapitel 2). Ferner haben sich in den bislang verfügbaren Studien zu datenbasierten Interpretationsleistungen von Lehrkräften aus dem allgemeinbildenden und sonderschulischen Bereich Laut-Denken-Protokolle als empirischer Zugriff bewährt (ESPIN et al., 2017; ZEUCH et al., 2017; WAGNER et al., 2017; BOSCH et al., 2017). Zu den auffälligsten Befunden dieser Studien gehört, dass sich praktizierende und angehende Lehrkräfte trotz unterrichtspraktischer Erfahrungsunterschiede bei der Interpretation der ihnen vorgelegten Lernverlaufsdaten nicht unterscheiden (z. B. ZEUCH et al., 2017). Stattdessen zeigte sich, dass alle Teilnehmenden häufig den wichtigen initialen Schritt der Beschreibung von Daten übersprangen (s. ZEUCH et al., 2017) und dazu neigten, Lernverlaufsdaten als Schulnoten zu deuten. Zudem beschränkten sich ihre abschließenden Ideen zu möglichen lernunterstützenden Maßnahmen auf allgemeine motivierende Anreize und blendeten (fach-)didaktische Strategien zur Förderung derjenigen Kenntnisse oder Fähigkeiten aus, auf die sich die Lernverlaufsdaten inhaltlich bezogen (ZEUCH et al., 2017). Verglichen mit den Laut-Denken-Protokollen von methodischen Expert:innen formativer Assessments erwiesen sich erwartungsgemäß die Protokolle von (praktizierenden) Lehrkräften als weniger vollständig und weniger genau (BOSCH et al., 2017).

Vor diesem Hintergrund kann die hier berichtete Studie als Replikationsstudie i. w. S. eingestuft werden, die mit ähnlichem Erhebungsdesign sowie einem kategorial verfeinerten Analyseraster auf Lernverlaufsdaten bezogene funktionale Lesefähigkeiten von angehenden und praktizierenden Lehrkräften an berufsbildenden Schulen untersucht. Dies erscheint nicht nur angesichts der Ausblendung eben jener Lehrpersonengruppe

in bisherigen Forschungsarbeiten, sondern auch angesichts der eingangs festgestellten Zunahme IT-basierter Lern- und Assessmentsysteme im beruflichen Bildungssektor als ein wichtiger Fokus berufs- und wirtschaftspädagogischer Forschung.

Aus diesem Grund ist es das vorrangige Ziel der vorliegenden Studie zu untersuchen, ob und inwieweit (angehende und praktizierende) Lehrkräfte an berufsbildenden Schulen die Informationen, die in diagrammförmig aufbereiteten Lernverlaufskurven enthalten sind, in einer der Daten angemessenen und erschöpfenden Weise interpretieren, und welche spontanen didaktischen Schlussfolgerungen sie hieraus ziehen. Ergänzend hierzu besteht ein weiteres Ziel darin, die Einstellung zu und bisherige Erfahrung mit ebenjenen Daten aufzuzeigen. Die Interpretation der aufbereiteten Lernverlaufskurve sowie die erfahrungs- und einstellungsbezogenen Aussagen der angehenden und praktizierenden Lehrkräfte werden schließlich vergleichend gegenübergestellt.

Dazu werden im theoretischen Teil das formative Assessment sowie die funktionale Lesefähigkeit näher betrachtet. Ergänzend zur Fähigkeit im Umgang mit (lernverlaufsdagnostischen) Daten werden bisherige Erkenntnisse über Einstellungen von Lehrkräften zur Datennutzung zusammengetragen. Anschließend wird das methodische Design der Untersuchung vorgestellt. Hierbei handelt es sich um die Think-Aloud-Methode mit einem anschließenden strukturierten Interview. Die Auswertungen der so gewonnenen Aussagen erfolgen anhand von deduktiven Kategoriensystemen (für die Think-Aloud-Protokolle) und induktiven Kategoriensystemen (für die Interviewtranskripte). Ergebnisse werden für angehende und praktizierende Lehrkräfte sowohl separat als auch vergleichend berichtet. Das Schlusskapitel bilanziert Fähigkeiten und Verbesserungspotenziale der Proband:innen mit Blick auf die Interpretation lernverlaufsdagnostischer Daten und beleuchtet zudem die festzustellenden bzw. fehlenden Abweichungen zwischen angehenden und praktizierenden Lehrkräften. Darüber hinaus werden die grundsätzlich positiven Äußerungen der Proband:innen zur *Verwertbarkeit* lernverlaufsdagnostischer Daten sowie deren skeptische Äußerungen mit Blick auf die praktische Erhebung dieser Daten diskutiert.

2. Hintergrund

2.1 Formatives Assessment und Lernverlaufsdagnostik

In Abgrenzung zum outputorientierten, summativen Assessment kann das formative Assessment¹ der Prozessdiagnostik zugeordnet werden (SCHÜTZE et al., 2018; INGENKAMP/LISSMANN, 2008). Es ist gekennzeichnet durch eine häufige, systematisch wiederholte Leistungsmessung (KLAUER, 2014) und erfolgt auf Klassen- oder Individualebene mit dem Ziel der Unterrichtverbesserung (SCHÜTZE et al., 2018). Beim for-

1 In der Literatur lassen sich neben Begriffen wie „formatives Assessment“ und „formative Evaluation“ auch die Begriffe „formative Leistungsbeurteilung“ (vgl. bspw. KLAUER, 2014; SCHÜTZE et al., 2018) und „classroom evaluation“ (BLACK/WILLIAM, 1998) finden. Diese sind hier gleichbedeutend zu verstehen.

mativen Assessment wird also das Erreichen von Lernzielen in kurzfristigen Zeitabständen während des Unterrichts mehrmals gemessen, sodass mit didaktischen Variationen wie etwa wiederholender Übung, schwierigkeitsdifferenzierter Aufgabenbearbeitung oder gezielten Fördermaßnahmen für schwächere Schüler:innen reagiert werden kann (INGENKAMP/LISSMANN, 2008; SCHÜTZE et al., 2018). Aus diesem Grund wird dem formativen Assessment ein hohes lernförderliches Potenzial zugeschrieben (SCHÜTZE et al., 2018).

Ein pädagogisch angereichertes Konzept zur Integration regelmäßiger formativer Assessments in den Unterricht ist in Abbildung 1 ersichtlich: Demnach sollten sowohl Lernziele als auch Erfolgskriterien von Beginn an transparent für Lehrkraft und Lernende sein. Darauf aufbauend wird zu mehreren Zeitpunkten der Lehr-/Lerneinheit festgestellt, auf welchem Leistungsstand sich ein:e Schüler:in jeweils befindet. Im dritten Schritt (nach jedem Erhebungszeitpunkt) sollen die in Schritt II generierten diagnostischen Informationen dazu verwendet werden, Lernenden eine Rückmeldung über bereits gemeisterte Lernschritte und bestehende Leistungsdefizite zu geben, sie auf dem Weg zum Lernziel individuell zu unterstützen und/oder didaktische Maßnahmen im Klassenverbund anzupassen. Die Schritte IV und V sind optional und können Lernprozesse zusätzlich anregen bzw. vertiefen. So sollten Schüler:innen zum einen aktiviert werden, ihren eigenen Lernprozess und gegenwärtigen Lernstand zu reflektieren, um ihre nachfolgenden Lernhandlungen zielbezogener zu steuern. Zum anderen sollten sie sich gegenseitig Feedback geben und sich so auf dem Weg zur Lernzielerreichung untereinander unterstützen (SCHÜTZE et al., 2018).

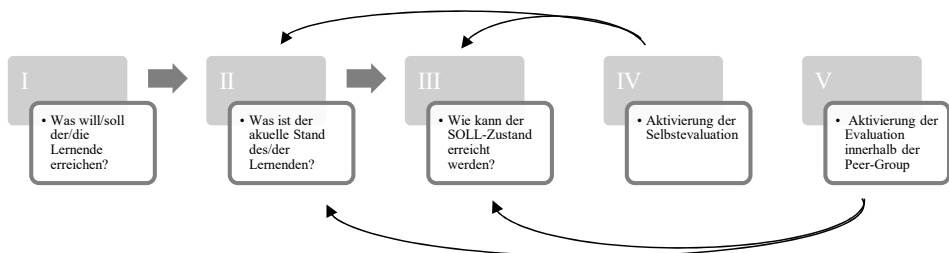


Abbildung 1: Schritte des formativen Assessments in Anlehnung an SCHÜTZE et al. (2018, S. 700 f.)

Die Lernverlaufsdiagnostik als Variante des formativen Assessments ist

eine in kurzen Zeitintervallen wiederholte standardisierte, psychometrisch fundierte Leistungsdiagnostik mit Paralleltests zur Begleitung von Lernprozessen, die Lehrkräften eine Rückmeldung über die Leistungsentwicklung ihrer Schülerinnen und Schüler und somit über den Erfolg ihrer Fördermaßnahmen liefert. (SCHÜTZE et al., 2018, S. 705)

Sie ermöglicht die Erfassung eines Lernverlaufs, also der Entwicklung einer Kompetenz von Lernenden, über eine längere Zeitspanne hinweg (KLAUER, 2014). Die Verwendung computerbasierter Testverfahren erlaubt es Lehrkräften, Zielsetzungen für Schüler:innen zu festzulegen und der jeweiligen Lernentwicklung anzupassen, den eigenen

Unterricht zu evaluieren und entsprechend zu verbessern, Interventionen bzw. didaktische Maßnahmen hinsichtlich ihres Erfolges miteinander zu vergleichen und Strategien zur weiteren Unterrichtsverbesserung zu identifizieren (FUCHS et al., 2015). Diagrammförmig aufbereitete Lernverlaufskurven bieten eine strukturierte Übersicht über den Lernverlauf von Schüler:innen sowie das gesetzte curriculare bzw. kriteriale Ziel. Die grafische Darstellung von Lernverläufen von Schüler:innen lässt zudem einen Vergleich zum Mittelwert der Klasse zu und eröffnet so zusätzliche Interpretationsmöglichkeiten. Die Einordnung individueller Fortschritte in den klasseninternen Bezugsrahmen stellt einen besonderen Informationsgehalt von diagrammförmig aufbereiteten Lernverlaufskurven dar und kann didaktische Entscheidungen (etwa zu binnendifferenzierendem Methodeneinsatz) zusätzlich erleichtern (KLAUER, 2014).

2.2 Funktionale Lesefähigkeit

Wie eben ausgeführt erfolgt das formative Assessment nicht zum Selbstzweck, sondern dient der Anpassung des Unterrichts an die Lernbedarfe der Schüler:innen (SCHÜTZE et al., 2018). Beruht diese Adaption auf zuvor erhobenen Daten, setzt sie neben (fach-) didaktischen und (fach-) inhaltlichen Kenntnissen (s. hierzu z. B. ACHTENHAGEN/WEBER, 2015) auch eine pädagogische *data literacy* der Lehrpersonen voraus (MANDINACH, 2012). Im deutschsprachigen Raum wird hierfür insbesondere der Begriff der funktionalen Lesefähigkeit verwendet. Funktionales Lesen zeichnet sich durch ein Lesen mit Informationsnutzung und zielgerichteter Informationssuche aus, wobei tendenziell Handlungsdruck besteht (ZIEGLER et al., 2016). Drei Repräsentationsformate von Informationen lassen sich dabei unterscheiden. Während die *deskriptionale* Repräsentation einem kontinuierlichen Text entspricht und daher den Anwendungskern der *data literacy* nur tangiert, scheinen *hybride* Repräsentationen wie z. B. Tabellen sowie *depiktionale* Dokumente wie Diagramme zentral (ebd.). Die konkreten Fähigkeiten funktionalen Lesens zeigen sich indes auf mehreren Anforderungsstufen: So müssen zunächst Informationen lokalisiert werden. Als weiterer Schritt sind Informationen zu integrieren, es werden also Bezüge zwischen Informationen hergestellt. Höhere Anforderungen werden auf der Stufe „Generieren“ gestellt, bei der eine mentale Modellierung erforderlich wird. Das heißt, dass der/die Leser:in in der Lage ist, Inhalte in verschiedene Modi der Informationsdarstellung zu übertragen (ebd.). In englischsprachiger Literatur und spezifisch auf Lernverlaufsdaten bezogen werden diese drei Anforderungsstufen mit den Begriffen *read the data* (Beschreiben von Daten), *read between the data* (Daten in Beziehung zueinander setzen, z. B. Vergleichen) und *read beyond the data* (z. B. Prognostizieren auf Basis der Daten, Annehmen geeigneter instruktorischer Maßnahmen) charakterisiert (CURCIO, 1981; ZEUCH et al., 2017). Sie können als Teil diagnostischer Kompetenz verstanden werden und sind somit maßgeblich für diagnostisch fundierte Entscheidungen in der Didaktik (FÖRSTER/SOUVIGNIER, 2017).

2.3 Einstellungen von Lehrkräften zu datenbasierten Entscheidungen

Die Nutzung von Daten, bspw. aus (Schul-)Evaluationen und Lernstandserhebungen, um evidenzbasierte Entscheidungen zu treffen, ist abhängig von persönlichen Eigenschaften einer Lehrkraft, der schulischen Struktur und Schulkultur (ERCAN et al., 2021; ZLATKIN-TROITSCHANSKAIA et al., 2016). Bezüglich der persönlichen Eigenschaften einer Lehrkraft ist neben ihren diagnostischen und didaktischen Fähigkeiten auch ihre Einstellung² zur Datennutzung von Bedeutung (ERCAN et al., 2021; REUSSER/PAULI, 2014). Zum einen können Lehrkräfte durch mangelndes Selbstvertrauen bezüglich der Interpretation von Daten in ihrer Datennutzung gehemmt werden (DATNOW/HUBBARD, 2016; kognitive Einstellungskomponente, hier der selbst- und methodenbezogenen Überzeugung im Umgang mit Daten). Zum anderen kann auch die Angst, erhobene Daten könnten durch Vorgesetzte dazu herangezogen werden, die Lehrkraft selbst zu bewerten, eine Abwehrhaltung gegenüber evidenzbasiertem Handeln befördern (ebd.; affektive Einstellungskomponente). Sind Lehrkräfte nicht bereit Daten als Entscheidungsgrundlage für ihr pädagogisch-didaktisches Handeln zu nutzen, werden sie diese auch seltener aktiv bzw. initiativ heranziehen als Kollegiumsmitglieder, die einer solchen Verwertung offener gegenüberstehen (VANLOMMEL et al., 2016; DEMSKI, 2017; verhaltensbezogene Komponente). Befragungsstudien legen in diesem Zusammenhang nahe, dass Lehrkräfte ihre pädagogisch-didaktischen Entscheidungen bevorzugt auf ihre persönliche Intuition stützen (VANLOMMEL et al., 2017). Dahinter verbirgt sich die Überzeugung, dass intuitionsbasierte Entscheidungen, welche in Erfahrungswerten und im persönlichen Kontakt zu Schüler:innen gründen, den Bedürfnissen der Schüler:innen gerechter werden als datenbasierte Entscheidungen (ebd.). So entsteht eine starke Tendenz, entweder direkt intuitionsbasiert zu entscheiden oder aber nur jene Daten selektiv zu verwerten, welche die eigene Intuition bestätigen (ebd.; DATNOW/HUBBARD, 2016). Ähnliches zeigte sich in einer Studie von SCHILDKAMP und KUIPER (2010) an niederländischen Schulen. Bei Proband:innen aus drei von sechs teilnehmenden Schulen konnte eine geringe Akzeptanz evidenzbasierter Entscheidungsfindung und damit verbunden eine geringere faktische Datennutzung festgestellt werden. So konstatierten die Lehrkräfte dieser Schulen, dass die Lehrerfahrung für die förderliche Gestaltung von Unterricht ausreichend sei und die Berücksichtigung von Daten daher unnötig ma-

2. Dem vorliegenden Beitrag liegt ein multidimensionales Verständnis des Begriffs „Einstellung“ zugrunde, demzufolge Einstellungen aus kognitiven, affektiven und verhaltensbezogenen Komponenten bestehen (HADDOCK/MAUI, 2014). Die kognitive Komponente der Einstellung im Sinne von Überzeugungen bringt zum Ausdruck, was eine Lehrkraft glaubt und subjektiv für richtig hält (REUSSER/PAULI, 2014). Ihr liegt stets ein Gegenstandsbezug zugrunde, welcher in der vorliegenden Untersuchung auf die Nutzung lernverlaufsdiagnostischer Daten gerichtet ist (vgl. ebd.). Eine strikte Abgrenzung gegenüber Wissen ist dabei nicht möglich (ebd.). Die affektive Komponente beschreibt Gefühle, die mit dem relevanten Gegenstand verbunden sind, etwa Unbehagen oder Begeisterung (HADDOCK/MAUI, 2014). Die verhaltensbezogene Komponente beschreibt die (auch nach außen demonstrierten) Verhaltenstendenzen, die für wiederkehrende oder zukünftige gegenstandsrelevante Aktivitäten bestehen, etwa im Wahlverhalten der Person.

che (SCHILDKAMP/KUIPER, 2010). Im Lichte der verfügbaren Studien herrscht damit gegenüber der Nutzung von diagnostischen Daten in etlichen Kollegien noch Skepsis.

3. Forschungsfragen und methodisches Vorgehen

Die vorliegende Studie besitzt zwei Betrachtungsschwerpunkte. Erstens interessiert, wie angehende und praktizierende Lehrkräfte beruflicher Schulen Diagramme mit Lernverlaufskurven (als typische depiktionale Repräsentationsformen im formativen Assessment) interpretieren, ob sie dabei den Informationsgehalt der grafisch verdichteten Daten erfassen und welche spontanen didaktischen Schlussfolgerungen sie hieraus ziehen. Diese Betrachtungen werden durch Analysekategorien strukturiert, die aus Modellierungen der funktionalen Lesefähigkeit und Vorgängerstudien aus dem allgemeinbildenden und sonderschulischen Bereich gewonnen wurden (zu Details der Operationalisierung siehe unten). Unter der forschungsleitenden Frage 1 *„Wie lesen angehende und praktizierende Lehrkräfte an beruflichen Schulen lernverlaufsdagnostische Daten?“* untersuchen wir dabei spezifisch,

- 1.1 ... auf welchen Anforderungsstufen die Proband:innen aus dem beruflichen Schulwesen sich bei der Interpretation eines gegebenen Diagramms bewegen und in welcher Folge sie diese durchlaufen;
- 1.2 ... inwieweit die vorgenommenen Interpretationen mit Blick auf den Informationsgehalt der dargestellten Daten als vollständig und akkurat bezeichnet werden können;
- 1.3 ... ob und in welcher Hinsicht sich praktizierende Lehrkräfte von angehenden Lehrkräften hinsichtlich des Vorgehens bei der Interpretation (1.1) sowie der Vollständigkeit und Akkuratheit (1.2) unterscheiden;
- 1.4 ... ob sich interindividuelle Unterschiede bei der Interpretation lernverlaufsdagnostischer Daten auf weitere Merkmale der Proband:innen, konkret das studierte Nebenfach, den erlernten Ausbildungsberuf, die geschilderte Erfahrung mit lernverlaufsdagnostischen Daten oder den Besuch bestimmter universitärer Veranstaltungen, zurückführen lassen.

Die hierzu durchgeführten Gruppenvergleiche gründen in der allgemeinen Annahme, dass mit verschiedenen Erfahrungshintergründen wie etwa der Belegung von Veranstaltungen zur pädagogischen Diagnostik im Studium oder der Arbeit mit formativen Assessments an der eigenen Schule auch unterschiedliche Lerngelegenheiten zum funktionalen Lesen wie auch pädagogisch-didaktischen Verwerten von Lernverlaufsdigrammen verknüpft sein können (s. hierzu vertieft SIEGRIED/ACKERMANN, 2022). Diese Erfahrungen könnten sowohl Einstellungen als auch die geäußerten (unterrichtlichen) Schlussfolgerungen aus dem vorgelegten Diagramm prägen.

Der zweite Betrachtungsschwerpunkt knüpft an die aktuelle Forschung zu Einstellungen von Lehrkräften zur Datennutzung an (s. Kapitel 2.3). Hierzu werden entsprechende Interviewpassagen im Rahmen einer strukturierenden Inhaltsanalyse

weitgehend induktiv ausgewertet. Unter der Forschungsfrage 2 „Welche Erfahrungen und Einstellungen mit/zur Nutzung lernverlaufsdiagnostischer Daten haben angehende und praktizierende Lehrkräfte beruflicher Schulen?“ bündeln wir die Antworten der Proband:innen innerhalb zweier Interviewbereiche:

- 2.1 Wie schildern die Proband:innen ihre praktischen Erfahrungen mit konkreten lernverlaufsdiagnostischen Informationen?
- 2.2 Welche Informationen über die Lernprozesse der Schüler:innen würden sie sich grundsätzlich in formativen Assessments wünschen und wofür würden sie diese nutzen?

Um diesen Fragen nachzugehen, wurde ein Mixed-Methods-Ansatz gewählt, welcher primär die strukturierende inhaltsanalytische Auswertung von Laut-Denken-Protokollen³ (vgl. DÖRING/BORTZ, 2016; KONRAD, 2020) (mit deduktiv entwickelten Kategorien) und von Interviews (mit induktiv gebildeten Kategorien) umfasst. Dabei wurden die Interviewtranskripte nach Sinneinheiten kodiert (im Ergebnisbericht zumeist als „Aussagen“ bezeichnet). Für ergänzende statistische Vergleiche zwischen Lehrkräfte-Gruppen mit unterschiedlichen Erfahrungshintergründen wurden nicht-parametrische statistische Verfahren verwendet.

Es nahmen 33 praktizierende Lehrkräfte von insgesamt sieben berufsbildenden Schulen und 30 Studierende der Wirtschaftspädagogik der Universitäten Göttingen und Kassel von Oktober 2019 bis Januar 2020 teil. Den Proband:innen wurde die Lernverlaufskurve der fiktiven Schülerin Lena im Vergleich zum Leistungsdurchschnitt der Klasse und einer Zielkurve vorgelegt (s. Abbildung 2):

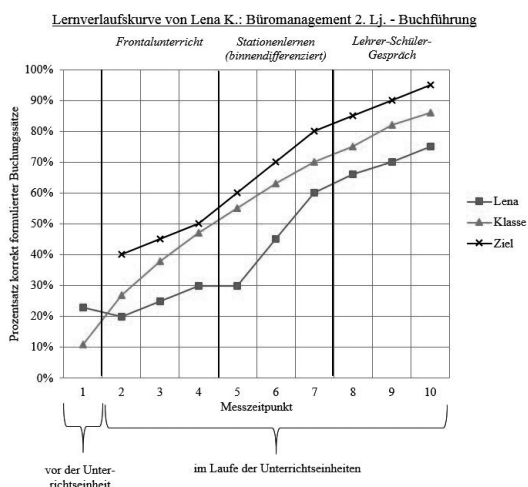


Abbildung 2: Lernverlaufskurve
(eigene Darstellung)

- 3 In Anlehnung an KONRAD (2020) kann das Verfahren des Lauten Denken im Rahmen dieses Beitrags als unstrukturiert und kontinuierlich klassifiziert werden. Dementsprechend wurden die Gedankenentäußerungen der Proband:innen frei und unmittelbar während der Diagrammbetrachtung vorgenommen. Sie wurden nicht durch sog. Prompts unterbrochen und/oder auf spezifische Diagrammelemente gelenkt.

Die Lernverlaufskurve erfüllt die von ZEUCH et al. (2017) identifizierten Charakteristika dieser Repräsentationsform schülerbezogener Informationen zu Zwecken des formativen Assessments. So gibt es mehrere Messzeitpunkte (hier: 10); der prozentuale Anteil korrekter Aufgabenlösungen (hier: formulierte Buchungssätze) wird für die jeweiligen Messzeitpunkte markiert; die Messungen sind miteinander verbunden. Zudem besteht die Möglichkeit die Leistung der Schülerin Lena mit den Leistungen der Klasse zu vergleichen. Der Lernverlauf der Schülerin lässt zudem Interpretationsspielraum, indem beispielsweise für den guten Start Lenas im Vergleich zur Klasse und der darauffolgenden Leistungsverschlechterung gemutmaßt werden kann, dass Lena möglicherweise schon Vorerfahrung mit dem Thema „Buchführung“ hat und sich hierauf zunächst ausruht, was eine negative Entwicklung begünstigt. Unter Berücksichtigung von ESPIN et al. (2017) wurde zudem eine Kurve mit den von der Lehrkraft für den Zeitpunkt definierten Leistungsanforderungen an die Schüler:innen (Ziel) eingezeichnet. Des Weiteren sind drei Interventionsabschnitte markiert, in denen die Lehrkraft unterschiedliche didaktische Maßnahmen im Unterricht eingesetzt hat. Als Einstimmung und Aufforderung zum freien und ausführlichen Entäußern ihrer Gedanken (Think-Aloud bzw. Lautes Denken) in der Auseinandersetzung mit dem Diagramm wurden den Proband:innen folgende Situation und Instruktion geschildert:

Sie sind Lehrerin/Lehrer in einer Klasse von Auszubildenden zu Kaufleuten für Büromanagement im zweiten Lehrjahr. Im Lernfeld 6 „Wertströme erfassen und beurteilen“ beschäftigen sich die Schülerinnen und Schüler nun erstmals mit dem Thema „Buchführung“.

Zur Unterstützung des Unterrichts wurden Ihnen Aufgaben zur Buchführung zur Verfügung gestellt, welche die Schülerinnen und Schüler wiederkehrend in kleinen Aufgabenblöcken am PC bearbeitet haben. Bei den Aufgaben geht es immer um das Aufstellen von Buchungssätzen, die im Einstiegsunterricht typisch sind. Der Test variiert die zu formulierenden Buchungen nur inhaltlich, das Anforderungsniveau bleibt jedoch durchweg konstant.

Die Software erstellt Ihnen automatisch eine Übersicht über die in jedem Aufgabenblock erzielten Ergebnisse für jedes Mitglied der Klasse. Zudem liefert sie Vergleichswerte mit. Ich zeige Ihnen nun die fiktive Kurve der Testergebnisse einer Schülerin. Bitte schauen Sie sich das Diagramm genau und in Ruhe an. Bitte sagen Sie mir dabei aber andauernd, was Sie sehen und denken. Nennen Sie mir, was Sie auf dem Diagramm ansehen und warum. Sagen Sie mir, welche Schlussfolgerungen Sie über den Lernprozess der Schülerin ziehen.

Da die (angehenden) Lehrkräfte dazu neigten, bereits nach einer eher oberflächlichen Betrachtung des Diagramms fragend innezuhalten, wurde in der Regel ein- bis zweimal nachgefasst, ob die/der Interviewte noch weitere Gedanken bzw. Ergänzungen hat.

Im Anschluss an die Interpretation des oben gezeigten Graphen wurden den Teilnehmenden Fragen zu ihren Erfahrungen mit und ihrer Haltung zu lernverlaufsdagnostischen Informationen gestellt. Außerdem wurden Merkmale der Person, wie z. B. Dauer der Lehrtätigkeit in Jahren, erfragt. Der Interviewleitfaden befindet sich in Tabelle 1.

Tabelle 1: Interviewleitfaden

Studierende	Lehrkräfte
1.1 Was studierst du?	1.1 Wie lange bist du bereits als Lehrer:in, inkl. Referendariat, tätig?
1.2 Welches Nebenfach studierst du?	1.2 Welche Ausbildungsberufe unterrichtest du schwerpunktmäßig?
1.3 Welchen Studienabschluss strebst du aktuell an? (Bachelor/Master)	
2. Hast du selbst einen kaufmännischen Ausbildungsberuf erlernt? Wenn ja, welchen?	2. Hast du selbst einen kaufmännischen Ausbildungsberuf erlernt? Wenn ja, welchen?
3. Im Folgenden interessieren wir uns besonders für bisherige Lerngelegenheiten zum Thema „Diagnostik“:	3. Im Folgenden interessieren wir uns besonders für Lerngelegenheiten zum Thema „Diagnostik“:
a. Hast du im Rahmen deines Studiums forschungsmethodische Veranstaltungen besucht? Wenn ja, welche?	a. Hast du im Rahmen deines Studiums forschungsmethodische Veranstaltungen besucht? Wenn ja, welche?
b. Hast du im Rahmen deines Studiums Veranstaltungen zur pädagogischen/psychologischen Diagnostik besucht? Wenn ja, welche?	b. Hast du im Rahmen deines Studiums Veranstaltungen zur pädagogischen/psychologischen Diagnostik besucht? Wenn ja, welche?
c. Hast du im Rahmen deines Studiums Veranstaltungen zur Allgemeinen Statistik besucht? Wenn ja, welche?	c. Hast du im Rahmen deines Studiums Veranstaltungen zur Allgemeinen Statistik besucht? Wenn ja, welche?
4. Hast du im Rahmen eines Schulpraktikums oder sonstiger Einsätze an (beruflichen) Schulen bereits Erfahrung mit <i>lernverlaufsdiagnostischen</i> Informationen gesammelt – sei es in der vorliegenden Form oder in einer anderen Form und eventuell auch anderen Bezeichnung? Wenn ja, welche?	4. Hast du in deiner bisherigen <i>Tätigkeit als Lehrkraft</i> bereits Erfahrung mit <i>lernverlaufsdiagnostischen</i> Informationen gesammelt – sei es in der vorliegenden Form oder in einer anderen Form und eventuell auch anderen Bezeichnung? Wenn ja, welche?
a. Mögliche Nachfrage bei Nutzung diagnostischer Daten (abhängig von vorheriger Schilderung): Warum wurden die von dir geschilderten diagnostischen Daten genutzt?	a. Mögliche Nachfrage bei Nutzung diagnostischer Daten (abhängig von vorheriger Schilderung): Warum nutzt du die von dir geschilderten diagnostischen Daten?
b. Nachfrage bei keiner Nutzung diagnostischer Daten: Welche anderen Informationen hast du für didaktische Entscheidungen genutzt?	b. Nachfrage bei keiner Nutzung diagnostischer Daten: Auf welche Einschätzungen stützt du dich bei deinen didaktischen Entscheidungen?
c. Nachfrage bei keiner Erfahrung im Sinne von Schulpraktika etc.: Was meinst du, worauf stützen Lehrkräfte ihre didaktischen Entscheidungen?	

5. Wie nützlich schätzt du Daten bzw. Graphen wie die präsentierte für die Unterrichtspraxis ein? Wofür könnten Lehrkräfte die Daten nutzen?
5. Stell dir vor, du könntest entscheiden, welche Daten wie oft und in welchen Klassen oder Lernfeldern generiert werden könnten: Welche Informationen über die Lernprozesse der Schüler/innen würdest du dir wünschen und wofür würdest du diese Informationen nutzen?

Die Analyse der Interview-Transkripte erfolgte im Rahmen einer strukturierenden Inhaltsanalyse (vgl. MAYRING, 2015). Das zugrunde gelegte Kategoriensystem basiert auf den verschiedenen Anforderungsstufen funktionalen Lesens (vgl. CURCIO, 1981; ZIEGLER et al., 2016). Zusätzlich zu den Kategorien *read the data*, *read between the data* und *read beyond the data* wird in Anlehnung an ESPIN et al. (2017) die Kategorie *framing* berücksichtigt. Diese beschreibt Aussagen zur Einbettung der gezeigten Daten. Hierzu gehören die Betrachtung der Überschrift und Legenden sowie grundlegende Aussagen über das Diagramm wie bspw., dass es drei Kurven enthält. *Framing* stellt gemäß ESPIN et al. (2017) den ersten Schritt im Rahmen der Interpretation lernverlaufsdiagnostischer Daten dar. Einer idealen Schrittweise folgend sind im Anschluss Beschreibungen konkreter Datenpunkte (*read the data*), im nächsten Schritt die Integration von Informationen in Form von Vergleichen (*read between the data*) und zuletzt die Interpretation im engeren Sinne, das Prognostizieren und Schlussfolgern weiterer instrukionaler Maßnahmen (*read beyond the data*) zu erwarten. Folgende Ankerbeispiele wurden für die Analysekategorien identifiziert (s. Tabelle 2):

Tabelle 2: Ankerbeispiele der Analysekategorien

Kategorie	Ankerbeispiel
<i>Framing</i>	„Also drei Kurven in drei verschiedenen Farben, einmal in dunkelblau, grün und rot. Die rote zeigt den Graphen von Lena, die ist ja auch die Lernverlaufskurve von Lena. Die grüne zeigt die Klasse insgesamt und die dunkelblaue ist dann das Ziel.“ (S03)
<i>Read the data</i>	„Dann macht es einen Sprung, Lena wird viel besser über 10 %, nämlich 15 % besser in der 6. Einheit. Dann springt sie nochmal richtig weit nach oben 60 % ihrer Buchungen sind jetzt richtig, wenn ich das jetzt hier richtig lese“ (L32)
<i>Read between the data</i>	„Das heißt ich sehe jetzt hier, dass die Klasse besser ist als Lena. Aber auch Lena noch nicht und auch die Klasse noch nicht so gut ist, wie das Ziel ist.“ (L09)
<i>Read beyond the data</i>	„Lena erreicht beim Stationenlernen, also bei den binnendifferenzierten Aufgaben die größten Sprünge von korrekt formulierten Buchungssätzen. Also der Lernzuwachs ist hier am größten [...] Also würde ich daraus schließen, ein weiterer Frontalunterricht hätte ihr hierbei nicht geholfen“ (L21)

Die Untersuchung der Schrittfolge bei der Interpretation lernverlaufsdiagnostischer Daten (vgl. Forschungsfrage 1.1) beruht in Anlehnung an BOSCH et al. (2017) auf einem Punktesystem zur Ermittlung des prozentualen Anteils sequentiell kohärenter Schritte. Dabei wird für jeden korrekten Schritt ein Punkt vergeben und die Summe schließlich durch die Anzahl der Kategorien bzw. Schritte insgesamt dividiert. Zur Untersuchung der Vollständigkeit und Akkuratheit wird ein weiteres Scoringssystem herangezogen. Dazu wurden verschiedene Diagrammkomponenten definiert, zu denen Aussagen der Proband:innen erwartet wurden. Für jede Komponente, wie z. B. die Identifizierung der für die betreffende Schülerin lernförderlichste didaktische Variation (Anforderungsstufe *read beyond the data*), konnte im Rahmen der auf Vollständigkeit fokussierten Analyse je ein Punkt erlangt werden. Die Vollständigkeitsbewertung erfolgte unabhängig davon, ob die zugehörige Diagrammkomponente von den Proband:innen auch akkurat erfasst wurde. Im Rahmen der Akkuratheitsbewertung konnten pro Komponente maximal 2 Punkte erreicht werden. Ein Punkt wurde vergeben, wenn die Äußerung zu einer Komponente lückenhaft oder nur oberflächlich war. Null Punkte wurden bei inhaltlich falschen Äußerungen vergeben. Ergänzt werden die qualitativen Analysen um nicht-parametrische Testverfahren (Mann-Whitney-U-Test; Kruskal-Wallis-Test) zum Vergleich verschiedener Personengruppen.

4. Ergebnisse

4.1 Zum Umgang mit dem vorgelegten Lernverlaufdiagramm

Insgesamt konnten 23.165 Wörter, gebündelt zu Sinneinheiten, den gewählten Kategorien zugeordnet werden. Betrachtet man in einem ersten Auswertungsschritt, mit dem näherungsweise die Ausführlichkeit der von den Proband:innen unternommenen Interpretationsschritte erfasst werden soll, die in jeder kategorial kodierten Sinneinheit entäußerten (bedeutungstragenden) Worte, so entfallen 22,4 % der Wörter auf die Kategorie *Framing*, 25,1 % auf die Kategorie *Read the data*, 31,5 % auf die Kategorie *Read between the data* und 21,1 % auf die Kategorie *Read beyond the data*. Sowohl praktizierende als auch angehende Lehrkräfte machten primär vergleichende Aussagen, die entsprechend als *read between the data* kodiert wurden (29,1 % bzw. 33,5 % des Gesagten). In der Gegenüberstellung beider Lehrkräftegruppen fällt weiterhin auf, dass praktizierende Lehrkräfte vermehrt das Diagramm einbettende Äußerungen tätigen. So wurden 26,1 % ihrer in Sinneinheiten kodierte Wörter der Kategorie *Framing* zugeordnet, wohingegen angehende Lehrkräfte hierauf 19,3 % ihrer verbalisierten Überlegungen verwenden. Folgende Tabelle bietet eine Übersicht über den relativen Anteil der Kategorien *Framing*, *Read the data*, *Read between the data* und *Read beyond the data* am Gesagten:

Tabelle 3: Übersicht über die Anteile der Analysekategorien an den in Sinneinheiten kodierten Wörtern

	Framing	Read the data	Read between the data	Read beyond the data
Praktizierende Lehrkräfte	26,07 %	23,89 %	29,11 %	20,93 %
Angehende Lehrkräfte	19,30 %	26,10 %	33,52 %	21,08 %
Gesamt	22,35 %	25,07 %	31,49 %	21,10 %

Mit Blick auf *Forschungsfrage 1.1* ergibt sich folglich, dass alle untersuchten Lehrkräfte insbesondere auf das Vergleichen gegebener Daten abstellen und sich damit auf das charakteristische Darstellungspotential von Lernverlaufskurven konzentrieren. Bezüglich der sequentiellen Kohärenz wäre idealtypisch zu erwarten, dass die Proband:innen die Daten zunächst rahmen (*Framing*), sie dann beschreiben (*Read the data*), anschließend vergleichen (*Read between the data*) und schließlich Schlussfolgerungen hieraus ziehen (*Read beyond the data*). Tatsächlich begannen die Proband:innen ganz überwiegend mit einem Einbetten der Daten (*Framing*), fuhren jedoch danach direkt mit einem Vergleich von Daten fort (*Read between the data*), bevor einzelne Datenpunkte separat betrachtet und beschrieben, Schlussfolgerungen gezogen oder weitere Zwischenschritte unternommen wurden. Die individuellen Abweichungen von einer idealtypischen interpretierenden Erschließung des Diagramms schwankten dabei erheblich. Bei den praktizierenden Lehrkräften lag der höchste Anteil kohärent aufeinander folgender Schritte bei 75 %, der geringste Anteil bei 20 %. Auf Basis des in 3 beschriebenen Punktesystems könnten die Interpretationen der praktizierenden Lehrkräfte als zu 38 % sequentiell kohärent beschrieben werden. Mit einer sequentiellen Kohärenz von durchschnittlich 42 % erreichen die Studierenden zwar leicht höhere Werte. Allerdings liegt bei ihnen auch eine höhere Spannweite im Anteil sequentiell kohärenter Äußerungen vor (13 % bis 75 %).

Bezüglich der Vollständigkeit und Akkuratheit (*Forschungsfrage 1.2*) erweisen sich Interpretationen angehender Lehrkräfte im Durchschnitt in ihrer Vollständigkeit als besser, nicht aber in ihrer Akkuratheit. Folgendes Diagramm zeigt die erreichbaren Punkte einer Kategorie sowie die von angehenden und praktizierenden Lehrkräften durchschnittlich erreichte Punktzahl (s. Abbildung 3):

Wie das Diagramm zeigt, erreichen die Studierenden im Durchschnitt 8,4 von 10 Punkte im Bewertungsaspekt der Vollständigkeit ihrer Interpretationen, während praktizierende Lehrkräfte im Durchschnitt 7,2 von 10 Punkten erlangen. Diese Abweichung erweist sich bei Durchführung eines Mann-Whitney-U-Tests als überzufällig. Es ergibt sich eine moderate Effektstärke r^4 in Höhe von 0,335.

4 Die Effektstärke $r = 0,1$ ist ein kleiner Effekt, $r = 0,3$ ein mittlerer Effekt und $r = 0,5$ ist ein starker Effekt (BÜHNER/ZIEGLER, 2017, S. 320).

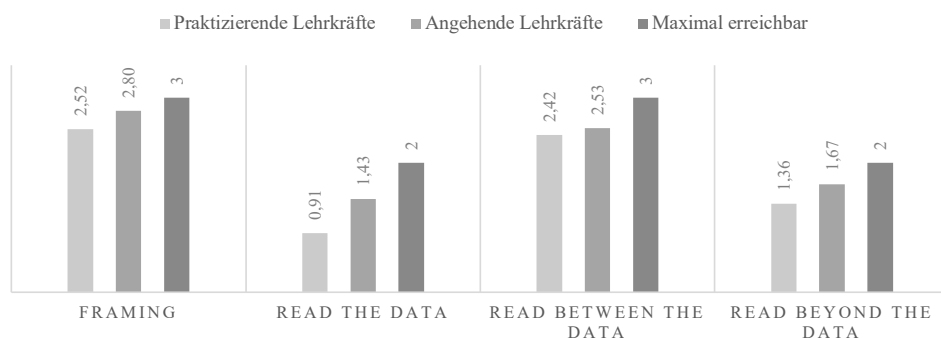


Abbildung 3: Bewertung der Vollständigkeit – Maximal erreichbare und durchschnittlich erreichte Punktzahl pro Analysekatgorie (eigene Darstellung)

Hinsichtlich der Akkuratheit lassen sich hingegen kaum Unterschiede zwischen praktizierenden und angehenden Lehrkräften feststellen. Um die Vergleichbarkeit beider Gruppen zu wahren, wurden hierbei nur Diagrammkomponenten berücksichtigt, zu denen überhaupt eine Äußerung erfolgte. Demnach erzielten die beiden Personengruppen 88,9 % (praktizierende Lehrkräfte) bzw. 87,4 % (angehende Lehrkräfte) der erreichbaren Punkte. Pro Kategorie (*Framing*, *Read the data*, *Read between the data*, *Read beyond the data*) konnten maximal zwei Punkte erreicht werden. Abbildung 4 zeigt die durchschnittlich erzielten Punkte praktizierender und angehender Lehrkräfte für jede Kategorie:

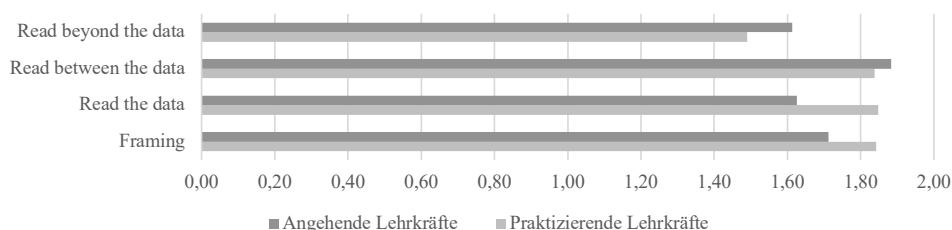


Abbildung 4: Bewertung der Akkuratheit – Durchschnittlich erreichte Punktzahl pro Analysekatgorie (eigene Darstellung)

Es zeigt sich, dass praktizierende Lehrkräfte insbesondere auf den Anforderungsstufen des Einbettens und Beschreibens der Daten (*Framing* und *Read the data*) ein besseres Ergebnis erzielen als angehende Lehrkräfte, obgleich Letztere vergleichsweise mehr Worte für die Beschreibung der Daten (*Read the data*) verwenden. Die Studierenden sind in ihren Äußerungen, die als *Read between the data* und *Read beyond the data* kodiert wurden, akkurater. Im Gesamtblick wird die geringste Akkuratheit auf der Stufe *Read beyond the data* erzielt.

Vergleicht man die Interpretationen der vorgelegten Lernverlaufskurve von praktizierenden und angehenden Lehrkräften, lassen sich im Gesamtblick auf alle kodierten Kategorien der funktionalen Lesefähigkeit keine gravierenden Unterschiede feststellen

(*Forschungsfrage 1.3*). Allerdings wurden Interpretationen angehender Lehrkräfte hinsichtlich sequentieller Kohärenz und Vollständigkeit besser bewertet, wohingegen die Akkuratheit der getätigten Äußerungen praktizierender Lehrkräfte marginal höher bewertet wurde. Diese Befunde erweisen sich hingegen nur mit Blick auf eine höhere Vollständigkeit der Interpretation durch angehende Lehrkräfte im Vergleich zu praktizierenden Lehrkräften als überzufällig. Vor dem Hintergrund des Zwecks lernverlaufsdiagnostischer Daten (vgl. Abschnitt 2.1) ist eine (vorläufige, rein diagrammgestützte) Schlussfolgerung wichtig, durch welche didaktische Variation der höchste Lernerfolg bei der fiktiven Schülerin Lena erzielt werden konnte und von welcher sie somit vermutlich in einer anschließenden Lernsequenz am meisten profitieren würde. Hinsichtlich der Beantwortung dieser grundlegenden Frage zeigt sich wiederum eine leichte Überlegenheit der angehenden Lehrkräfte im Vergleich zu praktizierenden Lehrkräften. So bezeichnen nur 60 % der praktizierenden Lehrkräfte, aber 77 % der angehenden Lehrkräfte das Stationenlernen als die für Lena beste didaktische Variation angesichts der verfügbaren diagnostischen Daten.

4.2 Interpretationsleistungen lernverlaufsdiagnostischer Daten in Abhängigkeit unterschiedlicher Erfahrungshintergründe

Ergänzend zu einem Vergleich von praktizierenden und angehenden Lehrkräften wurde statistisch geprüft, ob weitere erfahrungsbezogene Merkmale der Proband:innen systematische Abweichungen in der Interpretation lernverlaufsdiagnostischer Daten erklären (*Forschungsfrage 1.4*). Hierfür wurden die Proband:innen zum einen danach gruppiert, ob sie bereits konkrete Erfahrungen speziell mit lernverlaufsdiagnostischen Daten gesammelt hatten oder nicht. Zum anderen wurde untersucht, ob jeweils der Besuch bzw. Nicht-Besuch einer forschungsmethodischen Veranstaltung, einer Veranstaltung zur pädagogisch-psychologischen Diagnostik sowie zur Allgemeinen Statistik einen Effekt auf die Interpretation hat. Weiterhin wurde geprüft, ob sich Unterschiede in der Interpretation lernverlaufsdiagnostischer Daten je nach erlerntem Ausbildungsberuf sowie je nach studiertem Nebenfach ergeben. Für Zwei-Gruppen-Vergleiche wurde der Mann-Whitney-U-Test als non-parametrisches Verfahren gewählt, bei Gegenüberstellungen von mehr als zwei Gruppen kam der Kruskal-Wallis-Test zum Einsatz (vgl. BÜHNER/ZIEGLER, 2017). Im Gesamtblick ergeben sich allerdings keine überzufälligen Unterschiede mit Ausnahme des Besuchs von *Veranstaltungen zur pädagogisch-psychologischen Diagnostik*. Personen, die eine derartige Lehrveranstaltung besucht haben, erzielten bezüglich der sequentiellen Kohärenz ein signifikant besseres Ergebnis als Personen ohne entsprechende Veranstaltungsteilnahme (z -Wert = $-2,452$). Die Effektstärke r beträgt $0,314$ und ist somit als moderat einzuschätzen.

4.3 Zur praktischen Erfahrung (angehender) Lehrkräfte zu lernverlaufsdiagnostischen Daten

Forschungsfrage 2.1 zielt zunächst auf illustrierende Schilderungen der gesammelten Erfahrungen speziell mit lernverlaufsdiagnostischen Daten. Nur fünf der praktizierenden Lehrkräfte berichteten derartige direkte Erfahrungswerte (Lo2, Lo5, L18, L30, L32). Lehrkraft 2 berichtet aus dem Lernbüro:

Also ich habe kennengelernt, [...] dass man ein Kompetenzraster ausfüllt und [...] das vergleicht mit einer Schülereinschätzung. Dass sie sagen müssen, welche Aufgaben können sie sicher erledigen in der einen Einschätzung und Fremdeinschätzung, die auch überprüft wird im Rahmen von Tests. Da können wir das anhand von Aufgaben am PC [...] auch nachvollziehen, dass sie im Programm Lexware zum Beispiel, Sachen bearbeiten und wir können hinterher gucken, wie viel Prozent sind davon sicher. Und dann haben wir ein Kompetenzraster und stellen das mit dem Schüler gegenüber und kommen so dann in ein Gespräch über den Lernverlauf. (Lo2)

Eine weitere Lehrkraft berichtet aus dem Deutschunterricht, wobei die Schüler:innen in Vorbereitung für die Klassenarbeit über einen längeren Zeitraum hinweg immer wieder Texte formuliert haben und ein Feedback dazu bekamen. Allerdings sind die schriftlichen Ausarbeitungen auch in die mündlichen Noten eingeflossen. Ebenfalls im Deutschunterricht hat eine Lehrkraft (L18) die Website „orthografietrainer.net“ genutzt. Hierbei kann die Lehrkraft auch die Fortschritte der Schüler:innen einsehen. Lehrkraft 29, 30 und 32 haben ein Moodle-System mit Aufgaben genutzt und konnten auf diese Weise erfassen, ob die Schüler:innen sich verbesserten. Lehrkraft 32 schildert dazu:

Ich mache regelmäßig Moodle Tests. [...] Und da kann ich ja immer den Testverlauf angucken und dann kann ich ja sehen, was die Schüler verstanden haben und was nicht. Habe meine Prozentsätze von den Schülern und sehe halt, wo es noch hakt. Das ist ganz toll. Das ist transparent, schnell gemacht, man muss es nicht korrigieren, ein Mausklick und Ergebnis. Das ist super. (L32)

Vier praktizierende Lehrkräfte (L11, L16, L19, L31) sprechen von geringen Erfahrungen mit der Nutzung lernverlaufsdiagnostischer Daten. Hierbei wurden nicht systematisch Daten im Lernverlauf erhoben und analysiert, aber zumindest Vergleiche mit vorherigen Ergebnissen von Tests oder Hausaufgaben durchgeführt, um eine Tendenz des Lernverlaufs zu ermitteln (L11, L16, L19). Lehrkraft 19 stellt den Schüler:innen Selbsttests zur Verfügung, wertet diese aber nicht ihrerseits aus. Die Mehrheit (23 Lehrkräfte) gibt indes an, auf keine praktischen Erfahrungen mit lernverlaufsdiagnostischen Daten zurückblicken zu können.

Wurden lernverlaufsdiagnostische Daten ermittelt, so wurden diese von den Lehrkräften in erster Linie genutzt, um Schüler:innen individuell fördern zu können (Lo2, Lo5, L18, L19, L30). Zudem bestehe ein Interesse darin, den Wissenstand der Schüler:innen zu ermitteln (L16, L29, L32). Weitere Gründe der Datennutzung sind:

- Den Schüler:innen den eigenen Fortschritt sichtbar machen (L30, L31)
- Einstieg in ein Gespräch mit den Schüler:innen über Probleme etc. (L02, L30)
- Notengebung (L05, L30)
- Persönliches Interesse (L11)
- Motivation der Schüler:innen (L05)
- Tests stellen Übungen bzw. Vorbereitung auf Klassenarbeit für Schüler:innen dar (L05)

Hier zeigt sich allerdings, dass lernverlaufsdiagnostische Daten durch die Lehrkräfte nicht nur im Sinne eines formativen Assessments genutzt werden, sondern auch zur Notengebung (L05, L30).

Lehrkräfte, die lernverlaufsdiagnostische Daten nicht oder kaum nutzen, wurden ergänzend gefragt, worauf sie ihre didaktischen Entscheidungen stützen. Am häufigsten wurde hierzu die Klasse, deren Größe, Zusammensetzung oder anfängliches Leistungsniveau genannt (N=14). Des Weiteren führten Lehrkräfte die fachlichen Vorkenntnisse der Schüler:innen (N=9), Rückmeldungen der Schüler:innen (N=7), eigene Erfahrungen oder Erfahrungen anderer Lehrkräfte mit der Klasse (N=7), Ergebnisse von Klassenarbeiten (N=6) sowie Rahmenrichtlinien/ Rahmenlehrpläne bzw. schulinterne Lehrpläne (N=6) als relevante Bezugspunkte ihrer didaktischen Entscheidungen an. Eine Rolle spielen auch das Bauchgefühl (N=4) bzw. „Kopf“ und „Herz“ (LK01) sowie persönliche Gespräche mit Schüler:innen (N=4). Viermal wird zudem die mündliche Beteiligung bzw. Interaktion mit den Schüler:innen als entscheidungsbeeinflussend angegeben. Durch die Beiträge von Schüler:innen und ihre Rückfragen könne man erfahren, was diese bereits verstanden haben. Dementsprechend wird auch die Leistung in Übungsaufgaben als Stütze für didaktische Entscheidungen genannt (N=2). Daneben führen Lehrkräfte eigene Reflexionen nach einer Unterrichtsstunde, den beobachteten Umgang der Klasse mit verschiedenen Sozialformen sowie allgemeine Unterrichtsbeobachtungen an (jeweils N=3). Zwei Lehrkräfte nennen ihre Menschenkenntnis als eine Grundlage ihrer didaktischen Entscheidungen.

Auch die befragten Studierenden führen keine oder kaum Erfahrungen mit lernverlaufsdiagnostischen Informationen, in ihrem Fall während Schulpraktika oder sonstigen Einsätzen an (beruflichen) Schulen, an. Stattdessen schildern sie überwiegend das Handeln praktizierender Lehrkräfte, welche sie bspw. im Rahmen eines Schulpraktikums begleitet haben. Hier hätten Lehrkräfte z. T. Daten erhoben, um diese zur (mündlichen) Benotung zu verwenden (N=2). Lediglich zwei Studierende berichten davon, beobachtet zu haben, wie entsprechend der Grundidee formativer Diagnostik kleinere Aufgaben im Unterricht eingesetzt wurden, um ein Bild vom Lernstand der Schüler:innen zu erhalten (S18, S23). Ein:e Student:in (S11) erzählt von einer Lehrkraft, die vorliegende Daten nutzen wollte, um den Schüler:innen Rückmeldungen zu geben und diese zu motivieren. Ebenfalls einmal genannt wurde die Möglichkeit, sich selbst zu vergewissern, ob man Lerninhalte erfolgreich hat vermitteln können, um darauf aufbauend den weiteren Unterricht zu planen (So4). Im Anschluss wurden die Teilnehmenden gefragt, welche Informationen sie selbst für ihre didaktischen Entscheidungen genutzt

haben bzw. bei mangelnder schulpraktischer Erfahrung, worauf Lehrkräfte ihre didaktischen Entscheidungen stützen würden. Hierbei wurden in hoher Übereinstimmung mit ausgewerteten Antworten der Lehrkräfte die Erfahrung (N=15), das Bauchgefühl/die Intuition (N=10), allgemein Unterrichtsbeobachtungen (N=7) sowie die Zusammensetzung der Klasse (N=6) als häufige Stütze für Unterrichtsentscheidungen genannt. Außerdem wurden Klassenarbeitsergebnisse als didaktisch entscheidungsbeeinflussend identifiziert (N=2).

Im Gesamtblick zeigt sich zu Forschungsfrage 2.1, dass sowohl praktizierende als auch angehende Lehrkräfte über nur wenig anwendungspraktische Erfahrung lernverlaufsdiagnostischer Daten verfügen. Stattdessen stellen die von beiden Personengruppen häufig genannten Informationsgrundlagen für das Treffen didaktischer Entscheidungen die *persönliche* Erfahrung im pädagogischen Austausch mit den Schüler:innen und die Zusammensetzung der Klasse dar.

4.4 Zur Einstellung (angehender) Lehrkräfte zu lernverlaufsdiagnostischen Daten

Die angehenden und praktizierenden Lehrkräfte wurden im Sinne von *Forschungsfrage 2.2* aufgefordert, wünschenswerte Informationen über den Lernprozess von Schüler:innen zu benennen und zu erläutern, wie sie solche Informationen nutzen würden. Auch hierzu werden zunächst die Angaben praktizierender Lehrkräfte und im Anschluss diejenigen der angehenden Lehrkräfte inhaltlich strukturierend dargestellt. Auf diesem Wege soll sich der Einstellung der (angehenden) Lehrkräfte gegenüber der Nutzung lernverlaufsdiagnostischer Daten genähert werden. 18 von 33 Lehrkräften äußern sich auch skeptisch gegenüber der Nutzung lernverlaufsdiagnostischer Daten. Diese Skepsis gründet allerdings eher in der angezweifelte Gewinnung von Daten im laufenden Unterricht sowie (lerninhaltsabhängig) jeweils besonders interessierender Daten als einer Abwehr oder Abwertung der Nutzung diagnostischer Daten per se. So zweifeln neun Lehrkräfte an der Messbarkeit von Leistungen im Kontext der Lernverlaufsdiagnostik:

Ich sehe da allerdings einige Probleme bei der Operationalisierung, also wir haben jetzt hier bei dem Beispiel ging es jetzt ja um Buchungssätze. Das kann ich, glaube ich, ganz gut abbilden. Bei manchen Themen könnte das etwas schwieriger werden. (LK03)

Eine weitere Lehrkraft befürchtet Verzerrungen in Messungen bspw. aufgrund von Sprachschwierigkeiten (LK23). Ein weiterer Faktor, welcher den Nutzen der Lernverlaufsdiagnostik einschränkt, ist gemäß sechs praktizierenden Lehrkräften die Zeit, z. B. „Es ist aufgrund der Zeit, die man dafür aufwendet, im Moment nicht möglich für mich“ (LK05). Eine Lehrkraft (LK28) räumt zwar ein, dass man lernverlaufsdiagnostische Daten nutzen könne, um die Effektivität von Unterrichtsmethoden zu beurteilen, befürchtet aber, dass es die Methodenvielfalt zu stark einschränken würde, wenn immer nur noch die effektivste Unterrichtsmethode genutzt würde. Andere Lehrkräfte betonen, dass es in der Schule bzw. im Unterricht nicht nur um die Steigerung der Fachkompetenz ginge, sondern um die Interaktion in der Schule respektive das Lernen miteinander

(LK06) und die Sozialkompetenz (LK29). Zwei Lehrkräfte (LK22 & LK24) berichten von mangelnden Ressourcen zur Umsetzung von Lernverlaufsdagnostik und beziehen sich dabei auf die technische Umsetzung und eigene diagnostische Fähigkeiten. Die Auffassung, dass grundsätzlich nicht zu viele Daten erhoben werden sollten, vertreten drei Lehrkräfte (LK24, LK30, LK31), z. B. „ich bin grundsätzlich kein großer Freund davon viele Daten zu erheben [...]“ (LK24), „Also ich bin ja ein Freund von nicht zu vielen Daten erfassen, was ja in Big Data Welt, was dort ja so ein bisschen ausufert“ (LK30). Gemäß Lehrkraft 31 wären die Daten vielmehr für die Schüler:innen selbst interessant:

also ich würde erstmal selbst, glaube ich, gar nicht so viel von diesen Daten, wie es jetzt hier z. B. vorlieg[en], als wichtig erachten. Es ist, glaube ich, für die Schüler selbst eine schöne Möglichkeit sich selbst nochmal einzuschätzen. Schön ist auch irgendwo nochmal der Vergleich mit dem Klassenschnitt, wenn ich mir das Diagramm mir hier nochmal anschau, dann haben wir ja immer entsprechend eine Schülerin, die sich zum Klassenschnitt eben selbst einschätzen kann und dort eben nochmal sehen kann, wie sie im Vergleich zu den anderen liegt. (LK31)

Eine Lehrkraft (LK28) sieht den Nutzen von lernverlaufsdagnostischen Daten insbesondere in Lernfeldern, die man selbst noch nicht so häufig unterrichtet hat. In häufig unterrichteten Lernfeldern könnte man die Leistung der Schüler:innen schon gut durch „Herumgehen“ bei der Bearbeitung von Aufgaben beurteilen. Lediglich eine Lehrkraft bleibt allgemein skeptisch, lässt keine generell ablehnende Haltung erkennen, aber nennt gar keine positiven Aspekte im Umgang mit lernverlaufsdagnostischen Daten (LK09).

Insgesamt erkennen 32 von 33 Lehrkräften Nutzenpotentiale von lernverlaufsdagnostischen Daten. Knapp die Hälfte (15 von 33) sieht den Nutzen im Assessment von Unterrichtsmethoden, in der Analyse der eigenen Unterrichtsgestaltung bzw. in der hierauf basierenden Verbesserung derselben. Andere beschreiben

- die Aufnahme eines Ist-Zustandes des Leistungsstandes (LK01, LK03, LK11, LK12, LK14, LK23, LK28),
- die bei den Schüler:innen gesteigerte Motivation (LK05, LK13),
- das Feedback zur eigenen Einschätzung der Schüler:innen (im Sinne einer Selbst-evaluation durch die Schüler:innen) (LK30, LK31)
- die Sichtbarkeit der Lernentwicklung von Schüler:innen (LK05),
- die Information über Schwächen und Probleme der Schüler:innen (LK16)
- die mögliche Berücksichtigung des Lernzuwachses bei der Benotung (LK21),
- das Nutzen der Daten für ein individuelles Coaching-Gespräch (LK22),
- das Vergleichen der Daten mit Klassenarbeitsergebnissen (LK26),
- den verbesserten Austausch unter Lehrkräften zu verschiedenen Daten über Unterrichtsfächer und Lernfelder hinweg (LK17) und
- das Feedback über einzelne Schüler:innen und ihrem Lernen während bestimmter Unterrichtsmethoden (LK24)

als vorteilhaft.

Lehrkraft (LK06) zeigt ein Bewusstsein möglicher verzerrender Einflüsse auf die Erfassung fachlicher Fähigkeiten:

Ich würde aber gerne zusätzlich noch erfassen, wie ist das persönliche Empfinden irgendwie von der Person in den verschiedenen Phasen, also fühlt sich die Person wohl oder – ja also das finde ich auch noch wichtig, dass nicht nur die Leistung an sich gesehen wird in der fachspezifischen Sache, sondern das irgendwie noch erfasst wird, wie es dem Menschen geht in dieser Phase. Also das man das irgendwie an einem Punkt mit abfragt. Sind sie gestresst, gibt es Probleme im Betrieb, all so etwas. Vielleicht so 3, 4 Fragen hinterher. (LK06)

Drei Lehrkräfte (LK04, LK08, LK27) halten fest, dass sie sich solche Daten wünschen würden, wie sie auch im Erhebungsdesign dieser Studie dargelegt wurden. Lehrkraft 12 betont die Zeiteffizienz eines digitalen Instruments. Darüber hinaus wird eine breite Palette wünschenswerter Taktungen digital unterstützter Assessments thematisiert: Eine Lehrkraft (LK18) empfände eine Erhebung einmal im Monat, die ca. 10 Minuten Zeit in Anspruch nimmt, erstrebenswert. Eine andere Lehrkraft (LK19) würde sogar eine noch engmaschigere Datenerhebung (nach jeder Unterrichtsstunde) begrüßen, sieht darin allerdings die Gefahr der Monotonie und Langeweile bei den Schüler:innen. Eine weitere Lehrkraft (LK20) würde sich eine Art von Übersicht zum Wissensstand und Lernfortschritt alle zwei, drei Monate wünschen. Lehrkraft 30 würde eine Erhebung „ein paar Mal pro Schuljahr [als] so eine Art Fixstandpunkt“ (LK30) bevorzugen. Außerdem sollten Schüler:innen „sich [auch] selber mehr evaluieren [...] Denn gerade in der beruflichen Bildung ist das unser Ziel, dass sie alleine arbeiten können und dass wir sozusagen eher als Coaches im Hintergrund stehen und Unterstützung leisten“ (LK30). Mit Blick auf die diagnostischen Inhalte erstrecken sich die Vorstellungen ebenfalls auf ein breites Spektrum: Eine Lehrkraft (LK07) betont Erhebungen in prüfungs- und abschlussrelevanten Lernfeldern. Eine andere Lehrkraft (LK17) würde sich auch Daten von den Schulen, von denen die Schüler:innen kommen, wünschen, um bereits besser auf Schwierigkeiten von Schüler:innen reagieren zu können. Lehrkraft 32 würde sich vielmehr wünschen zu wissen, an welchen Unterrichtsmethoden die Schüler:innen Freude haben: „Für mich wäre der Spaßfaktor eigentlich das wichtigste, wenn die Spaß dran haben, dann funktioniert das auch“ (LK32). Lehrkraft 33 sieht Unterstützungsbedarf bei der Vergabe von mündlichen Noten und würde sich hierzu ein Tool wünschen.

Im Gesamtblick können über die erfassten Daten insbesondere Aussagen im Sinne der kognitiven Komponente der Einstellung getroffen werden. So speist sich die geäußerte Skepsis ganz überwiegend aus Überzeugungen hinsichtlich der Umsetzbarkeit, d. h. der Messbarkeit und Erhebung lernverlaufsdiagnostischer Daten, und nicht auf die Sinnhaftigkeit solcher Daten per se oder aber auf eine affektiv aufgeladene, grundsätzliche Ablehnung derselben. Die Mehrheit der Lehrkräfte sieht einen Nutzen in lernverlaufsdiagnostischen Daten, wobei die meisten diesen zur Adaption des Unterrichts an die Bedürfnisse der Schüler:innen nutzen würden. Dabei äußern die Lehrkräfte u. a. verschiedene Vorstellungen hinsichtlich des zeitlichen Rahmens für die Erfassung lernverlaufsdiagnostischer Daten und würden sich z. T. ergänzende Fragen abseits der fachspezifischer Aufgabenstellungen wünschen.

Auch die angehenden Lehrkräfte wurden befragt, wie nützlich sie Daten bzw. Graphen wie in dieser Studie einschätzen und wofür Lehrkräfte die Daten nutzen könnten. Ein Drittel der befragten Studierenden äußert sich skeptisch gegenüber der Nutzung lernverlaufsdiagnostischer Daten. Während ein:e Student:in allgemein der Meinung ist, man müsse nicht „immer versuchen alles messbar zu machen“ (S19), fragt ein:e andere:r Student:in nach der Sinnhaftigkeit der Erfassung lernverlaufsdiagnostischer Daten (So5). Student:in 09 hält lernverlaufsdiagnostische Daten nur für „bedingt nützlich, da [er/sie] glaub[t], dass viel mehr über Erfahrung und das direkte Gespräch und den direkten Kontakt mit den Schülern abläuft“ (So9). Sechs Studierende äußern Bedenken hinsichtlich der zeitlichen und technischen Umsetzbarkeit sowie des zusätzlichen Aufwands für die Lehrkraft (So4, So5, S12, S18, S22, S27). Proband:in S14 sieht eine Gefahr in lernverlaufsdiagnostischen Informationen, wenn diese vor dem Kennenlernen der Schüler:innen einsehbar sind und zu Vorurteilen führen. Ferner wird kritisch angemerkt, dass ein Test kein Gesamtbild der Schülerleistung bieten kann (S23). Außerdem ist ein:e Student:in der Meinung, dass lernverlaufsdiagnostische Daten erst dann hilfreich seien, wenn „signifikante Unterschiede“ (S27) zwischen Schülerleistungen bestehen würden. Im Gesamtblick zeigt sich, dass auch in der Gruppe der angehenden Lehrkräfte keine abwehrende Haltung gegenüber lernverlaufsdiagnostischen Informationen eingenommen wird. Soweit Bedenken geäußert werden (hier von einem Drittel der Befragten), betreffen diese aber nicht nur organisatorische Aspekte der Umsetzung, sondern auch die Aussagekraft und Interpretierbarkeit von diagnostischen Daten.

Trotz soeben genannter Kritikpunkte sehen *alle* befragten angehenden Lehrkräfte (auch) einen positiven Nutzen von lernverlaufsdiagnostischen Daten. Zu den mehrfach genannten Aussagen gehören:

- Wahl oder Anpassung der Unterrichtsform (So1, So8, S10, S11, S15, S26, S28)
- Leistung eines/r Schüler:in im Vergleich zur Klasse/zu anderen Klassen (So1, So9, S15, S17, S23)
- Erkennen, welchen Schüler:innen welche didaktische Variation am besten liegt (So2, So3, So6, So8, S12)
- Individuelle Förderung von Schüler:innen (So7, S18, S20, S24, S28)
- Veranschaulichung von Daten (So3, So5, S25, S28)
- Kontrolle der Zielerreichung (So6, S13, S17, S29)
- Analyse von Lernschwächen (So4, S10, S18, S27)
- Schüler:innen Feedback geben (So4, S21, S22)

Der besondere Nutzen lernverlaufsdiagnostischer Daten liegt gemäß den angehenden Lehrkräften damit insbesondere in der Anpassung bzw. reflektierten Wahl von Unterrichtsformen (N=5). Weitere genannte Vorzüge gehen in eine ähnliche Richtung, wenn bspw. gesagt wird, dass Schüler:innen individuell gefördert werden könnten, erkannt werden kann, welche didaktische Variation am lernförderlichsten ist etc.

Im Gesamtblick zeigt sich zu Forschungsfrage 2.2, dass lernverlaufsdiagnostischen Daten insbesondere hinsichtlich ihrer Erhebung mit Skepsis entgegengetreten wird; dennoch sprechen beide Gruppen lernverlaufsdiagnostischen Daten ganz überwiegend

einen pädagogisch-didaktischen Nutzen zu. Sowohl praktizierende als auch angehende Lehrkräfte würden die Daten primär zur Unterrichtsverbesserung nutzen.

5. Schlussteil

5.1 Zusammenfassung und Diskussion der zentralen Ergebnisse

In unserer Studie zur Interpretation lernverlaufsdiagnostischer Daten durch angehende und praktizierende Lehrkräfte an beruflichen Schulen zeigte sich übereinstimmend mit den Ergebnissen von ZEUCH et al. (2017), dass die Proband:innen beider Gruppen häufig den Schritt des Beschreibens der Daten (*Read the data*) übersprangen und direkt Vergleiche vornahmen (*Read between the data*). Vergleiche der Leistung der fiktiven Schüler:in Lena mit der Klasse waren dabei vorrangig. Dadurch liegt der Schluss nahe, dass die Lehrkräfte unserer Studie bevorzugt auf eine *soziale Bezugsnorm* zur Beurteilung der Leistungsentwicklung eines/einer Schüler:in zurückgreifen, indem sie diese schnell in den klasseninternen Referenzrahmen einordnen, statt die betreffende Schüler:in zunächst individuell über verschiedene Messzeitpunkte hinweg zu betrachten und dann (auch oder vorrangig) in Relation zu curricularen bzw. kriterialen Zielen zu setzen.

Die Aussagen zum Diagramm sind mit durchschnittlich 40 % der interpretativen Schritte in korrekter (idealtypischer) Reihenfolge nur eingeschränkt *sequentiell kohärent*. Dieses Ergebnis korrespondiert mit Befunden von BOSCH et al. (2017), scheint mit Blick auf den inhaltlichen Gehalt dieser Aussagen aber weniger interessant als deren *Vollständigkeit* und *Akkuratheit*. Diese Kategorien funktionaler Lesefähigkeiten sind nicht nur empirisch separat analysierbar, sondern ebenso wie ein strukturell sauberes Vorgehen gezielt in der Lehrkräfteaus- und -fortbildung förderbar. Eine direkte Gegenüberstellung beider Lehrkräftegruppen legt dabei nahe, dass sich deren inhaltliche Interpretationsleistungen auf ähnlich hohem Niveau bewegen und eher punktuelle Defizite in bestimmten Schritten bzw. Stufen (*framing, read the data, read between the data* und *read beyond the data*) aufweisen: Über alle Stufen hinweg konnte eine Akkuratheit von durchschnittlich 88,9 % (praktizierende Lehrkräfte) bzw. 87,4 % (angehende Lehrkräfte) festgestellt werden. Diese Werte korrespondieren mit den in der Studie von ESPIN et al. (2017) ermittelten Ergebnissen für Lehrkräfte aus dem allgemeinbildenden Schulsektor. In Abweichung zu der von WAGNER et al. (2017) ermittelten Vollständigkeit der Interpretation lernverlaufsdiagnostischer Daten durch Lehrkräfte im Bereich der Sonderpädagogik in Höhe von durchschnittlich 34,6 % der erreichbaren Punktzahl, konnte in der vorliegenden Studie eine höhere Vollständigkeit der Interpretation festgestellt werden. Angehende Lehrkräfte im Bereich der Wirtschaftspädagogik erreichten durchschnittlich 84,3 % der Punkte im Rahmen der Vollständigkeit, praktizierende Lehrkräfte erreichten durchschnittlich 72,1 %.

Auch wenn diese Durchschnittswerte im Vergleich zu Vorgängerstudien bemerkenswert positiv ausfallen, lassen sich zwei punktuelle Schwächen konkret identifizieren: Allen Teilnehmenden mangelt es am ehesten auf der höchsten Anforderungsstufe der

funktionalen Lesefähigkeit (*read beyond the data*) an Vollständigkeit und Akkuratheit (75,75 % und 77,5 %). Ferner übersehen auf der basalen Anforderungsstufe „*read the data*“ sowohl angehende als auch praktizierende Lehrkräfte grundlegende Informationen grafisch verdichteter Lernverlaufsdaten (durchschnittliche Vollständigkeit: 58,5 %). Dies könnte schlussfolgernde Überlegungen natürlich von vorneherein beeinträchtigen. Vor dem Hintergrund, dass jede Lehrkraft in der Lage sein sollte mit lernverlaufsdagnostischen Informationen angemessen umzugehen und diese für den weiteren Unterricht gewinnbringend zu nutzen, gibt es damit konkrete Verbesserungsmöglichkeiten. Mit Blick auf die primäre (nämlich lernunterstützende) Funktion der Lernverlaufsdagnostik ist zudem festzuhalten, dass lediglich 60 % bzw. 77 % der praktizierenden bzw. angehenden Lehrkräfte die für die fiktive Schülerin lernförderlichste didaktische Variation identifizieren und erste, d. h. auf die verfügbaren Daten gestützte, Überlegungen zur Gestaltung nachfolgender Unterrichtseinheiten anstellen. Hervorzuheben ist hier außerdem, dass nicht relativ mehr praktizierende, sondern relativ mehr angehende Lehrkräfte diesen relevanten Aspekt thematisieren. Dies widerspricht der gängigen Annahme, praktizierenden Lehrkräften pauschal eine höhere Expertise im Umgang mit lernverlaufsdagnostischen Informationen zuzuschreiben. ZEUCH et al. (2017) vermuten angesichts der auch in ihrer Studie fehlenden Unterschiede zwischen praktizierenden und angehenden Lehrkräften, dass Erstere ihr reichhaltiges erfahrungsbasiertes (und wohl teilweise implizites) Wissen nicht mit den formalisierten Diagnosedaten verknüpfen können. Die Think-Aloud-Protokolle unserer Studie legen eine andere Hypothese nahe: Tatsächlich schaffen die Proband:innen der vorliegenden Studie häufig explizite Bezüge zu eigenen Unterrichtserfahrungen. Diese führen jedoch eher dazu, dass Gründe für eine Dynamik in den Lernverlaufskurven außerhalb der im Diagramm verdichteten Informationen gesucht werden, indem bspw. über die Klassenmitglieder spekuliert wird, anstelle die Kurvenverläufe selbst detaillierter auszuwerten. ZEUCH et al. (2017) stellten in ihrer Untersuchung weiterhin fest, dass Teilnehmende mit Studium der Mathematik bessere Ergebnisse erzielten als Teilnehmende mit anderen fachlichen Vertiefungen. Da Studierende und Absolvent:innen der Wirtschaftspädagogik wenigstens im Rahmen ihres Wirtschaftsstudiums ebenfalls verstärkt mit Mathematik in Berührung kommen bzw. kamen, ist es denkbar, dass diese auf Grund ihres möglicherweise besseren mathematischen Verständnisses weniger Schwierigkeiten bei der Interpretation lernverlaufsdagnostischer Daten haben als andere Proband:innen. Hierzu sind jedoch weitere Untersuchungen vonnöten. Eine Überlegenheit von Studierenden mit Nebenfach Mathematik gegenüber Studierenden mit anderen Nebenfächern zeigte sich in der vorliegenden Stichprobe jedenfalls nicht.

Bezüglich der Einstellungen angehender und praktizierender Lehrkräfte im berufsbildenden Bereich zeichnet sich in unseren interviewbasiert gewonnenen Befunden ab, dass beide Gruppen vielfältige pädagogisch-didaktische Verwertungspotentiale lernverlaufsdagnostischer Daten erkennen. Korrespondierend mit den in der Fachliteratur definierten Kernfunktionen formativer Assessments (vgl. Kapitel 2) sehen sie den Nutzen lernverlaufsdagnostischer Daten vor allem bei der Evaluation des eigenen Unterrichts und der Identifikation von Ansatzpunkten einer verbesserten Unterrichtsge-

staltung. Aber auch die Bestimmung individueller Bedarfe der Lernunterstützung oder die Kontrastierung von Selbst- und Fremdeinschätzungen der Schüler:innen gehören hierzu. Vereinzelt werden jedoch auch Fehlvorstellungen (kognitive Komponente der Einstellung) geäußert, etwa die Verwendung formativer Assessments zu Zwecken der Notengebung oder ihre Gleichsetzung mit bilanzierenden Lernstandserhebungen mit nur wenigen Erhebungszeitpunkten während eines Schuljahres.

Trotz dieser geteilten und vielfältigen Nutzenzuschreibungen durch alle Proband:innen werden lernverlaufsdiagnostische Daten zumindest in deren berufsschulischer Praxis

- (a) ... *nur vereinzelt erhoben und noch eher unsystematisch ausgewertet* (vgl. Kapitel 4.2). Hierzu nannten die praktizierenden Lehrkräfte von den Schüler:innen auszufüllende Kompetenzraster im Lernbüro, (Recht-)Schreibübungen im Deutschunterricht und Moodle-Tests. Studierende berichteten aus ihren Schulpraktika bspw. den Einsatz kleiner Übungsaufgaben, um ein Bild vom aktuellen Lern-/Leistungsstand der Schüler:innen zu erhalten.
- (b) ... *skeptisch bewertet hinsichtlich ihres Planungs- und Durchführungsaufwands*. Einige Lehrkräfte berichten Bedenken mit Blick auf einen Mangel an Ressourcen (z. B. zeitlich, technisch). Einzelne Studierende äußern zudem Sorgen über ergebnisverzerrende Einflüsse oder eine eingeschränkte Validität erhobener Daten.
- (c) ... *nicht nur als weniger einfach zugänglich, sondern auch weniger vertrauenswürdig oder zumindest keinesfalls vorzugswürdig gegenüber persönlichen Eindrücken von Schüler:innen und erfolgreichen Unterrichtsmethoden bewertet*. Unabhängig von Fragen der Begründetheit und Validität eben dieser im Zuge unterrichtlicher Interaktionen reifenden Eindrucksbildung äußern die befragten angehenden und praktizierenden Lehrkräfte an berufsbildenden Schulen damit eine Präferenz für eine intuitive didaktische Entscheidungsfindung gegenüber einer datenbasierten Unterrichtsplanung (vgl. Kapitel 2.1 und 4.3). Dabei gelten die identifizierten „Bedürfnisse“ der Schüler:innen als wesentlicher Bezugspunkt.

5.2 Limitationen der Studie und Implikationen für weitere Forschung

Einschränkend ist zu den Ergebnissen der vorliegenden Studie anzumerken, dass diese sich geografisch auf ein Bundesland und überwiegend auf Studierende der Universität Göttingen beschränkt. Obgleich die Methode *Think Aloud* bei der Diagramminterpretation den Vorteil hat, dass Proband:innen kaum vom/von der Interviewer:in beeinflusst werden, birgt dies auch den Nachteil, dass keine inhaltlichen Nachfragen zugelassen werden. So bleibt offen, ob Proband:innen bei gezieltem Nachfassen in der Lage gewesen wären, die abgebildeten didaktischen Variationen zu beurteilen, obwohl sie sich im Zuge des lauten Denkens nicht eigeninitiativ dazu äußern. So kann für rund ein Drittel der Proband:innen, die keine didaktischen Überlegungen anstellten, nicht eindeutig beurteilt werden, ob sie diesen Überlegungen bei der Betrachtung von Lernverlaufskurven „nur“ keinerlei Aufmerksamkeit schenken oder ob es ihnen schwerfällt, die im

Kurvendiagramm enthaltenen Informationen zu differenziellen Effekten eingesetzter Unterrichtsmethoden für verschiedene Schüler:innen zu enkodieren. In einer weiteren Untersuchung wären demnach halbstrukturierte Interviews mit Interviewleitfäden gewinnbringend (vgl. DÖRING/BORTZ, 2016).

Ferner wurde den Proband:innen nur ein Diagramm mit lernverlaufsdiagnostischen Informationen vorgelegt. Im Rahmen einer weiteren Untersuchung könnten den Teilnehmenden mehrere, unterschiedlich komplexe Diagramme gezeigt werden, die noch differenziertere Einblicke in funktionale Lesefähigkeiten gewähren und zudem erweiterte Möglichkeiten einer systematischen Analyse von Bedingungsfaktoren unterschiedlicher Interpretationsleistungen bieten. Weiterhin wurde das vorgelegte Diagramm für fiktive Schüler:innen erstellt. Weitere Forschung könnte sich auf reale lernverlaufsdiagnostische Daten, bestenfalls der eigenen Schüler:innen teilnehmender Lehrkräfte stützen. Eine besondere Herausforderung dieser alternativen Vorgehensweise läge allerdings darin, auch die realen Lernvoraussetzungen und Klassenbedingungen zu kennen und in die Analysen einfließen zu lassen. Um Aussagen etwa über die funktionalen Lesefähigkeiten von Lehrkräften auf der Anforderungsstufe *read beyond the data* treffen zu können, müsste stets auch berücksichtigt werden, ob die über das Diagramm hinaus einbezogenen Informationen auch sachlich zutreffend und relevant für die Urteilsbildung bzw. Entscheidungsfindung sind.

Schließlich können unsere vergleichenden Auswertungen mit Blick auf spezifische Erfahrungshintergründe der Proband:innen (z.B. studiertes Nebenfach, bereits eingesetzte Diagnoseinstrumente) und deren potentielle Effekte auf Interpretationsleistungen von oder Einstellungen gegenüber lernverlaufsdiagnostischen Daten aufgrund der Stichprobengröße nur der Hypothesengenerierung dienen und noch keinen erklärenden Anspruch erheben. So wären bspw. die hier ermittelten systematischen und mittelstarken Abweichungen im sequentiell kohärenten Vorgehen zwischen Personen, die Lehrveranstaltungen zur pädagogisch-psychologischen Diagnostik nicht oder eben schon besucht haben, in Folgestudien mit höherer Proband:innenzahl ebenso zu erhärten wie die in einer Vorgängerstudie ermittelten Unterschiede in den Interpretationsleistungen von Personen mit oder ohne mathematische Vertiefungsrichtung.

Literaturverzeichnis

- ACHTENHAGEN, F. / WEBER, S. (2015). „Ck“, „pck“, „pk“ – Oder: Was Sie schon immer über Fachdidaktik wissen wollten ... In A. RAUSCH / J. WARWAS / J. SEIFRIED / E. WUTTKE (Hrsg.), *Konzepte und Ergebnisse ausgewählter Forschungsfelder der beruflichen Bildung – Festschrift für Detlef Sembill* (S. 89–105). Baltmannsweiler: Schneider Verlag Hohengehren.
- BEHRMANN, L. / SOUVIGNIER, E. (2013). The Relation Between Teachers' Diagnostic Sensitivity, their Instructional Activities, and their Students' Achievement Gains in Reading. *Zeitschrift für Pädagogische Psychologie*, 27(4), 283–293.
- BLACK, P. / WILIAM, D. (1998). *Assessment and Classroom Learning*. Abgerufen am 11. Januar 2022 <https://www.gla.ac.uk/t4/learningandteaching/files/PGCTHE/BlackandWiliam1998.pdf>.

- BÖHMER, M. / ENGLISH, B. / BÖHMER, I. (2017). Schülerbeurteilungen aus der Perspektive dualer Prozessmodelle der sozialen Urteilsbildung. In A. SÜDKAMP / A.-K. PRAETORIUS (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 50–54). Waxmann.
- BÜHNER, M. / ZIEGLER, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler*. (2. Aufl.). Pearson Deutschland.
- DATNOW, A. / HUBBARD, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, 17(1), 7–28.
- DEMSKI, D. (2017). *Evidenzbasierte Schulentwicklung: Empirische Analyse eines Steuerungsparadigmas*. Springer VS.
- DÖRING, N. / BORTZ, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. (5. Aufl.). Springer.
- ERCAN, H. / HARTMANN, U. / RICHTER, D. / KUSCHEL, J. / GRÄSEL, C. (2021). Effekte von integrierter Führung auf die Datennutzung von Lehrkräften. *Die deutsche Schule* 113 (1), 85–100.
- ESPIN, C. / WAYMAN, M. / DENO, S. / MCMASTER, K. / DE ROOIJ, M. (2017). Data-Based Decision-Making: Developing a Method for Capturing Teachers' Understanding of CBM Graphs. *Learning Disabilities Research/Practice*, 32(1), 8–21.
- FÖRSTER, N. / SOUVIGNIER, E. (2017). Förderung diagnostischer Kompetenz durch Bereitstellung formative Diagnostik. In A. SÜDKAMP / A.-K. PRAETORIUS (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 231–239). Waxmann.
- FUCHS, L. S. / FUCHS, D. / HAMLETT, C. L. (2015). Republication of "Curriculum-Based Measurement: A Standardized, Long-Term Goal Approach to Monitoring Student Progress". *Intervention in School and Clinic*, 50(3), 185–192.
- HERITAGE, M. (2007). Formative Assessment: What Do Teachers Need to Know and Do?. *Phil Delta Kappan*, 89. Abgerufen am 03. April 2019 von http://easlinstitute.org/wp-content/uploads/Heritage_formative_assessment.pdf.
- INGENKAMP, K. / LISSMANN, U. (2008). *Lehrbuch der pädagogischen Diagnostik*. (6. Aufl.). Weinheim und Basel: Beltz Verlag.
- KÄRNER, T. / WARWAS, J. / SCHUMANN, S. (2021). A learning analytics approach to address heterogeneity in the classroom: The teachers' diagnostic support system. *Technology, Knowledge and Learning*, 26(1), 31–52.
- KONRAD, K. (2020). Lautes Denken. In G. MEY / K. MRUCK (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie*. Springer.
- KULTUSMINISTERKONFERENZ (2014). *Standards für die Lehrerbildung: Bildungswissenschaften*. Abgerufen am 03. April 2019 von https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf.
- MANDINACH, E. B. (2012). A Perfect Time for Data-Use: Using Data-Driven Decision Making to Inform Practice. *Educational Psychologist*, 47(2), 71–85.
- MAYRING, P. (2002). *Einführung in die qualitative Sozialforschung*. (5. Aufl.). Weinheim und Basel: Beltz Verlag.
- SCHILDKAMP, K. / KUIPER, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education* (26), 482–496.
- SCHUMANN, M. (2021) Chancen und Herausforderungen IT-basierter Lern- und Assessmentssysteme in der kaufmännischen Ausbildung. In K. BECK / F. OSER (Hrsg.), *Resultate und Probleme der Berufsbildungsforschung* (S. 121–140). Wbv.
- SCHÜTZE, B. / SOUVIGNIER, E. / HASSELHORN, M. (2018). Stichwort – Formatives Assessment. *Zeitschrift für Erziehungswissenschaft*, 21, 697–715.
- SIEGFRIED, C. / ACKERMANN, N. (2020). Die Bedeutung curricularer Lerngelegenheiten im gymnasialen Bereich für die wirtschaftsbürgerliche Kompetenz. *Zeitschrift für ökonomische Bildung*, 9, 1–31.

- SOUVIGNIER, E. / FÖRSTER, N. / SCHULTE, E. (2014). Wirksamkeit formativen Assessments: Evaluation des Ansatzes der Lernverlaufsdiagnostik. In M. HASSELHORN / W. SCHNEIDER / U. TRAUTWEIN (Hrsg.), *Lernverlaufsdiagnostik* (S. 221–237). Hogrefe Verlag.
- STIGLER, J. W. / MILLER, K., F. (2018). Expertise and Expert Performance in Teaching. In K. ERICSSON / R. HOFFMAN / A. KOZBELT / A. WILLIAMS (Hrsg.), *The Cambridge Handbook of Expertise and Expert Performance* (Cambridge Handbooks in Psychology, pp. 431–452). Cambridge: Cambridge University Press. doi:10.1017/9781316480748.024.
- VAN DEN BOSCH, R. M. / ESPIN, C. A. / PAT-EL, R. J. / SAAB, N. (2019). Improving Teachers' Comprehension of Curriculum-Based Measurement Progress-Monitoring Graphs, *Journal of Learning Disabilities*, 52(5), 413–427.
- VAN DEN BOSCH, R. M. / ESPIN, C. A. / CHUNG, S. / SAAB, N. (2017). Data-based decision-making: Teachers' comprehension of curriculum-based measurement progress-monitoring graphs, *Learning Disabilities Research/practice*, 32(1), 46–60.
- VAN OPHUYSEN, S. / LINTORE, K. (2013). Pädagogische Diagnostik im Schulalltag. In S.-I. BEUTEL / W. BOS / R. PORSCH (Hrsg.): *Lernen in Vielfalt: Chance und Herausforderung für Schul- und Unterrichtsentwicklung*. Waxmann.
- WAGNER, D. / HAMMERSCHMIDT-SNIDARICH, S. / ESPIN, C. / SEIFERT, K. / MCMASTER, K. (2017). Pre-service Teachers' Interpretation of CBM Progress Monitoring Data. *Learning Disabilities Research/Practice*, 32(1), 22–31.
- WARWAS, J. / KÄRNER, T. / GOLYSZNY, K. (2015). Diagnostische Sensibilität von Lehrpersonen im Berufsschulunterricht: Explorative Prozessanalysen mittels Continuous-State-Sampling. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111(3), 437–454.
- WARWAS, J. / SCHIRMER, H., / HELM, C. (2020). Auf dem Weg zu einem kognitiven Diagnosemodell für eine gezielte Lernunterstützung im Rechnungswesen-Unterricht. *berufsbildung* 74, 29–32.
- ZEUCH, N. / FÖRSTER, N. / SOUVIGNIER, E. (2017). Assessing Teachers' Competencies to Read and Interpret Graphs from Learning Progress Assessment: Results from Tests and Interviews. *Learning Disabilities Research/Practice*, 32(1), 61–70.
- ZIEGLER, B. / FREY, A. / SEEGER, S. / BALKENHOL, A. / BERNHARDT, R. (2016). Adaptive Messung allgemeiner Kompetenzen (MaK-adapt). In K. BECK / M. LANDENBERGER / F. OSER (Hrsg.): *Technologiebasierte Kompetenzmessung in der beruflichen Bildung*. W. Bertelsmann Verlag.
- ZLATKIN-TROITSCHANSKAIA, O. / FÖRSTER, M. / PREUSSE, D. / MATER, O. (2016). The relationship between teachers' evidence-based actions and communication, cooperation, and participation structures at schools. *Journal for Educational Research*, 8(3), 59–79.

WIEBKE VORPAHL

Universität Hohenheim, Professur für Wirtschaftspädagogik, insbesondere Theorie und Didaktik beruflicher Bildung, Fruwirthstr. 47, 70593 Stuttgart,
wiebke.vorpahl@uni-hohenheim.de

PROF. DR. JULIA WARWAS

Universität Hohenheim, Professur für Wirtschaftspädagogik, insbesondere Theorie und Didaktik beruflicher Bildung, Fruwirthstr. 47, 70593 Stuttgart,
julia.warwas@uni-hohenheim.de