



Förderung diagnostischer Kompetenz im Rahmen der Ausbildung von Lehrkräften für die Primarschule – Eine Studie zum Erkennen von Stärken und Schwächen von Schülerinnen und Schülern am Beispiel von Größen

K. Philipp · I. Gobeli-Egloff

Eingegangen: 28. Dezember 2020 / Angenommen: 25. Februar 2022 / Online publiziert: 4. März 2022
© Der/die Autor(en) 2022

Zusammenfassung Diagnostische Kompetenz umfasst eine Vielzahl unterschiedlicher Fähigkeiten, die eine hohe Praxisrelevanz besitzen. Die Fähigkeit, Stärken und Schwächen von Schülerinnen und Schülern zu erkennen, wird insbesondere für die individuelle Förderung als fundamental betrachtet. Der diagnostische Prozess kann als Informationsverarbeitungsprozess verstanden werden, bei dem sowohl vorhandenes (fachbezogenes) Wissen als auch Informationen aus einer diagnostischen Situation aufeinander bezogen werden müssen, um Stärken und Schwächen von Schülerinnen und Schülern erkennen zu können. Ein Konzept zur Förderung dieser Teilfähigkeit diagnostischer Kompetenz am Beispiel Größen wurde im Rahmen einer Interventionsstudie mit angehenden Lehrkräften für die Primarschule ($N=181$) hinsichtlich seiner Wirksamkeit quantitativ untersucht. Es zeigte sich, dass das Erkennen von Stärken und Schwächen durch die Intervention nachhaltig gefördert werden konnte. Die Art und die Nutzung von Informationen beim Diagnostizieren von Stärken und Schwächen von Schülerinnen und Schülern wurde in einer qualitativen Analyse schriftlicher Dokumente der Studierenden untersucht. Wesentliches Ergebnis der qualitativen Herangehensweise ist ein Kategoriensystem, das über die Art der verwendeten Informationen Aufschluss gibt, und möglicherweise auf weitere Inhaltsbereiche im Fach Mathematik übertragen werden kann. Insgesamt geben die Ergebnisse Hinweise auf den Erwerb diagnostischer Kompetenz.

Schlüsselwörter Diagnosekompetenz · Größenvorstellungen · Intervention · Ausbildung von Lehrkräften

Mathematics Education Subject Classification 97B50 · 97D72

K. Philipp (✉) · I. Gobeli-Egloff
Pädagogische Hochschule Nordwestschweiz, Hofackerstr. 30, 4132 Muttenz, Schweiz
E-Mail: kathleen.philipp@fnw.ch

Fostering Diagnostic Competence in Primary Teacher Education – a Study of Recognizing Strengths and Weaknesses of Students in the Field of “Measurement”

Abstract Diagnostic competence includes many different abilities that are highly relevant for practice. The ability to recognize the strengths and weaknesses of students is seen as fundamental for individual support of students. The diagnostic process can be understood as information processing in which it is necessary to interlink the (subject-related) knowledge and information from a diagnostic situation to recognize the strengths and weaknesses of students. A concept to foster this facet of diagnostic competence in the field of “measurement” was investigated with quantitative methods regarding its effectiveness in an intervention study with pre-service teachers for primary school ($N=181$). The results indicate that the facilitation of the recognition of strengths and weaknesses was successful and sustainable. A qualitative analysis based on written documents from students was derived to examine the nature and the use of information in assessing the strengths and weaknesses of students. Main result of the qualitative analyses is a category system that sheds light on the nature of used information and which can possibly be transferred to other fields in the subject of mathematics. Overall, the results provide indications for the acquirement of diagnostic competence.

Keywords Diagnostic competence · Conceptions of measurement · Intervention study · Teacher education

1 Einleitung

Diagnostische Kompetenz hat in der Diskussion um professionelle Kompetenzen von Lehrkräften gegenwärtig einen hohen Stellenwert. Sie gilt als Voraussetzung dafür, Unterricht an die Bedürfnisse der Lernenden anpassen zu können (Anders et al. 2010; Helmke et al. 2004; Schwarz et al. 2008). Ihr wird damit eine zentrale Bedeutung als Komponente der Professionalität von Lehrkräften beigemessen (von Aufschnaiter et al. 2015; Baumert und Kunter 2006; Helmke 2009; Weinert 2000). Die Anpassung des Unterrichts an die Lernvoraussetzungen der Lernenden wiederum wird als Voraussetzung für deren Lernerfolg gesehen (Schrader 2013). Diagnostische Kompetenz soll dabei helfen, einerseits Schwierigkeiten, andererseits aber auch Kompetenzen bzw. Potenziale von Lernenden zu identifizieren, um sie für eine adäquate individuelle Förderung nutzen zu können (Helmke 2009; Rösike und Schnell 2017). Die Fähigkeit, sowohl Stärken als auch Schwächen von Schülerinnen und Schülern zu erkennen, wird daher als Teil diagnostischer Kompetenz und als grundlegend für die Begleitung und Unterstützung von Lernprozessen im Sinne der individuellen Förderung aufgefasst. Das Erkennen von Stärken und Schwächen wird als Fähigkeit betrachtet, die einerseits das Identifizieren von fehlerhaften bzw. richtigen Überlegungen von Schülerinnen und Schülern beinhaltet und andererseits deren Interpretation, beispielsweise auf welche dahinterliegenden (Fehl-)Vorstellungen ihre Überlegungen verweisen können (Philipp 2018). Die Fähigkeit wird benötigt,

um den Grad des Verständnisses von Schülerinnen und Schülern zu erfassen und bildet die Voraussetzung dafür, angepasste Fördermaßnahmen ergreifen zu können, auch wenn sich aus einer Diagnose selbst nicht direkt Maßnahmen ableiten lassen (Moser Opitz und Nührenböcker 2015; Wember 1998). Das Ziel einer solchen Diagnostik ist die Lernförderung (assessment for learning) im Gegensatz zur summativen Bewertung einer Leistung (assessment of learning) (Black und Wiliam 2009).

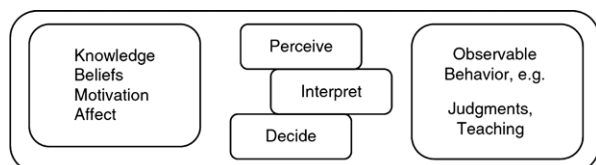
Der vorliegenden Studie liegt die Annahme zugrunde, dass der Aufbau fachbezogenen Wissens und die Auseinandersetzung mit praxisnahen diagnostischen Situationen den Erwerb diagnostischer Kompetenz begünstigt. Konkret wird am Beispiel Größen untersucht, (1) inwiefern sich im Rahmen der Ausbildung von Lehrkräften für die Primarschule durch eine entsprechend gestaltete Intervention die Fähigkeit des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern fördern lässt und (2) welche Informationen die Studierenden beim Diagnostizieren von Stärken und Schwächen von Schülerinnen und Schülern nutzen. Ziel des Beitrags ist es, das Verständnis des Aufbaus diagnostischer Kompetenz im Hinblick auf die benötigte Informations- und Wissensbasis zu erweitern.

2 Theoretischer Hintergrund

2.1 Diagnostische Kompetenz im Fach Mathematik

Blömeke et al. (2015) betrachten Kompetenz als Kontinuum bzw. einen Prozess und schlagen ein Modell vor, das die Transformation von Kompetenz in Performanz über situationsbezogene Fähigkeiten der Wahrnehmung, Interpretation und Entscheidungsfindung beschreibt. Diese Fähigkeiten vermitteln zwischen dem beobachtbaren Verhalten (Performanz) und den kognitiven und affektiv-motivationalen Personeneigenschaften. Bezogen auf diagnostische Kompetenz umfassen solche Personeneigenschaften Wissen, Einstellungen sowie affektive und motivationale Aspekte. Die situationsbezogenen Fähigkeiten sind abhängig von den Personeneigenschaften und führen zu beobachtbarem Verhalten in diagnostischen Situationen, z. B. diagnostischen Urteilen (vgl. Abb. 1). Damit bietet das Modell einen strukturellen Rahmen, der eine breite Auffassung von diagnostischer Kompetenz und verschiedene Forschungsansätze zulässt: (1) die Untersuchung von Personeneigenschaften, die ein erfolgreiches Agieren in diagnostischen Situationen ermöglichen (2) die Untersuchung kognitiver Prozesse, die das Zustandekommen diagnostischer Urteile erklären und (3) die Untersuchung von diagnostischen Urteilen und Tätigkeiten von Lehrkräften in Bezug auf das Lernen von Schülerinnen und Schülern (Leuders et al. 2018).

Abb. 1 Diagnostische Kompetenz als Kontinuum. (Aus: Leuders et al. 2018, S. 9)



Je nach Untersuchungsgegenstand kann daher das Verständnis von diagnostischer Kompetenz und ihrer Bestandteile unterschiedlich sein (Förster und Karst 2017; Leuders et al. 2018). Eine dahingehend offene Definition liefert Weinert (2000): Diagnostische Kompetenz kann verstanden werden als ein „Bündel von Fähigkeiten, um den Kenntnisstand, die Lernfortschritte und die Leistungsprobleme der einzelnen Schüler sowie die Schwierigkeiten verschiedener Lernaufgaben im Unterricht fortlaufend beurteilen zu können, sodass das didaktische Handeln auf diagnostischen Einsichten aufgebaut werden kann“ (Weinert 2000, S. 14 f.). In dieser Definition wird deutlich, dass verschiedene diagnostische Anforderungen auch verschiedene Fähigkeiten erfordern können. Dass es sich bei diagnostischer Kompetenz um ein „Bündel“ von Fähigkeiten anstelle einer allgemeinen (fachübergreifenden) Kompetenz handelt, wird durch empirische Befunde untermauert (Brunner et al. 2011; Spinath 2005). In Anbetracht der großen Vielfalt an diagnostischen Situationen (Ingenkamp und Lissmann 2008; Karst et al. 2017) erscheint es plausibel, dass verschiedene Fähigkeiten erforderlich sind.

Die Qualität eines diagnostischen Urteils wird meist mittels dessen Urteilsakkuratheit bestimmt (Hoge und Coladarci 1989; Südkamp et al. 2012), d.h. es wird die möglichst genaue Übereinstimmung des Urteils einer Lehrkraft mit der tatsächlichen Merkmalsausprägung (z. B. Lösungshäufigkeit bei einer konkreten Aufgabe) betrachtet. Dies setzt voraus, dass das interessierende Merkmal gemessen werden kann. Im schulischen Kontext sind diagnostische Urteile jedoch häufig komplex und eine Einschätzung der Urteilsakkuratheit nicht sinnvoll möglich (van Ophuysen und Behrmann 2015). Beim Erkennen von Stärken und Schwächen von Schülerinnen und Schülern ist es wichtig, dass eine Lehrkraft anhand fachbezogener Kriterien möglichst viele Aspekte, die zu richtigen bzw. fehlerhaften Überlegungen und Lösungen führen können, berücksichtigen kann. Dazu muss sie diese sowohl identifizieren als auch interpretieren können (Philipp 2018). Das bedeutet einerseits, dass die Überlegungen der Schülerinnen und Schüler als richtig oder falsch beschrieben, aber darüber hinaus auch im Hinblick auf das zugrundeliegende mathematische Verständnis gedeutet werden müssen (Jacobs et al. 2010), wobei solche Deutungen durch angehende Lehrkräfte häufig noch ungenau und eher allgemein sind (Pott 2019). Vor diesem Hintergrund ist davon auszugehen, dass diagnostische Kompetenz domänen-, wenn nicht gar inhaltspezifisch ist (z. B. Lorenz und Artelt 2009). Insbesondere mit Blick auf das Ziel der individuellen Förderung spielen fachbezogene Aspekte bei einer Diagnose und damit das fachbezogene Wissen eine Rolle (Chernikova et al. 2020). Bei einer solchen Diagnose muss der (individuelle) Lernprozess der Schülerinnen und Schülern selbst fokussiert werden, der eine genaue Analyse des Lernstands ermöglicht, indem beispielweise Aussagen oder schriftliche Produkte herangezogen werden. Prediger et al. (2012) sprechen in diesem Zusammenhang von „diagnostischer Tiefenschärfe“, die unter anderem auch fachdidaktisches Wissen wie beispielweise Wissen über Grundvorstellungen oder typische Fehlvorstellungen voraussetzt.

Modelle zur Konzeptualisierung diagnostischer Kompetenz im Fach Mathematik verdeutlichen die enge Verknüpfung von diagnostischen und (fach)didaktischen Kompetenzen, mitunter scheinen sie kaum voneinander trennbar, was insbesondere bei der Erfassung diagnostischer Kompetenz herausfordernd sein kann (Kaiser

et al. 2017; Ufer und Leutner 2017). Bass und Ball (2004) identifizieren in ihrer „job analysis“, einer Tätigkeitsanalyse von Lehrkräften im Mathematikunterricht der Primarschule, Kerntätigkeiten, die auch diagnostische Tätigkeiten enthalten (z. B. Lernprozesse evaluieren). In ihrem Modell zum Professionswissen von Mathematiklehrkräften beschreiben sie verschiedene fachwissenschaftliche und fachdidaktische Wissensfacetten, die für diese Kerntätigkeiten relevant sind (Ball et al. 2008). Im COACTIV-Modell zur professionellen Kompetenz von Lehrkräften werden diagnostische Fähigkeiten explizit mit fachdidaktischem Wissen in Verbindung gebracht (Wissen über das mathematische Denken von Schülerinnen und Schülern sowie Wissen über mathematische Aufgaben). Daneben wird ein Bereich pädagogischen Wissens, das Wissen um Leistungsbeurteilung, als weitere Wissensfacette genannt. Bei einer Diagnose (z. B. bei der Beurteilung von lern- und leistungsrelevanten Merkmalen von Schülerinnen und Schülern) müssen diese verschiedenen Wissensfacetten integriert werden (Brunner et al. 2011). Sie dienen damit als Voraussetzung für diagnostische Tätigkeiten. Insbesondere bei der Operationalisierung fachdidaktischer bzw. diagnostischer Kompetenz wird in den genannten Ansätzen deutlich, dass diagnostische Kompetenz fachbezogenes Wissen erfordert, etwa bei der Analyse typischer Fehler von Schülerinnen und Schülern oder bei der Einschätzung des diagnostischen Potenzials von Aufgaben. Beide Modelle betonen damit die Fachspezifität diagnostischer Kompetenz. Es handelt sich um die Wissensbasis, die zur Bewältigung (fachspezifischer) diagnostischer Anforderungen erforderlich ist.

Für die Förderung der Fähigkeit des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern bedeutet das, dass das benötigte fachbezogene Wissen im Rahmen der Intervention auf- bzw. ausgebaut werden muss, damit es in diagnostischen Situationen aktiviert werden kann. Diese Wissensbasis ist erforderlich, um Überlegungen und Lösungsansätze von Schülerinnen und Schülern deuten zu können. Dabei müssen Informationen aus Aussagen oder Lösungen von Schülerinnen und Schülern mit fachbezogenem Wissen verknüpft werden. Welche Informationen jedoch in einem diagnostischen Prozess herangezogen werden und wie sie verarbeitet werden, ist noch weitgehend unklar (Leuders et al. 2020). Es wird angenommen, dass im Diagnoseprozess verschiedene kognitive Prozesse ablaufen. Informationen müssen wahrgenommen und interpretiert werden, um Entscheidungen treffen zu können (z. B. über die Relevanz von Informationen und deren Gewichtung). Diese Prozesse sind abhängig von in der Person liegenden Faktoren (z. B. Wissen und Einstellungen) und zeigen sich in beobachtbarem Verhalten (Blömeke et al. 2015; Leuders et al. 2020). Im Hinblick auf die Qualität der diagnostischen Urteile ist davon auszugehen, dass eine „hohe Urteilsqualität nur möglich ist, wenn adäquate Verarbeitungsprozesse stattfinden, die auf qualitativ hochwertigen diagnostischen Informationen basieren“ (van Ophuysen und Behrmann 2015, S. 89).

2.2 Diagnose als Informationsverarbeitungsprozess

Die Sicht auf Diagnose als Prozess steht in der Tradition der psychologischen Urteilsforschung (Bless et al. 2004; Hastie und Dawes 2001). Im Kern stellen sich dabei Fragen nach der Art der verarbeiteten Informationen, der Informationsbeschaffung und der Verarbeitung von Informationen bei der Bildung eines Urteils. Informa-

tionsverarbeitungsprozesse stehen zwischen der Aufnahme von Informationen und einem Urteil. Eine wesentliche Grundannahme besteht darin, dass Urteile durch die Interaktion zwischen neu aufgenommenen Informationen und bereits vorhandenem Wissen gebildet werden (Plessner 2011). Das bedeutet, dass beispielweise Informationen über Schülerinnen und Schülern aus Arbeitsprodukten oder aus einem Gespräch mit bereits vorhandenem fachdidaktischem Wissen verknüpft werden, um zu einer Diagnose zu gelangen.

Die Heuristik in einem solchen Prozess kann als Verankerung und Anpassung gedeutet werden (Tversky und Kahneman 1974). Dabei werden Informationen als Anker für eine erste Annäherung zur Lösung eines Problems genutzt. Weitere Informationen dienen dazu, diese erste Annäherung anzupassen. Nickerson (1999) beschreibt einen solchen Prozess bei der Einschätzung des Wissens eines Laien durch eine Expertin oder einen Experten, bei dem der Ausgangspunkt das eigene Wissen darstellt. Unter Berücksichtigung unterschiedlicher Informationen wird das Modell immer weiter modifiziert: Zunächst führen „unübliche Aspekte“ des eigenen Wissens zu einer Adaption des Modells, die Expertin oder der Experte kann die eigene Expertise berücksichtigen. Im nächsten Schritt wird Wissen über die Zugehörigkeit zu einer Gruppe der anderen Person bzw. bereits vorhandenes Wissen über die andere Person berücksichtigt. Mit der Verarbeitung von neu hinzukommenden Informationen (z. B. in einem Gespräch) wird die Einschätzung des Wissens der anderen Person weiter verfeinert (Nickerson 1999). Richtet man den Blick auf die Art der verschiedenen Informationen, die im Modell von Nickerson (1999) beschrieben werden, so wird deutlich, dass sie verschiedenen Ursprungs sind. Einerseits wird auf Informationen zurückgegriffen, über die die urteilende Person bereits verfügt, und andererseits werden Informationen genutzt, die im Prozess neu hinzukommen. Im Prozess der Bildung eines Urteils werden also Informationen aus unterschiedlichen Quellen integriert. Eine Diagnose hängt damit wesentlich vom Wissen der urteilenden Person ab, aber auch von der diagnostischen Situation, die die Art der Informationen erheblich mitbestimmt. Folgt man dem Modell von Nickerson (1999), so nimmt insbesondere das eigene Wissen einen hohen Stellenwert ein. Es dient als Anker und damit als Voraussetzung für den Diagnoseprozess. Damit bildet es die Grundlage dafür, neue Informationen zu interpretieren (Tversky und Kahneman 1974). Bezogen auf fachbezogene diagnostische Prozesse ist daher davon auszugehen, dass das eigene fachbezogene Wissen eine bedeutende Rolle spielt. Allerdings geht es dabei nicht nur um Inhaltswissen, sondern insbesondere um Wissen über mathematisches Denken und Lernen von Lernenden. So konnte Ostermann (2018) zeigen, dass fachdidaktisches Wissen wie etwa typische (Fehl-)Vorstellungen von Schülerinnen und Schülern bei der Schwierigkeitseinschätzung von Aufgaben eine bedeutende Rolle spielt, aber auch fachliches Wissen, das für Lehrkräfte spezifisch ist (z. B. Grundvorstellungen). Während das eigene Wissen im Nickerson-Modell damit dem mathematischen Fachwissen entspricht, wird das spezifische fachliche Wissen als „unübliche Aspekte“ des eigenen Wissens verstanden. Wissen über die Wechselwirkung von Inhalt und Lernendem (wie z. B. typische Fehler von Schülerinnen und Schülern) wird genutzt, wenn Wissen über die Zugehörigkeit einer Gruppe (z. B. Klassenstufe) berücksichtigt wird (Ostermann 2018). Das Nutzen fachbezogenen Wissens bei der Genese eines diagnostischen Urteils konnte, ebenfalls für

das Fach Mathematik, im Rahmen einer qualitativen Studie aufgezeigt werden. Die Funktion des eigenen (fachlichen) Wissens als Grundlage (Anker) für eine Diagnose zeigte sich darin, dass die befragten Lehrkräfte eigene Lösungsansätze nutzten, um Herausforderungen einer Aufgabe zu identifizieren. Außerdem griffen die Lehrkräfte beim Analysieren von Lösungen von Schülerinnen und Schülern auf verschiedene Aspekte fachbezogenen Wissens zurück, wie etwa Grundvorstellungen, typische Fehler oder Fehlkonzepte sowie mathematische Denkweisen und Strategien von Schülerinnen und Schülern (Philipp 2018).

Grundsätzlich kann Diagnose also als Prozess betrachtet werden, bei dem Informationen erfasst und verarbeitet werden, um zu einer Einschätzung eines Merkmals oder zu einer Entscheidung zu gelangen (van Ophuysen und Behrmann 2015). Die Informationsbasis umfasst Informationen aus diagnostischen Situationen und aus eigenem (fachbezogenen) Wissen. Allerdings ist bislang noch wenig darüber bekannt, wie genau diagnostische Prozesse und solche Wissensressourcen zusammenwirken. Das Modell von Nickerson (1999) liefert einen allgemeinen Ansatz, das Zusammenspiel von diagnostischem Prozess und der Nutzung von Wissensarten zu beleuchten. Welche Informationen beim Erkennen von Stärken und Schwächen von Schülerinnen und Schülern genutzt werden, ist Gegenstand der Untersuchung.

2.3 Erwerb diagnostischer Kompetenz

Aufgrund der hohen Bedeutung für die Unterrichtsqualität, die diagnostischer Kompetenz zugeschrieben wird, erscheint deren Förderung und Verbesserung zentral für das Lernen von Schülerinnen und Schülern (z. B. Schrader 2013). Es liegt die Annahme zugrunde, dass sich eine höhere diagnostische Kompetenz positiv auf die Unterrichtsqualität und somit auch auf Leistungen von Schülerinnen und Schülern auswirkt, indem beispielweise Schwierigkeiten von Schülerinnen und Schülern frühzeitig erkannt werden (Anders et al. 2010; Helmke et al. 2004; Helmke 2009).

Diagnostische Kompetenz gilt als prinzipiell erlernbar, auch wenn die genauen Bedingungen des Erwerbs diagnostischer Kompetenz noch unklar sind (Herppich et al. 2017; Krauss et al. 2008). In diesem Zusammenhang stellt sich die Frage, wie diagnostische Kompetenz gefördert bzw. überhaupt erworben werden kann. Betrachtet man Konzepte zur Förderung diagnostischer Kompetenz, so lassen sich zahlreiche Interventionsstudien finden, die die Entwicklung diagnostischer Kompetenz durch den Aufbau spezifischen Wissens für diagnostische Situationen anregen, beispielweise in themenspezifischen Fortbildungen für Lehrkräfte (z. B. Busch et al. 2015). Ansätze zur Förderung diagnostischer Prozesse nehmen die gezielte Sammlung und theoriegeleitete Verarbeitung von diagnostischen Informationen in den Blick, beispielweise durch das Führen diagnostischer Interviews (Clarke et al. 2018; Wollring et al. 2013). Gemeinsam ist diesen Ansätzen die Annahme, dass sich diagnostische Kompetenz durch die Verknüpfung von theoretischem Wissen einerseits und reflektierter Praxiserfahrung andererseits erwerben lässt (Hascher 2008). Damit sind beispielsweise das Arbeiten an Beispielen oder tatsächliche Praxiserfahrungen im Rahmen der Ausbildung von Lehrkräften gemeint, die systematisch reflektiert werden. Solche praxisnahen Erfahrungen als „approximations of practice“ bilden insbesondere in der Ausbildung von Lehrkräften häufig Lernanlässe (Grossman und

McDonald 2008, S. 190). Chernikova et al. (2020) stützen diese Sichtweise in Bezug auf diagnostische Kompetenz: Das fachbezogene Wissen von Lehrkräften spielt eine Rolle bei der Förderung diagnostischer Kompetenz, es wird gesehen als „prerequisite for further development of skills and competences“ (Chernikova et al. 2020, S. 160), d.h. es beeinflusst die Entwicklung diagnostischer Kompetenz. Dabei sollen Möglichkeiten zu praktischen Erfahrungen mit diesem Wissen verknüpft werden: „practice opportunities, combined with sufficient professional knowledge, can facilitate diagnostic competences in higher education“ (Chernikova et al. 2020, S. 161). Weiterhin stellen die Autorinnen und Autoren förderliche Aspekte im Rahmen von Scaffolding heraus wie das angeleitete Arbeiten an Beispielen, Bereitstellung von „Prompts“ (Lernhinweisen), die Übernahme einer bestimmten Rolle sowie die Bedeutung von Reflexionsphasen. Daraus lässt sich für die vorliegende Studie ableiten, dass authentische diagnostische Situationen und Beispiele Praxiserfahrungen ermöglichen, die angeleitet und mit theoretischem Wissen verknüpft sowie reflektiert werden müssen, um den Aufbau diagnostischer Kompetenz zu fördern. Insbesondere kann die Förderung diagnostischer Kompetenz besonders wirksam sein, wenn interaktionsbasierte diagnostische Situationen einbezogen werden (Chernikova et al. 2020).

Mit Blick auf Informationsverarbeitungsprozesse lassen sich je nach Expertisegrad Unterschiede finden, die weitere Hinweise auf den Erwerb diagnostischer Kompetenz geben können. Kellman und Massey (2013) stellen zusammenfassend fest, dass Novizinnen und Novizen eher alle verfügbaren Informationen beachten, während Expertinnen und Experten relevante Informationen auswählen können und nicht relevante Informationen ausblenden. Außerdem verarbeiten Novizinnen und Novizen eher einzelne Aspekte nacheinander und können noch nicht größere Zusammenhänge oder Strukturen bilden (Kellman und Massey 2013). Begründen lassen sich solche Unterschiede auch durch das unterschiedliche Wissen, wobei es nicht nur um „mehr“ Wissen geht, sondern auch darum, dass das Wissen besser organisiert werden kann (z.B. Bromme 2008). In Bezug auf Mathematik wird berichtet, dass angehende Lehrkräfte Schwierigkeiten damit haben, Lösungen von Schülerinnen und Schülern zu interpretieren, wobei ihr eigenes mathematisches Wissen eine bedeutende Rolle spielt (Stahnke et al. 2016). Für die Förderung des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern könnte das bedeuten, dass die Auswahl von Informationen anhand fachbezogener Kriterien gezielt unterstützt werden muss.

Bei der Entwicklung von diagnostischer Kompetenz ist von einem längerfristigen Lernprozess auszugehen, der eine kontinuierliche Auseinandersetzung mit diagnostischen Fragen erfordert. In der Ausbildung von Lehrkräften kann der Erwerb diagnostischer Kompetenz angebahnt werden, im Rahmen der Berufstätigkeit wird sie weiter ausgebaut (Buholzer und Zulliger 2013; Chernikova et al. 2020; Hascher 2008). Wichtig für die Kompetenzentwicklung ist, dass die Vielzahl an diagnostischen Situationen, mit denen Lehrkräfte im Alltag konfrontiert sind, und ihre Bewältigung nicht unreflektiert bleibt. Die Förderung diagnostischer Kompetenz ist somit in allen Phasen der Ausbildung, aber auch in der Fortbildung von Lehrkräften möglich.

3 Ziele der Studie und Forschungsfragen

Die hier berichtete Interventionsstudie zielt auf die Förderung diagnostischer Kompetenz im Rahmen der Ausbildung von Lehrkräften für die Primarschule. Im Fokus steht eine zentrale Fähigkeit (als Teilfähigkeit diagnostischer Kompetenz) bei der Beurteilung mathematischer Kompetenzen von Schülerinnen und Schülern, das Erkennen von Stärken und Schwächen. Die Entwicklung diagnostischer Kompetenz soll einerseits durch die Vermittlung fachbezogenen Wissens zum Inhaltsbereich Größen und andererseits durch die Auseinandersetzung mit unterschiedlichen diagnostischen Situationen (Aufgabenanalyse, Analyse von Arbeitsprodukten von Schülerinnen und Schülern, Analyse von Videos und Führen eines diagnostischen Interviews) und deren systematische Reflexion angeregt werden. Damit soll auch der Breite der verschiedenen Quellen für diagnostische Informationen Rechnung getragen werden (Chernikova et al. 2020).

Zunächst soll in einem quasi-experimentellen Setting geklärt werden, ob die Förderung der genannten Teilfähigkeit diagnostischer Kompetenz gelingt. Es wird davon ausgegangen, dass die Intervention die Entwicklung der Fähigkeit des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern positiv beeinflusst. Nachdem sich in bisherigen Befunden meist kurzfristige Wirkungen aufzeigen ließen (Karing und Seidel 2017), ist auch die Frage nach der Nachhaltigkeit einer solchen Förderung von Relevanz:

1. Lässt sich die Fähigkeit des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern durch die Vermittlung fachbezogenen Wissens und durch die Konfrontation mit verschiedenen diagnostischen Situationen prinzipiell und nachhaltig fördern?

Da das vermittelte Wissen beim Erkennen von Stärken und Schwächen von Schülerinnen und Schülern in einem Informationsverarbeitungsprozess mit Informationen aus einer diagnostischen Situation verknüpft werden muss und es Hinweise darauf gibt, dass angehende Lehrkräfte damit Schwierigkeiten haben, soll explorativ untersucht werden, welche Informationen die Studierenden nutzen:

2. Auf welche Informationen beziehen sich die Studierenden bei der schriftlichen Analyse eines selbst durchgeführten diagnostischen Interviews?
3. Welche Informationen werden bei der Formulierung und Begründung von Stärken und Schwächen von Schülerinnen und Schülern miteinander verknüpft?

4 Förderung und Untersuchung diagnostischer Kompetenz

4.1 Design

In einer quasi-experimentellen Studie wurde eine Intervention im Umfang von fünf Seminarsitzungen à 90 min im Rahmen regulärer Lehrveranstaltungen durchgeführt, die die Förderung fachbezogenen Wissens als Basis für das Erkennen von Stärken

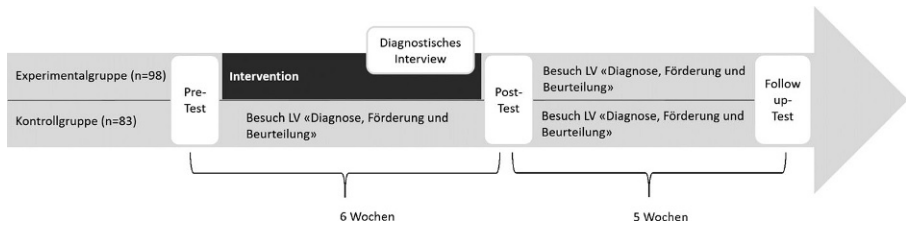


Abb. 2 Design der Studie

und Schwächen von Schülerinnen und Schülern (als Teilfähigkeit diagnostischer Kompetenz) im Bereich Größen beinhaltete (vgl. Abb. 2). Eine Kontrollgruppe nahm in der gleichen Zeit an einer Lehrveranstaltung zum Thema „Diagnose, Förderung und Beurteilung“ ohne inhaltlichen Bezug zum Thema Größen teil. Die Gruppen unterschieden sich damit hinsichtlich der Spezifität des thematisierten fachdidaktischen Wissens, gemeinsam ist ihnen die Auseinandersetzung mit diagnostischen Fragestellungen. Ein weiteres Unterscheidungsmerkmal der beiden Gruppen war das Führen eines diagnostischen Interviews mit einer Schülerin oder einem Schüler der Primarstufe als Bestandteil der Intervention. Das Interview wurde von den Studierenden vor dem zweiten Messzeitpunkt geführt und ausgewertet.

Zur Erfassung diagnostischer Kompetenz wurde ein Testinstrument entwickelt, das zu drei verschiedenen Messzeitpunkten eingesetzt wurde. Neben der Wirksamkeit der Intervention (Forschungsfrage 1) wurde die Verwendung von Informationen der Studierenden der Experimentalgruppe untersucht (Forschungsfragen 2 und 3), indem schriftliche Dokumente der Studierenden (Analyse des selbst durchgeführten diagnostischen Interviews) inhaltsanalytisch ausgewertet wurden.

4.2 Stichprobe

Die Stichprobe ($N = 181$) bildeten Studierende des Studiengangs Bachelor Primarstufe an der Pädagogischen Hochschule Nordwestschweiz. Davon waren 142 weiblich und 39 männlich. Das durchschnittliche Alter der Studierenden betrug 24,3 Jahre ($SD = 4,40$). Die Lehrveranstaltungen konnten von Studierenden ab dem 3. Semester belegt werden, die bereits die Einführungsveranstaltung in die Mathematikdidaktik erfolgreich abgeschlossen hatten. Alle Studierenden stimmten einer Teilnahme an einem Forschungsprojekt im Rahmen der regulären Lehrveranstaltung zu (ohne Kenntnis über die Gruppenzuteilung), indem sie sich für eine der angebotenen Lehrveranstaltungen anmeldeten. An der Intervention nahmen insgesamt 98 Studierende teil, in der Kontrollgruppe waren es 83 Studierende. Die Lehrveranstaltungen waren Wahl-Pflichtveranstaltungen für die Studierenden. Bei den Lehrveranstaltungen handelte es sich um Seminare mit maximal 30 zugelassenen Studierenden. Vier Lehrveranstaltungen waren der Experimentalgruppe zugeordnet, die von derselben Dozentin durchgeführt wurden, während die Kontrollgruppe auf vier weitere Seminare mit insgesamt drei verschiedenen Dozierenden verteilt war.

4.3 Intervention

Aufgrund der Annahme, dass das eigene fachbezogene Wissen in diagnostischen Situationen eine bedeutende Rolle spielt und für das Erkennen von Stärken und Schwächen von Schülerinnen und Schülern grundlegende Voraussetzung ist, war fachdidaktisches Wissen wesentliches Element der Intervention. Als Inhaltsbereich wurden Größen gewählt, weil sie einen unmittelbaren Bezug zur Lebenswelt haben und als „Verbindung von arithmetischen und geometrischen Inhalten und Kompetenzen“ (Peter-Koop und Nührenböcker 2008, S. 89) eine bedeutende Rolle im Mathematikunterricht spielen. Außerdem wurde in diesem Bereich das fachdidaktische Vorwissen der Studierenden als geringer eingeschätzt als in Bezug auf arithmetische Inhalte, die im ersten Studienjahr sowohl in der Fachdidaktik als auch in der Fachwissenschaft im Zentrum stehen, sodass (unerwünschte) Wechselwirkungen hätten entstehen können. Hinzu kommt, dass zahlreiche Studierende in der deutschsprachigen Schweiz bereits ab Studienbeginn als Lehrkräfte tätig sind und vorwiegend im Bereich Arithmetik Praxiserfahrungen sammeln, die zu Unterschieden im Vorwissen hätten führen können. Die Studierenden setzten sich im Rahmen der Intervention mit allen in der Primarstufe in der Schweiz vorkommenden Größenbereichen (Längen, Zeit, Geld, Gewichte, Hohlmaße und Flächeninhalte) auseinander. Es ging um den Aufbau von Größenvorstellungen, die Behandlung von Größen im Unterricht, Herausforderungen der verschiedenen Größenbereiche sowie typische Schwierigkeiten und Vorstellungen von Schülerinnen und Schülern. Durch die Analyse von Aussagen und Lösungen von Schülerinnen und Schülern und durch die originale Begegnung mit einer Schülerin oder einem Schüler der Primarstufe in einem diagnostischen Interview wurde die Nutzung fachbezogenen Wissens in diagnostischen Situationen angeregt.

Leitende Gestaltungsprinzipien der Intervention waren die Auseinandersetzung mit Aufgabenanforderungen, dem diagnostischen Potenzial von Aufgabenstellungen und Aussagen und Lösungen von Schülerinnen und Schülern in Form von Videos oder schriftlichen Dokumenten, die ein mehrfaches Betrachten und eine vertiefte Analyse erlaubten. Ergänzend wurde das geführte (und videografierte) diagnostische Interview von den Studierenden analysiert. Dazu wurde der Größenbereich Gewichte¹ ausgewählt, da dieser insbesondere beim Aufbau von Größenvorstellungen große Herausforderungen für Schülerinnen und Schüler mit sich bringt: So sind Gewichte nur in einem kleinen Bereich erfahrbar, da sehr kleine Gewichte kaum noch unterscheidbar sind und Gegenstände für Kinder recht schnell zu schwer werden, was den Aufbau von Größenvorstellungen erschwert. Auch kann das unterschiedliche Druckgefühl von Gegenständen bei der Wahrnehmung von Gewichten per Hand zu falschen Einschätzungen führen. Hinzu kommt, dass Gewichte nicht visuell wahrnehmbar sind, d. h. die Größe des Gegenstands lässt keinen Rückschluss auf sein Gewicht zu (Bräunling und Reuter 2015; Franke und Ruwisch 2010; Reuter 2011). Ziel des diagnostischen Interviews für die Studierenden war es, Stärken

¹ Umgangssprachlich und auch im Mathematikunterricht werden „Gewicht“ und „Masse“ gleichbedeutend verwendet. Auch wenn das physikalisch nicht korrekt ist, wird der Begriff „Gewicht“ auch in Anlehnung an fachdidaktische Literatur beibehalten.

und Schwächen des Kindes in Bezug auf Größenvorstellungen im Bereich Gewichte zu benennen und mit Bezug auf Aussagen oder Lösungen des Kindes zu begründen. Dazu standen den Studierenden acht Aufgaben mit diagnostischem Potenzial zur Verfügung, die sie im Interview einsetzten. Dieses dauerte ca. 45 min und wurde nicht im Rahmen der fünf Seminarsitzungen durchgeführt und ausgewertet. Eine schriftliche Bearbeitung (eigene Lösung) und Analyse des diagnostischen Potenzials der Aufgaben im Vorfeld diente der vertieften Auseinandersetzung mit fachbezogenen Kriterien zu Größenvorstellungen bei Gewichten. Die schriftliche Dokumentation der Studierenden bestand aus insgesamt vier Teilen: einem (vorbereitenden) Interviewleitfaden, einer Analyse der Aussagen und Lösungen des Kindes (aufgabenweise), einer Rückmeldung an die unterrichtende Lehrkraft und einer Reflexion des eigenen Verhaltens im Interview.

4.4 Instrumente

Die eingesetzten Instrumente unterscheiden sich hinsichtlich der diagnostischen Situation. Das eingesetzte Testinstrument zur Überprüfung der Wirksamkeit der Intervention enthielt schriftliche Lösungen von verschiedenen Schülerinnen und Schülern, die die Studierenden einschätzen sollten. Dies ermöglichte einen direkten Vergleich der Analysen der Studierenden pro Aufgabe hinsichtlich der genannten Stärken und Schwächen der Schülerinnen und Schüler und deren Interpretationen. Die Analyse der diagnostischen Interviews ermöglichte dagegen eine umfassende Beurteilung der Kompetenzen eines einzelnen Kindes über mehrere Aufgaben hinweg. Der Fokus der Auswertung lag hier auf den Informationen, die bei der Analyse eines Gesprächs mit einer „echten“ Schülerin oder einem Schüler genutzt wurden. Bei beiden Instrumenten war das Diagnoseziel für die Studierenden dasselbe: Die Stärken und Schwächen von Schülerinnen und Schülern in Bezug auf Größenvorstellungen im Bereich Gewichte sollten benannt und aufgrund der jeweiligen Datenlage begründet werden.

4.4.1 Testinstrument

Zur Überprüfung der Wirksamkeit der Intervention (Forschungsfrage 1) wurde ein Testinstrument entwickelt. Ziel war es, Veränderungen der Fähigkeit, Stärken und Schwächen von Schülerinnen und Schülern zu erkennen, zu erfassen. Die Studierenden sollten anhand schriftlich vorliegender Aufgabenlösungen von Schülerinnen und Schülern der Primarstufe (4. Klasse) zu zehn Aufgaben zu Größenvorstellungen von Gewichten (1) *richtige* und *fehlerhafte* Überlegungen *identifizieren* sowie (2) die Lösungen der Schülerinnen und Schüler *interpretieren* (z.B. mögliche Ursachen nennen). Die Aufgaben waren so konzipiert, dass Teilkompetenzen für Größenvorstellungen umfassend abgedeckt waren (Franke und Ruwisch 2010; Reuter 2011; Schipper 2009): Standardeinheiten kennen; Größen vergleichen, messen und schätzen; Repräsentanten für Standardeinheiten kennen; Größenangaben umwandeln und deren Invarianz erkennen. Als Antwortformat wurden offene Items gewählt. Bei der Auswahl der Lösungen von Schülerinnen und Schülern für das Testinstrument wur-

de darauf geachtet, dass der Interpretationsspielraum groß war, um möglichst viele (plausible) Vermutungen zu Fehlerursachen zuzulassen.

4.4.2 Schriftliche Dokumente der Studierenden (*Analysen diagnostischer Interviews*)

Im Rahmen der Intervention führten alle Studierenden der Experimentalgruppe ein diagnostisches Interview mit einer Schülerin oder einem Schüler der Primarstufe der Klassenstufen 3 bis 5 durch. Dazu erhielten sie ein Set von acht Aufgaben mit diagnostischem Potenzial (vgl. Emmrich 2008; Prediger et al. 2017), d.h. bei der Bearbeitung der Aufgaben sollten sowohl Vorstellungen als auch Lösungswege der Lernenden sichtbar werden, die Rückschlüsse auf deren Größenvorstellungen zu Gewichten zuließen. Die Auswertung der diagnostischen Interviews durch die Studierenden erfolgte schriftlich und beinhaltete mehrere Teile (vgl. Abschn. 4.3). Die schriftlichen Dokumente der Studierenden dienten als Datengrundlage zur Untersuchung der Frage nach der Verwendung von Informationen bei der Diagnose (Forschungsfragen 2 und 3). Dazu wurde ausschließlich ein Teil der schriftlichen Dokumente, die Analyse der Aussagen und Lösungen des Kindes, herangezogen. Die Verwendung von Informationen konnte so differenziert betrachtet werden: auf welche Informationen die Studierenden bei der Analyse zurückgriffen und welche sie bei der Formulierung und Begründung von Stärken und Schwächen nutzten.

4.5 Auswertung

4.5.1 Wirksamkeit der Intervention

Die Fähigkeit des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern wurde bei der Auswertung den zwei Bereichen *Stärken/Schwächen identifizieren* (Wert I) und *Stärken/Schwächen interpretieren* (Wert II) zugeordnet: Stärken und Schwächen zu benennen wird als erster wichtiger Schritt im Diagnoseprozess aufgefasst (identifizieren). Interpretieren zu können, auf welche Teilkompetenzen von Größenvorstellungen (vgl. Abschn. 4.4.1) die Lösung der Schülerinnen und Schüler Rückschlüsse zulässt, setzt voraus, dass tieferliegende Strukturen in ihnen erkannt werden. Die Anzahl der schlüssigen Antworten der Studierenden wurden für die Auswertung herangezogen. Dahinter steckt die Überlegung, dass auf der Basis einzelner Arbeitsprodukte von Schülerinnen und Schülern keine eindeutige Diagnose gestellt werden kann. Insofern war für die Qualität der Diagnose maßgebend, dass die Studierenden möglichst viele Stärken und Schwächen identifizieren konnten und dazu auch mögliche Ursachen (hypothetisch) breit benennen konnten. Jede korrekt als fehlerhaft oder richtig identifizierte Überlegung bzw. Lösung wurde mit einem Punkt bewertet. Ebenso wurde jede schlüssige interpretierende Aussage mit einem Punkt bewertet. Sinngemäß gleiche Aussagen wurden nur einfach gewertet. Anschließend wurden die Punkte summiert. Eine Maximalpunktzahl pro Aufgabe war nicht festgelegt.

In Abb. 3 ist ein Beispielitem und eine Studierendenantwort sowie deren Auswertung dargestellt: Die oder der Studierende identifiziert die Gewichtsangabe für

Abb. 3 Beispielitem und Auswertung der Studierendenantwort (Hervorhebungen nachträglich hinzugefügt)

2) Was schätzt du, wie viel wiegen diese Dinge in Wirklichkeit? Begründe Deine Antwort.

Gewicht einer Kuh:

Gewicht eines Stuhls:

Gewicht einer Butter:

Begründe Deine Antwort:

→ Folgende Schüler*in - Antwort wurde gegeben:

Gewicht einer Kuh: 200t
 Gewicht eines Stuhls: 2,15kg
 Gewicht einer Butter: 250gr.
 Begründe Deine Antwort:
 Ich backe viel mit meinem Mami. Daher weiss ich, das eine Butter 250gr wiegt. Unser Stuhl Zuhause wiegt ungefähr 20mal so viel!

(INFO: Eine Kuh wiegt zwischen 400-700kg.)

	Welche richtigen Überlegungen erkennen Sie?	Welche fehlerhaften Überlegungen erkennen Sie?
Wert I	• das Gewicht der Butter	• das Gewicht der Kuh wird deutlich überschätzt
Wert II	Wie interpretieren Sie die Lösungen des Kindes? Welche Ursachen können für die Lösungen / den Lösungsweg eine Rolle gespielt haben? • obwohl das Kind die Kuh überschätzt, ist es in der Lage → Vergleiche zu machen (die Butter mit dem Stuhl) • Außerdem ist es ein aufgewecktes Kind, denn es kann sich das Gewicht der Butter merken, da es ja oft mit der Butter backt.	

die Butter als richtig und die für die Kuh als „deutlich überschätzt“. Diese *Identifikationen* wurden mit 2 Punkten bewertet (Wert I). Des Weiteren gibt die oder der Studierende an, dass das Kind in der Lage ist, Vergleiche anzustellen, da es den Stuhl mit der Butter vergleicht. Diese *Interpretation* ergab 1 Punkt (Wert II). Dass das Kind aufgeweckt sei, wurde nicht bewertet, da dies eine Aussage über einen Charakterzug des Kindes ist. Ebenso wurde die Wiederholung der Antwort der Schülerin oder des Schülers nicht gewertet, da sie keine Interpretation enthält. Insgesamt wurden hier also 3 Punkte vergeben.

Die Güte des Testinstruments wurde mittels einer Itemanalyse überprüft. Dazu wurden die Werte des Posttests verwendet, um die größtmöglichen Unterschiede in den Antworten der Studierenden als Auswirkung der Intervention zu erfassen. In Tab. 1 sind Spannweite, Mittelwert und die Standardabweichung jedes Items abgebildet sowie Itemschwierigkeit und Trennschärfe. Die Trennschärfe wurde für jedes Item der Skala berechnet. Weil Item 8 einen sehr tiefen Wert ($r_{ii}=0,311$) aufwies, wurde es (nach Prüfung der inhaltlichen Validität der verbliebenen Items) eliminiert. Die Trennschärfen der restlichen neun Items lagen zwischen $r_{ii}=0,343$ und $r_{ii}=0,590$, was als „mittelmäßig“ eingeordnet werden kann (Döring und Bortz 2016). Für die Bestimmung der Itemschwierigkeit wurde der Maximalwert als Durchschnitt

Tab. 1 Überprüfung der Güte der Items

	<i>N</i>	Mini- mum	Maxi- mum	Mittel- wert	Std. Abwei- chung	Itemschwie- rigkeit (%)	Trenn- schärfe
Item 1	181	0	15	6,34	2,942	52	0,475
Item 2	181	1	18	7,70	3,064	54	0,522
Item 3	181	0	17	5,86	3,229	50	0,379
Item 4	181	1	16	6,60	2,586	56	0,554
Item 5	181	0	13	5,14	2,703	50	0,590
Item 6	181	0	12	4,42	1,773	57	0,523
Item 7	181	0	8	3,59	1,570	54	0,346
Item 8	181	1	10	4,44	1,967	53	0,311
Item 9	181	1	8	4,07	1,362	61	0,343
Item 10	181	0	6	3,66	1,312	62	0,415

der Summe der erreichten Punkte der besten 10% der Studierenden pro Item berechnet. Die Itemschwierigkeit aller Items lag zwischen 50 und 62% und somit in einem gewünschten Bereich (Döring und Bortz 2016). Mit dem Ausschluss von Item 8 lag das Reliabilitätsmaß der Skala, der Alpha-Koeffizient von Cronbach, bei 0,764.

Die Unterschiede in der Standardabweichung der verschiedenen Items lassen sich dadurch erklären, dass bei den Aufgaben die Anzahl der oben genannten Teilkompetenzen, zu denen eine Aussage gemacht werden konnte, variierte.

4.5.2 Verwendung von Informationen

Um Forschungsfrage 2 zu beantworten, auf welche Informationen sich die Studierenden in ihrer Analyse beziehen, wurden Kategorien entwickelt. Diese geben Aufschluss über die Informationsbasis, auf deren Grundlage eine Diagnose zustande kommt. Die Kategorien sollten die ganze Breite an unterschiedlichen Informationen widerspiegeln. Das Zusammenspiel verschiedener Kategorien sollte nachfolgend anhand der deduktiv und induktiv gebildeten Kategorien analysiert werden (Forschungsfrage 3). Als methodischer Zugang wurde eine inhaltlich-strukturierende Inhaltsanalyse (Kuckartz 2016; Mayring 2010) gewählt. Ziel dieser Vorgehensweise ist es, zentrale Aspekte zu identifizieren und das Datenmaterial anhand dieser Aspekte systematisch zu beschreiben. Das Verfahren ist gekennzeichnet durch interpretatives Vorgehen und das Merkmal der Kategoriengeleitetheit (Mayring und Fenzl 2019). Das Vorgehen im Rahmen der berichteten Studie orientiert sich am Ablaufschema von Kuckartz (2016), da es der induktiven Entwicklung von Kategorien ein größeres Gewicht beimisst, während das Modell von Mayring (2010) stärker auf die deduktive Entwicklung von Kategorien ausgerichtet ist. Das Vorgehen wurde modifiziert und an die Bedürfnisse der Studie angepasst:

1. Die Entwicklung der Hauptkategorien erfolgte deduktiv, d. h. die Kategorien wurden auf der Basis theoretischer Vorüberlegungen gebildet. Das beinhaltete die Unterscheidung von Informationen, die die Studierenden zur Verfügung hatten (*ver-*

Tab. 2 Kategoriensystem zur Analyse der Informationsnutzung bei der Diagnose

Hauptkategorie	Subkategorien	Beschreibung	Ankerbeispiel
V: Verfügbare Information	V1: Indikatoren für Größenvorstellungen	Ein fachbezogenes Kriterium wird genannt wie Messinstrumente kennen, passende Maßeinheiten finden, passende Repräsentanten finden, Gewichte vergleichen und ordnen, über Stützpunktvorstellungen verfügen, Beziehungen zwischen Maßeinheiten oder Invarianz von Gewichten erkennen	„Wenn sie diese jedoch schätzen muss, fällt es ihr schwer auf <i>Stützpunktvorstellungen</i> zurückzugreifen“ (Dfa_S1, S. 11: 1550)
	V2: Herausforderungen bei Gewichten	Typische Schwierigkeiten werden genannt wie die visuelle Wahrnehmbarkeit von Gewichten oder das Problem geringer Gewichtsunterschiede sowie der begrenzte Erfahrungsbereich bei Gewichten	„Ihr war nicht klar, dass das <i>Gewicht visuell nicht wahrnehmbar</i> ist und somit die Grösse keinen Rückschluss auf das Gewicht zulässt.“ (Dfa_S1, S. 11: 1814)
	V3: Mathematikunterricht	Fachbezogene Aspekte über Mathematikunterricht, Lehrmittel oder Lehrpläne werden genannt	„[...] dass <i>oftmals im Mathematikunterricht das Rechnen mit Grössen fokussiert wird</i> . Folglich können die Schülerinnen und Schüler schon recht früh Einheiten umrechnen, ohne jedoch genauere Vorstellungen zu haben, was die Masseinheiten beinhalten.“ (DId_S11, S. 8: 535)

fügbare Information), und den Informationen, die sie der Interviewsituation mit der Schülerin oder dem Schüler entnahmen (*neu hinzukommende Information*).

- Die schriftlichen Dokumente (Analysen diagnostischer Interviews) der Studierenden ($N=98$) wurden in einer ersten Kodierung anhand der beiden Hauptkategorien grob kodiert.
- Für die weitere Feinanalyse (Bildung von Subkategorien) wurden insgesamt 12 Dokumente aufgrund verschiedener Kriterien ausgewählt. Es wurde darauf geachtet, dass Studierende aus allen vier Lehrveranstaltungen sowie männlichen und weiblichen Geschlechts vertreten sind. Zudem sollten alle acht Aufgaben im Interview eingesetzt worden sein. Schließlich richtete sich die Anzahl der Interviews danach, dass im Verlauf des Kodierprozesses keine neuen Aspekte mehr hinzukamen (theoretische Sättigung).
- Die Subkategorien wurden sowohl deduktiv als auch induktiv (am Material), geleitet durch die Forschungsfrage, gebildet. Dazu gehörte der Bezug zu Informationen, die die Studierenden in ihrer Analyse nutzten, beispielsweise der Bezug zu fachbezogenen Kriterien zu Größenvorstellungen von Gewichten, die Bestandteil der Intervention waren und zur deduktiven Kategorienbildung genutzt wurden (als Subkategorien zur Hauptkategorie *verfügbare Informationen*). In einer zweiten Kodierung wurde das Material anhand der Subkategorien kodiert und das Kategoriensystem fortlaufend angepasst.

Tab. 2 (Fortsetzung)

Hauptkategorie	Subkategorien	Beschreibung	Ankerbeispiel
N: Neu hinzukommende Information	N1: Richtige Lösung	Eine korrekte Antwort oder Lösung wird genannt	„Sie schätzte die 3900 g korrekt für das Gewicht eines Neugeborenen.“ (DIc_S8, S. 8: 2673)
	N2: Fehler	Ein Fehler oder eine Fehlüberlegung des Kindes wird genannt	„Bei dieser Aufgabe liegt er mit seinen angegebenen Gegenständen meistens falsch.“ (DIb_S5, S. 6: 388)
	N3: Beispiel	Ein Beispiel des Kindes wird genannt	„Bei der Frage, woher er die Gewichte kenne, kamen Aussagen, wie: ‚Mein Cousin, der ein Jahr alt ist, wiegt 10kg.‘“ (DIa_S3, S. 6: 895)
	N4: Argument	Eine Begründung des Kindes wird genannt, beispielweise warum ein Gegenstand schwer oder leicht ist	„Er antwortet: ‚Also erstens ist in einem Brötchen viel Luft und zweitens besteht die Münze aus Metall, und Metall ist schwer.‘“ (DIa_S2, S. 7: 2776)
	N5: Strategisches Vorgehen	Das strategische Vorgehen des Kindes wird beschrieben wie das Vergleichen, Abschätzen, Raten oder das Zuordnen von Gewicht und Größe des Gegenstandes	„In Aufgabe zwei vergleicht sie die Gegenstände untereinander und im Vergleich mit Stützpunktvorstellungen und erstellt einer imaginären Rangfolge nach Gewicht.“ (DIb_S4, S. 8: 2111)
	N6: Bearbeitungsverhalten	Aspekte bezüglich der Bearbeitung werden genannt wie das Revidieren der Lösung, Nachfragen des Kindes, Zeigen von (Un-)Sicherheit, Reihenfolge der Bearbeitung und Bearbeitungsdauer	„[...] auffällig, dass F. lange zögert um einzelne Gegenstände für die Standardgrößen zu bestimmen.“ (DIb_S5, S. 6: 110)
	N7: Vorwissen	Aspekte aus dem Vorwissen des Kindes werden genannt wie Alltagserfahrung, Messerfahrung und Erfahrungen aus dem Unterricht	„Der Umgang mit der Küchenwaage kenne er vom Kochen mit seiner Mutter, [...]“ (DIc_S7, S. 6: 1707)

5. Um die Stabilität des Kodierens zu gewährleisten, wurde nach Abschluss der Analyse mit einem Abstand von drei Wochen ein Teil des Materials (25 %) erneut kodiert, um die Intrakoderübereinstimmung ($\kappa_n = 0,92$, nach Brennan und Prediger 1981) sicherzustellen (Rädiker und Kuckartz 2019).
6. Bezüglich der Forschungsfrage 3 wurden Zusammenhänge zwischen Kategorien untersucht. Dazu wurden Muster im Auftreten der Kategorien untersucht. Umgesetzt wurde dieser Auswertungsschritt über die Untersuchung von Beziehungen zwischen den Kategorien anhand der Überschneidungen (gemeinsames Auftreten von Kategorien) in einem Textabschnitt (Rädiker und Kuckartz 2019). Zu diesem Zweck wurde die Formulierung von Stärken und Schwächen von Schülerinnen und Schülern und deren Begründung ergänzend kodiert.

Bei der Entwicklung des Kategoriensystems stand die Frage, auf welche Informationen die Studierenden in ihrer Analyse Bezug nehmen, im Zentrum (vgl. Tab. 2). Die Frage nach der Nutzung von Informationen, etwa bei der Formulierung von Hypothesen zu Stärken und Schwächen des interviewten Kindes, wurde anhand von Analysen der Beziehungen der Kategorien untersucht.

5 Ergebnisse

5.1 Wirksamkeit der Intervention

Mithilfe des eingesetzten Testinstruments konnte die Frage geklärt werden, ob die Intervention bei der Experimentalgruppe zu einer höher ausgeprägten Fähigkeit des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern geführt hat als bei der Kontrollgruppe. Dazu wurden die Daten varianzanalytisch ausgewertet. Die Voraussetzungen dafür waren erfüllt: Die Unabhängigkeit der Messwerte war durch die Stichprobenwahl und das Design gegeben. Die Normalverteilung der abhängigen Variablen – dem Erkennen von Stärken und Schwächen – konnten durch den Saphiro-Wilk-Test ($p > 0,05$), mögliche Ausreißer durch die visuelle Überprüfung des Boxplots und die Sphärizität mit dem Mauchly-Test ($p > 0,05$) überprüft werden. Die Varianzgleichheit der Residuen mit dem Levene-Test war zum zweiten und dritten Messzeitpunkt nicht erfüllt. Varianzheterogenität entsteht jedoch gemäß Bryk und Raudenbush (1988) oft durch die Einwirkung eines Treatments, in diesem Fall der Intervention, was sich in unterschiedlichen Varianzen niederschlagen kann. Die Varianzanalyse gilt bei etwa gleich großen Stichproben als robust und Bortz und Schuster (2010) empfehlen nur bei kleinen Stichproben ($n < 10$) und Verletzung mehrerer Voraussetzungen ein anderes Verfahren.

In Tab. 3 sind die Mittelwerte sowie die Standardabweichung der erreichten Punkte der Experimental- und Kontrollgruppe zu allen drei Messzeitpunkten dargestellt (vgl. auch Abb. 4). Ebenso ist ersichtlich, wie sich die Gruppen zu den Messzeitpunkten unterschieden.

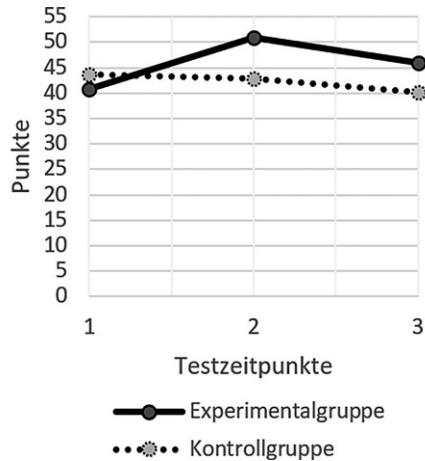
Deskriptiv lässt sich feststellen, dass die Experimentalgruppe beim zweiten Messzeitpunkt den höchsten Wert erzielte ($M = 50,82$; $SD = 14,22$). Dieser nahm zum dritten Messzeitpunkt wieder ab ($M = 45,98$; $SD = 13,58$), lag jedoch höher als zum ersten Messzeitpunkt. Die Kontrollgruppe hatte beim Pretest einen höheren Mittelwert als die Experimentalgruppe ($M = 43,69$; $SD = 9,56$). Der Mittelwert der Kontroll-

Tab. 3 Gruppenunterschiede zu den drei Messzeitpunkten

Messzeitpunkt	Experimentalgruppe ($n = 98$)		Kontrollgruppe ($n = 83$)		ANOVA			
	M	SD	M	SD	Effekt	df	F	η^2
Pre	40,78	8,55	43,69	9,56	Gruppe	1	6,53*	0,035
Post	50,82	14,22	42,83	9,19	Zeit	2	21,96**	0,109
Follow-up	45,98	13,58	40,07	9,98	Gruppe \times Zeit	2	30,50**	0,146

* $p < 0,05$; ** $p < 0,01$

Abb. 4 Mittelwerte der Experimental- und Kontrollgruppe



gruppe nahm dann zum zweiten Messzeitpunkt ($M=42,83$; $SD=9,19$) sowie zum dritten Messzeitpunkt ab ($M=40,07$; $SD=9,98$).

In einem nächsten Schritt wurden Gruppenunterschiede unter Einbezug aller drei Messzeitpunkte varianzanalytisch untersucht. Die Gruppen unterschieden sich bei einem Signifikanzniveau von $p < 0,05$ zu allen drei Testzeitpunkten signifikant (vgl. Tab. 3). Allerdings lag der Mittelwert der Kontrollgruppe zu Beginn der Intervention höher als der der Experimentalgruppe. Dies ist insofern von Bedeutung, weil damit mögliche Effekte der Intervention nicht auf bereits höhere Werte der Experimentalgruppe zu Beginn der Intervention zurückzuführen wären. Die Analyse zeigte eine statistisch signifikante Interaktion von Gruppe und Zeit ($F(2, 358) = 30,50$; $p < 0,001$). Das bedeutet, dass sich die erfasste Fähigkeit in Abhängigkeit von der Gruppenzugehörigkeit unterschiedlich entwickelte. Die Effektstärke ($\eta^2 = 0,146$) kann als stark interpretiert werden (Cohen 1988). Aufgrund der Abnahme der Mittelwerte der Experimentalgruppe zwischen Post- und Follow-up-Test, wurde zur Beurteilung der Nachhaltigkeit der Intervention eine Varianzanalyse mit Messwiederholung mit dem ersten und dritten Messzeitpunkt durchgeführt. Es zeigte sich ein statistisch signifikanter Unterschied ($F(1, 179) = 32,02$; $p < 0,001$) mit einem starken Effekt ($\eta^2 = 0,152$).

Da zwei Werte zur Erfassung der Fähigkeit des Erkennens von Stärken und Schwächen zur Messung herangezogen wurden, ist hier zur differenzierten Beurteilung der Wirksamkeit der Intervention eine getrennte Betrachtung der beiden Werte sinnvoll. Mit Wert I wurde das *Identifizieren* von Stärken und Schwächen erfasst. Interessant ist, dass kein statistisch signifikanter Interaktionseffekt nachweisbar war ($F(1,92; 342,83) = 0,03$; $p = 0,963$). Beide Gruppen haben sowohl vor als auch nach der Intervention ähnliche Werte erreicht (vgl. Tab. 4 und Abb. 5).

Mit Wert II wurde das *Interpretieren* von Stärken und Schwächen von Schülerinnen und Schülern erfasst (vgl. Tab. 5 und Abb. 6). Es zeigte sich ein statistisch signifikanter Interaktionseffekt ($F(2, 358) = 40,37$; $p < 0,001$) und ein starker Effekt ($\eta^2 = 0,184$).

Tab. 4 Gruppenunterschiede beim Identifizieren von Stärken und Schwächen

Messzeitpunkt	Experimentalgruppe (n = 98)		Kontrollgruppe (n = 83)		ANOVA			
	M	SD	M	SD	Effekt	df	F	η^2
Pre	16,70	5,48	18,59	5,45	Gruppe	1,00	7,21	0,039
Post	17,29	5,50	19,18	5,24	Zeit	1,92	2,28	0,013
Follow-up	17,40	6,41	19,46	5,88	Gruppe \times Zeit	1,92	0,03	0,000

* $p < 0,05$; ** $p < 0,01$

Tab. 5 Gruppenunterschiede beim Interpretieren von Stärken und Schwächen

Messzeitpunkt	Experimentalgruppe (n = 98)		Kontrollgruppe (n = 83)		ANOVA			
	M	SD	M	SD	Effekt	df	F	η^2
Pre	24,21	6,89	25,23	6,55	Gruppe	1	26,99*	0,131
Post	33,56	12,29	23,82	6,67	Zeit	2	25,99**	0,127
Follow-up	28,60	10,31	20,76	6,70	Gruppe \times Zeit	2	40,37**	0,184

* $p < 0,05$; ** $p < 0,01$

5.2 Informationen im diagnostischen Prozess

Nachdem die Förderung diagnostischer Kompetenz, bezogen auf die Teilfähigkeit des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern, hinsichtlich ihrer Wirksamkeit betrachtet wurde, steht im Folgenden die Verwendung von Informationen im Fokus. Die Datengrundlage hierfür bildeten schriftliche Dokumente (Analysen diagnostischer Interviews) der Studierenden der Experimentalgruppe von selbst durchgeführten Interviews im Verlauf der Intervention.

5.2.1 Art der Informationen

In einem ersten Schritt der Analyse wurde untersucht, welche Informationen die Studierenden bei ihrer Diagnose nutzten (Forschungsfrage 2).²

Die verfügbaren Informationen (V), auf die die Studierenden Bezug nahmen, entstammten fachdidaktischem Wissen zu Größenbereichen, wie verschiedene Indikatoren für Größenvorstellungen (VI) sowie typische Herausforderungen des Größenbereichs Gewichte (V2), die Schwierigkeiten beim Aufbau von Größenvorstellungen verursachen können. Anhand der acht eingesetzten Aufgaben mit diagnostischem Potenzial konnten die Studierenden verschiedene Teilfähigkeiten von Schülerinnen und Schülern in Bezug auf den Umgang mit Gewichten einschätzen. Dazu zogen die Studierenden in ihrer Analyse verschiedene Kriterien heran. Die Studierenden schätzten ein, inwiefern bereits Kenntnisse zu Messinstrumenten vorhanden waren. Beim Schätzen von Gewichten diente den Studierenden als Kriterium, ob die Schülerin oder der Schüler passende Repräsentanten für Standardgrößen kannte und ob er oder sie auf Referenzgrößen (Stützpunktvorstellungen) zurückgreifen konnte. Ob

² Bezüge zu Kategorien aus Tab. 2 sind kursiv hervorgehoben.

Abb. 5 Gruppenunterschiede beim Identifizieren von Stärken und Schwächen

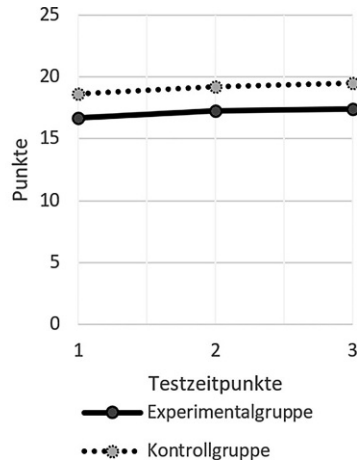
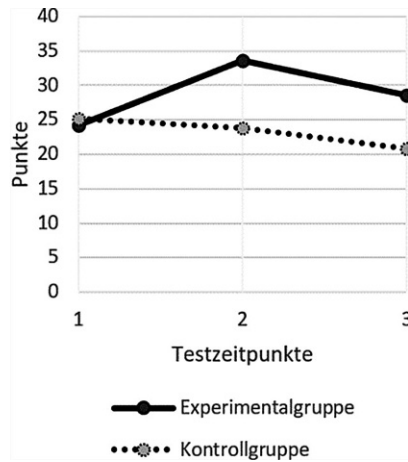


Abb. 6 Gruppenunterschiede beim Interpretieren von Stärken und Schwächen



die Schülerin oder der Schüler passende Maßeinheiten für Gegenstände auswählen konnte, ob diese zueinander in Beziehung (vergrößern/verfeinern) gesetzt werden konnten und die Unveränderlichkeit von Gewichten (Invarianz) erkannt wurde, waren weitere Kriterien, auf die sich die Studierenden in ihrer Analyse bezogen. Außerdem orientierten sich die Studierenden daran, ob das Kind Gegenstände nach Gewicht vergleichen und ordnen konnte.

Es wurde sichtbar, dass die Studierenden in ihrer Analyse die fachlichen Kriterien zur Beurteilung von Größenvorstellungen, die in den Interviewaufgaben angelegt waren, auch tatsächlich als Informationen nutzten. In den Analysen der Studierenden wurde aber auch deutlich, dass sie sich auf weiteres fachdidaktisches Wissen bezogen, das im Rahmen der Intervention thematisiert wurde. Dies zeigte sich beispielweise darin, dass sie auf *Herausforderungen des Größenbereichs Gewichte (V2)* Bezug nahmen. Sie nannten im Zusammenhang mit einer Tendenz von Schülerinnen und Schülern, das Gewicht mit der Größe des Gegenstandes in Verbindung zu

bringen, die nicht-visuelle Wahrnehmbarkeit von Gewichten. Auch das Problem, dass geringe Gewichtsunterschiede mit der Hand kaum wahrnehmbar sind, wurde als Information genutzt. Daneben verwendeten die Studierenden Informationen, die sich auf ihre Kenntnisse über gängigen *Mathematikunterricht* (V3), Lehrmittel bzw. auf den Lehrplan bezogen, beispielweise welche Repräsentanten für Gewichte typischerweise vorkommen.

Die *neu hinzukommenden Informationen* (N) sind dadurch gekennzeichnet, dass die Studierenden sie aus der Gesprächssituation mit der Schülerin oder dem Schüler entnahmen. Hierbei bezogen sie sich sowohl auf *richtige Lösungen* (N1) als auch auf *Fehler* (N2) der Schülerinnen und Schüler sowie genannte *Beispiele* (N3). Diese stellten Informationen dar, die im Gespräch über die Aufgaben zunächst konkret sichtbar und damit schnell erfassbar waren. Weiterhin wurden aber auch Informationen herangezogen, die von einer tiefergehenden Analyse zeugten. So wurden *Argumente* (N4) von Schülerinnen und Schülern berücksichtigt, die das Zustandekommen ihrer Lösung erklärten, beispielweise, dass die Materialbeschaffenheit oder die Größe von Gegenständen bei der Einschätzung von Gewichten eine Rolle spielte oder die Kategorisierung von Gegenständen in „schwer“ und „leicht“, z. B. bei der Zuordnung von Maßeinheiten. Informationen von den Schülerinnen und Schülern bezüglich ihres *Vorwissens* (N7), differenzierten die Studierenden in ihrer Analyse einerseits danach, woher dieses Vorwissen stammte, z. B. aus dem Unterricht oder aus dem Alltag und spezifizierten es, wenn es sich um eine konkrete Messerfahrung handelte. Neben der Argumentation der Schülerinnen und Schüler fokussierten die Studierenden auf Informationen zum *strategischen Vorgehen* (N5) der Kinder. So wurden verschiedene Strategien identifiziert, die bei der Aufgabebearbeitung sichtbar wurden. Die Studierenden bezogen sich auf Strategien wie raten, abschätzen (beispielweise durch das Vorstellen des Gewichts eines Gegenstandes in der Hand), systematisches Vergleichen und Ordnen von Gegenständen nach Gewicht. Neben diesen eher fachlich ausgerichteten Informationen, wurden auch Informationen zu allgemeineren Merkmalen genannt, wie dem *Bearbeitungsverhalten* (N6). Die Studierenden nahmen hier Bezug zur Bearbeitungsdauer und zur Reihenfolge der Bearbeitung einer Aufgabe, berücksichtigten aber auch, ob das Kind Sicherheit bzw. Unsicherheit zeigte und ob es gegebenenfalls seine Lösungen selbst korrigierte. Auch das Nachfragen der Schülerinnen und Schüler wurde in diesem Zusammenhang angeführt.

Es zeigte sich insgesamt, dass die Studierenden in ihrer Analyse Bezug auf ein breites Spektrum an Informationen nahmen. Die Informationen unterschieden sich nicht nur hinsichtlich der Quellen, sondern auch in ihrer unmittelbaren Wahrnehmbarkeit. Während etwa richtige Lösungen, Fehler oder eine länger andauernde Bearbeitungsdauer prinzipiell direkt wahrgenommen werden können, erfordert beispielweise die Identifikation von strategischem Vorgehen oder der Bezug zu Indikatoren für Größenvorstellungen eine Wahrnehmung tiefer liegender Strukturen.

5.2.2 Nutzung von Informationen bei der Formulierung von Stärken und Schwächen

Nachdem im ersten Schritt der qualitativen Auswertung der Fokus darauf lag, auf welche Informationen die Studierenden in ihrer Analyse Bezug nahmen, sollte in ei-

nem zweiten Schritt die Nutzung von Informationen genauer betrachtet werden (Forschungsfrage 3). Das Ziel dieser Analyse war es, zu beleuchten, inwiefern Informationen, die grundsätzlich in den Dokumenten genannt wurden, bei der Formulierung und Begründung von Stärken und Schwächen von den Studierenden herangezogen wurden.

Es zeigte sich, dass zunächst auf *verfügbare Informationen (V)* zurückgegriffen wurde. Dabei wurden die verschiedenen *Indikatoren für Größenvorstellungen (VI)* als Kriterien genutzt, um Kompetenzen der Schülerinnen und Schüler einschätzen zu können. Die fachlichen Kriterien dienten somit als Leitfaden für die Formulierung von Stärken und Schwächen, wurden jedoch nicht differenziert in Begründungen genutzt. Als Begründungen für Aussagen über Stärken und Schwächen wurden größtenteils Bezüge zu vermutetem *Vorwissen (N7)* der Kinder hergestellt, insbesondere fehlende bzw. vorhandene Alltags- und Messerfahrungen. Die typischen *Herausforderungen (V2)* des Größenbereichs Gewichte oder das Wissen über *Mathematikunterricht (V3)* wurden überwiegend nicht genutzt, um beispielweise mögliche Ursachen für Schwächen zu benennen, selbst wenn sie zuvor im Rahmen von Beschreibungen bei der Analyse verwendet wurden. Es zeigte sich außerdem, dass Aussagen zu Stärken und Schwächen häufig direkt aus einem *Fehler (N2)* oder aus einer *richtigen Lösung (N1)* des Kindes abgeleitet wurden, bei denen zwar der Bezug zu fachlichen Kriterien gegeben war, jedoch keine weiteren Informationen berücksichtigt wurden, um die Aussagen zu bestärken. Dies ließ sich ebenso feststellen, wenn Aussagen zu Stärken oder Schwächen des Kindes mit dem *Bearbeitungsverhalten (N6)* begründet wurden (z. B. einer langen bzw. kurzen Bearbeitungsdauer). Auch Bezüge zum *strategischen Vorgehen (N5)* der Kinder wurden in Begründungen für Stärken und Schwächen sichtbar, insbesondere das Ordnen von Gegenständen nach Größe. Allerdings wurde hier deutlich, dass das Benennen der Strategie nicht mit der nicht-visuellen Wahrnehmbarkeit von Gewichten (als typische *Herausforderung (V2)* der Größenbereichs Gewichte) in Verbindung gebracht wurde, wenn dies zu Fehleinschätzungen führte. Bei der Formulierung von Stärken und Schwächen fiel weiterhin auf, dass keine Bezüge zu *Argumenten (N4)* der Schülerinnen und Schüler hergestellt wurden. Auch *Beispiele (N3)* der Schülerinnen und Schüler wurden nicht herangezogen. Beim Bezug auf das *Bearbeitungsverhalten (N6)* zeigte sich, dass ausschließlich die Bearbeitungsdauer als Begründungen für eine Stärke oder Schwäche genutzt wurde, nicht jedoch beispielsweise das Revidieren einer Lösung oder die Reihenfolge der Bearbeitung, obwohl diese Aspekte in der Analyse genannt wurden.

Die Nutzung von Informationen bei der Beurteilung von Stärken und Schwächen von Schülerinnen und Schülern kann dahingehend gedeutet werden, dass die Studierenden in der Formulierung und Begründung von Stärken und Schwächen eines Kindes inhaltspezifisches fachdidaktisches Wissen benötigen, um sie mit Informationen, die in der Interviewsituation neu hinzukommen, zu fundierten Aussagen verknüpfen zu können. An vielen Stellen wurde damit die Verknüpfung von *verfügbaren Informationen (V)* und *neu hinzukommenden Informationen (N)* explizit. Dies lässt den Schluss zu, dass die fachlichen Kriterien zur Beurteilung von Kompetenzen von Schülerinnen und Schülern die Studierenden in ihrer Diagnose unterstützen.

Die Ergebnisse legen nahe, dass die Studierenden mehr Informationen wahrgenommen haben, als sie für die Formulierung von Stärken und Schwächen nutzten. Dies zeigte sich darin, dass zuvor thematisierte Informationen in den Analysen bei der Formulierung und Begründung von Stärken und Schwächen nicht mehr aufgegriffen wurden. Daher kann davon ausgegangen werden, dass die Studierenden die Informationen bezüglich ihrer Relevanz für die Beurteilung von Stärken und Schwächen werteten.

6 Diskussion

Ziel der Intervention war es, diagnostische Kompetenz im Rahmen der Ausbildung von Lehrkräften für die Primarschule zu fördern. Der Fokus lag auf einer zentralen Teilfähigkeit diagnostischer Kompetenz, der Fähigkeit des Erkennens von Stärken und Schwächen von Schülerinnen und Schülern. Diese Fähigkeit kann als grundlegend dafür betrachtet werden, dass auf der Grundlage (begründeter) Hypothesen bei der Beurteilung von Kompetenzen von Schülerinnen und Schülern auch individuelle Fördermaßnahmen eingeleitet werden können (Philipp 2018). Zielsetzung der Studie war es, einerseits die Wirksamkeit der Intervention zu prüfen (Forschungsfrage 1) und andererseits die Verwendung von Informationen im diagnostischen Prozess zu untersuchen (Forschungsfragen 2 und 3), ausgehend von der Annahme, dass das inhaltspezifische fachdidaktische Wissen, das Bestandteil der Intervention war, sichtbar wird.

6.1 Wirksamkeit der Intervention

Zur Überprüfung der Wirksamkeit der Intervention wurde ein Testinstrument eingesetzt, das es erlaubt, die Güte diagnostischer Urteile neben dem vorherrschenden Paradigma der Urteilsgenauigkeit zu bestimmen. Dies hat den Nachteil, dass die Ergebnisse aufgrund fehlender Korrespondenzmaße weniger gut mit Befunden anderer Studien vergleichbar sind. Von Vorteil ist jedoch, dass über die Anzahl von Aussagen bezüglich fachbezogener Kategorien besser Rückschlüsse auf das inhaltliche Lernen der Studierenden gezogen werden kann und damit auch über den Erwerb diagnostischer Kompetenz.

Die quantitativen Analysen zeigten einen deutlichen Zuwachs der erfassten Fähigkeit (das Erkennen von Stärken und Schwächen von Schülerinnen und Schülern) in der Experimentalgruppe über die drei Messzeitpunkte, während in der Kontrollgruppe kein Zuwachs zu verzeichnen war. Der Unterschied konnte mittels einer Varianzanalyse auf den Haupteffekt der Intervention zurückgeführt werden, der als groß zu werten ist. Das zeigt, dass die Förderung diagnostischer Kompetenz erfolgreich und nachhaltig war. Allerdings war ein Abfall der Werte zwischen dem zweiten und dritten Messzeitpunkt zu verzeichnen, d.h. die erreichten hohen Werte direkt nach der Intervention blieben nicht stabil. Das könnte daran liegen, dass zwischen den beiden letzten Messzeitpunkten das Thema Größen im Rahmen der Lehrveranstaltung nicht mehr behandelt wurde und somit das weitere Übungsfeld in diesem Bereich fehlte. Es könnte auch ein Hinweis darauf sein, dass diagnostische

Tätigkeiten kontinuierlicher angeregt werden müssen, um diagnostische Kompetenz nachhaltig zu fördern. Durch die Erfassung zweier Werte, das *Identifizieren* sowie das *Interpretieren* von Stärken und Schwächen, konnte die Wirksamkeit der Förderung diagnostischer Kompetenz differenzierter betrachtet werden. Beide Gruppen erreichten im Bereich Identifizieren über alle drei Messzeitpunkte hinweg ähnliche Werte. Daraus ist zu schließen, dass das Identifizieren von fehlerhaften und richtigen Überlegungen bzw. Lösungen durch die Intervention nicht gefördert werden konnte, wenngleich höhere Werte hätten erreicht werden können. Die Wirkung der Intervention konnte jedoch im Bereich der Interpretation von Stärken und Schwächen nachgewiesen werden. Das bedeutet, dass die Studierenden nach der Intervention eine größere Anzahl schlüssiger Interpretationen der Lösungen von Schülerinnen und Schülern, beispielweise über mögliche Ursachen für Schwierigkeiten, formulieren konnten. Dies zeigt, dass die Studierenden bei etwa gleichbleibender Anzahl von Identifikationen in der Lage waren, diese breiter zu interpretieren. Sie konnten also jeweils mehr fachbezogene Kriterien zur Beurteilung der Größenvorstellungen von Schülerinnen und Schülern heranziehen. Für eine förderorientierte Diagnostik stellt dies einen wesentlichen Schritt dar und dient als Basis für eine weiterführende, prozessorientierte Diagnostik, die theoriegeleitet ist (Philipp 2018; Scherer und Moser Opitz 2010).

Das fachdidaktische Wissen kann als Voraussetzung für die Entwicklung diagnostischer Kompetenz verstanden werden, die sich als beobachtbares Verhalten in der Formulierung von Aussagen zu Stärken und Schwächen von Schülerinnen und Schülern zeigte, also als Personeneigenschaft, die die situationsbezogenen Fähigkeiten beeinflusst (Blömeke et al. 2015). Der Einfluss des inhaltspezifischen fachdidaktischen Wissens auf die Güte des diagnostischen Urteils lässt sich durch den Vergleich mit der Kontrollgruppe untermauern, die sich bezüglich der besuchten Lehrveranstaltung in diesem Punkt unterschied. Dies könnte ebenfalls als Hinweis darauf gedeutet werden, dass diagnostische Urteile nicht nur domänen-, sondern auch inhaltspezifisch sein können bzw. durch solches Wissen beeinflusst werden. Steht dieses Wissen nicht zur Verfügung, so ist die Beurteilung von Kompetenzen von Schülerinnen und Schülern auf der Basis allgemeinen fachdidaktischen und diagnostischen Wissens erschwert. Dies könnte auch daran liegen, dass die Studierenden noch am Beginn der Entwicklung diagnostischer Kompetenz stehen und inhaltspezifische fachliche Kriterien benötigen, während erfahrenere Lehrkräfte möglicherweise flexibler mit fachdidaktischem Wissen umgehen können. Hinweise aus der Experten-Novizen-Forschung deuten an, dass Expertinnen und Experten im Vergleich zu Novizinnen und Novizen passende Kategorien aktivieren können (Krolak-Schwerdt et al. 2009). Möglicherweise unterstützte auch das Führen des diagnostischen Interviews die Verarbeitung des fachdidaktischen Wissens auf andere Weise als das in der Kontrollgruppe der Fall war.

Einschränkend ist zu sagen, dass aufgrund des Designs der Studie keine Aussage dazu gemacht werden kann, welches Element der Intervention (Auseinandersetzung mit diagnostischen Situationen und deren Reflexion, Aufbau fachdidaktischen Wissens) besonders förderlich war. Die Aussagen zur Wirksamkeit beziehen sich also auf das Gesamtkonzept der Intervention. In künftigen Studien könnte dies anhand

einer systematischen Variation verschiedener Elemente mit mehreren Gruppen in experimentellen Designs untersucht werden.

6.2 Verwendung von Informationen

Neben der Beurteilung der Wirksamkeit der Intervention, wurde im Rahmen der berichteten Studie auch der diagnostische Prozess selbst in den Blick genommen. Die Frage danach, auf welche Informationen überhaupt in einem solchen Prozess zurückgegriffen wird und wie sie genutzt werden, waren zentrale Fragestellungen. Im Gegensatz zu dem eingesetzten Testinstrument zur Überprüfung der Wirksamkeit der Intervention lag hier der Fokus auf einer Beurteilung von Kompetenzen eines Kindes über mehrere Aufgaben hinweg. Die Informationen mussten hier also integriert werden, um zu einer Gesamteinschätzung der Größenvorstellungen des Kindes zu kommen. Zentrales Ergebnis ist ein deduktiv sowie induktiv entwickeltes Kategoriensystem (vgl. Tab. 2), das Aufschluss über die Art der verwendeten Informationen gibt und zwischen Informationen, die bereits verfügbar sind (fachdidaktisches Wissen) und Informationen, die in einer diagnostischen Situation neu hinzukommen, unterscheidet. Eine Übertragbarkeit auf andere Inhaltsbereiche wäre bei Anpassung der inhaltsspezifischen Kategorien prinzipiell möglich. Des Weiteren zeigte sich, dass sowohl auf unmittelbar wahrnehmbare Informationen (z. B. Bearbeitungsdauer) zurückgegriffen wurde als auch auf Informationen, die nur durch die Wahrnehmung tieferliegender Strukturen erfasst werden können (z. B. strategisches Vorgehen der Schülerinnen und Schüler). Diese wurden jedoch nicht konsequent zur Begründung von Aussagen zu Stärken und Schwächen von Schülerinnen und Schülern genutzt. Von Novizinnen und Novizen wird berichtet, dass sie sich häufig an Oberflächenmerkmalen orientieren (Gruber 1994). Kellman und Massey (2013) weisen in diesem Zusammenhang darauf hin, dass Novizinnen und Novizen sowie Expertinnen und Experten durchaus dieselben Informationen berücksichtigen, sie aber erst mit zunehmender Expertise interpretieren können.

Die Nutzung von Informationen im Hinblick auf das Erkennen von Stärken und Schwächen von Schülerinnen und Schülern wurde über das gemeinsame Auftreten von Kategorien bei der Formulierung von Stärken und Schwächen des interviewten Kindes betrachtet. Hier zeigte sich insbesondere die Verknüpfung von verfügbaren Informationen und neu hinzukommenden Informationen. Dies wird als ein Beleg dafür gewertet, dass einzelne Informationen nicht getrennt voneinander betrachtet wurden, sondern dass sie zu Einheiten verknüpft wurden, bei denen Informationen aus beiden Informationsquellen integriert wurden. In den schriftlichen Analysen wurde deutlich, dass die Studierenden zunächst beschreibend vorgehen. Erst in der Formulierung von Kompetenzen und Schwierigkeiten des interviewten Kindes wurden Interpretationen deutlich. Es zeigte sich, dass die Studierenden dabei nicht alle Informationen nutzten, die sie bei der Analyse nannten. Das kann dahingehend gedeutet werden, dass die Studierenden im diagnostischen Prozess Informationen gewichteten. Allerdings wurde deutlich, dass aus Fehlern oder richtigen Lösungen direkt Formulierungen zu Stärken und Schwächen abgeleitet wurden. Solche vorschnellen Beurteilungen, die sich an Korrektheit orientieren, werden von Novizinnen und Novizen berichtet (Streit et al. 2019; Reinhold 2018). Zudem bezogen

sich die Studierenden stark auf das Bearbeitungsverhalten, insbesondere auf die Bearbeitungsdauer, sodass davon ausgegangen werden kann, dass der Prozess des Gewichtens gesonderter Unterstützung bedarf und in künftigen Studien fokussiert werden könnte.

Auch bei der qualitativen Analyse zeigte sich die Bedeutung des fachdidaktischen Wissens insofern, als dass die Studierenden auf das in der Intervention vermittelte fachdidaktische Wissen zurückgriffen, um Kompetenzen von Schülerinnen und Schülern zu beurteilen. Es kann daher als Ankerfunktion (Nickerson 1999) dienen und den Analyseprozess der Studierenden leiten. So wurden Teilkompetenzen von Größenvorstellungen (vgl. Abschn. 4.4.1) als Ankerpunkte verwendet. Dass die Studierenden verfügbare Informationen mit neu hinzukommenden Informationen verknüpften, um Aussagen über die Stärken und Schwächen des interviewten Kindes zu machen, kann als Anpassungsprozess (Nickerson 1999) betrachtet werden, bei dem die Menge an Informationen über mehrere Aufgaben hinweg verdichtet werden musste. Die Art der Informationen, die in einem solchen Anpassungsprozess verwendet wurden, konnten ergänzend zu den Befunden von Ostermann (2018) durch das Kategoriensystem konkretisiert werden.

Einschränkend ist anzumerken, dass nicht ausgeschlossen werden kann, dass die Studierenden bestimmte Informationen in der Interviewsituation gar nicht wahrgenommen (und damit auch nicht verschriftlicht) haben. Da die Studierenden unterschiedliche Kinder befragten, gab es keine „Vergleichsnorm“, an der die Güte der diagnostischen Urteile gemessen werden konnte. Der Fokus der Studie lag allerdings nicht auf der Akkuratheit der Urteile, sondern darauf, ob (1) sich die Anzahl der Aussagen zu Stärken und Schwächen von Schülerinnen und Schülern erhöhen ließ und (2) welche Informationen im diagnostischen Prozess genutzt wurden. Des Weiteren kann nicht mit Sicherheit gesagt werden, dass die Studierenden die verfügbaren Informationen tatsächlich verinnerlicht hatten, da die schriftlichen Analysen durch Hinzuziehen von Literatur entstanden sein könnten. Insofern müssen die Dokumente als Prozessdokumente im Lernprozess der Studierenden aufgefasst werden. Auch die Verallgemeinerbarkeit der qualitativen Befunde ist durch die reduzierte Stichprobe eingeschränkt.

6.3 Fazit

Die Befunde der hier berichteten Interventionsstudie weisen darauf hin, dass die Entwicklung diagnostischer Kompetenz durch die Förderung inhaltspezifischen fachdidaktischen Wissens bereits im Rahmen der Ausbildung angeregt werden kann, auch wenn davon auszugehen ist, dass sich diagnostische Kompetenz über einen längeren Zeitraum entwickelt (z. B. Hascher 2008). Die erfolgreiche Förderung diagnostischer Kompetenz im Rahmen der Interventionsstudie stützt damit auch die weit verbreitete Annahme, dass diagnostische Kompetenz wie andere Kompetenzen prinzipiell erlernbar ist (Hartig und Klieme 2006; Herppich et al. 2017) und kann dies konkretisieren: Förderung ist möglich durch Aufbau des fachbezogenen Wissens und durch die Konfrontation mit diagnostischen Situationen. Insgesamt geben die Ergebnisse der Studie wertvolle Hinweise auf den Erwerb diagnostischer Kompetenz und lassen darauf schließen, dass angehende Lehrkräfte nach entsprechender

Förderung in der Lage sind, fachbezogenes Wissen in diagnostischen Situationen zu nutzen, um Stärken und Schwächen von Schülerinnen und Schülern beurteilen können. Dabei wurde deutlich, welche Informationen die Studierenden bei der Diagnose berücksichtigen. Als Beitrag zur Förderung diagnostischer Kompetenz konnte somit ein Konzept entwickelt werden, das die Verbindung von fachbezogenem Wissen mit Informationen aus diagnostischen Situationen bei den Studierenden unterstützt.

Funding Open access funding provided by FHNW University of Applied Sciences and Arts Northwestern Switzerland

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 3, 175–193.
- v. Aufschnaiter, C., Cappell, J., Dübbelde, G., Ennemoser, M., Mayer, J., Stiensmeier-Pelster, J., Sträßer, R., & Wolgast, A. (2015). Diagnostische Kompetenz. Theoretische Überlegungen zu einem zentralen Konstrukt der Lehrerbildung. *Zeitschrift für Pädagogik*, 61, 738–758.
- Ball, D.L., Thames, M.H., & Phelbs, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, 59(5), 389–407.
- Bass, H., & Ball, D.L. (2004). A practice-based theory of mathematical knowledge for teaching: the case of mathematical reasoning. In W. Jianpan & X. Binyan (Hrsg.), *Trends and challenges in mathematics education* (S. 107–123). Shanghai: East China Normal University Press.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5–31.
- Bless, H., Fiedler, K., & Strack, F. (2004). *Social cognition—How individuals construct social reality*. Hove: Psychology Press.
- Blömeke, S., Gustafsson, J.E., & Shavelson, R.J. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3–13.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin Heidelberg: Springer.
- Bräunling, K., & Reuter, D. (2015). Messen und Größen. In J. Leuders & K. Philipp (Hrsg.), *Mathematik – Didaktik für die Grundschule* (S. 44–57). Berlin: Cornelsen.
- Brennan, R.L., & Prediger, D.J. (1981). Coefficient kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687–699.
- Bromme, R. (2008). Lehrerexpertise. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie* (S. 159–167). Stuttgart: Hogrefe.

- Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften* (S. 215–234). Münster: Waxmann.
- Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: a challenge to conventional interpretations. *Psychological Bulletin*, *104*(3), 396.
- Buholzer, A., & Zulliger, S. (2013). Die Entwicklung von diagnostischen Fähigkeiten bei Studierenden der pädagogischen Hochschulen im Laufe ihrer Ausbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, *31*(2), 186–197.
- Busch, J., Barzel, B., & Leuders, T. (2015). Die Entwicklung eines kategorialen Kompetenzmodells zur Erfassung diagnostischer Kompetenzen von Lehrkräften im Bereich Funktionen. *Journal für Mathematik-Didaktik*, *36*(2), 315–338.
- Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2020). Facilitating diagnostic competences in higher education—a meta-analysis in medical and teacher education. *Educational Psychology Review*, *32*, 157–196.
- Clarke, D. M., Roche, A., & Clarke, B. (2018). Supporting Mathematics Teachers' Diagnostic Competence Through the Use of One-to-One, Task-Based Assessment Interviews. In T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic Competence of Mathematics Teachers: Unpacking a Complex Construct in Teacher Education and Teacher Practice* (Bd. 11, S. 173–192). Cham: Springer.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2. Aufl.). New York: Academic Press.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin Heidelberg: Springer.
- Emmrich, A. (2008). Ein Vortest zu Gewichtsvorstellungen. *Grundschule Mathematik*, *5*(19), 14–19.
- Förster, N., & Karst, K. (2017). Modelle diagnostischer Kompetenz: Gemeinsamkeiten und Unterschiede. In A. Südkamp & A. K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 63–65). Münster: Waxmann.
- Franke, M., & Ruwisch, S. (2010). *Didaktik des Sachrechnens in der Grundschule*. Heidelberg: Spektrum.
- Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*(1), 184–205.
- Gruber, H. (1994). *Expertise: Modelle und empirische Untersuchungen*. Beiträge zur psychologischen Forschung, Bd. 34. Opladen: Westdeutscher Verlag.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Berlin, Heidelberg: Springer.
- Hascher, T. (2008). Diagnostische Kompetenzen im Lehrberuf. In C. Kraler & M. Schratz (Hrsg.), *Wissen erwerben, Kompetenzen entwickeln* (S. 71–86). Münster: Waxmann.
- Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world: the psychology of judgment and decision making*. Thousand Oaks: SAGE.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (2. Aufl.). Seelze: Klett.
- Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119–144). Hohengehren: Schneider.
- Herppich, S., Praetorius, A. K., Hetmanek, A., Glogger-Frey, I., Ufer, S., Leutner, D., Behrmann, L., Böhrer, I., Böhrer, M., Förster, N., Kaier, J., Karing, C., Karst, K., Klug, J., Ohle, A., & Südkamp, A. (2017). In A. Südkamp & A. K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 75–93). Münster: Waxmann.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, *59*(3), 297–313.
- Ingenkamp, K., & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6. Aufl.). Weinheim Basel: Beltz.
- Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for research in mathematics education*, *41*(2), 169–202.
- Kaiser, J., Praetorius, A. K., Südkamp, A., & Ufer, S. (2017). Die enge Verwobenheit von diagnostischem und pädagogischem Handeln als Herausforderung bei der Erfassung diagnostischer Kompetenz. In A. Südkamp & A. K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 114–123). Münster: Waxmann.
- Karing, C., & Seidel, T. (2017). Ausblick zur Förderung diagnostischer Kompetenz. In A. Südkamp & A. K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 240–246). Münster: Waxmann.

- Karst, K., Klug, J., & Ufer, S. (2017). Strukturierung diagnostischer Situationen im inner- und außerunterrichtlichen Handeln von Lehrkräften. In A. Südkamp & A. K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 102–114). Münster: Waxmann.
- Kellman, P. J., & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. *Psychology of learning and motivation*, 58, 117–165. Academic Press.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716–725.
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess: Der Lehrer als „flexibler Denker“. *Zeitschrift für Pädagogische Psychologie*, 23(34), 175–186.
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (3. Aufl.). Weinheim Basel: Beltz.
- Leuders, T., Dörfler, T., Leuders, J., & Philipp, K. (2018). Diagnostic competence of mathematics teachers: unpacking a complex construct. In T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic competence of mathematics teachers: unpacking a complex construct in teacher education and teacher practice* (Bd. 11, S. 3–31). Cham: Springer.
- Leuders, T., Loibl, K., & Dörfler, T. (2020). Diagnostische Urteile von Lehrkräften erklären – Ein Rahmenmodell für kognitive Modellierungen und deren experimentelle Prüfung. *Unterrichtswissenschaft*, 48(4), 493–502.
- Lorenz, C., & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23(3), 211–222.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (11. Aufl.). Weinheim, Basel: Beltz.
- Mayring, P., & Fenzl, T. (2019). Qualitative Inhaltsanalyse. In N. Bauer & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (2. Aufl. S. 633–648). Wiesbaden: Springer VS.
- Moser Opitz, E., & Nührenbörger, M. (2015). Diagnostik und Leistungsbeurteilung. In R. Bruder, L. Heffendehl-Hebeker, B. Schmidt-Thieme & H. G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 491–512). Berlin: Springer Spektrum.
- Nickerson, R. S. (1999). How we know-and sometimes misjudge-what others know: imputing one's own knowledge to others. *Psychological Bulletin*, 125(6), 737–759.
- van Ophuysen, S., & Behrmann, L. (2015). Die Qualität pädagogischer Diagnostik im Lehrerberuf-Anmerkungen zum Themenheft „Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz“. *Journal for educational research online*, 7(2), 82–98.
- Ostermann, A. (2018). Factors influencing the accuracy of diagnostic judgments. In T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic competence of mathematics teachers: unpacking a complex construct in teacher education and teacher practice* (Bd. 11, S. 95–108). Cham: Springer.
- Peter-Koop, A., & Nührenbörger, M. (2008). Größen und Messen. In G. Walther, M. van den Heuvel-Panhuizen, D. Gränzer & O. Köller (Hrsg.), *Bildungsstandards für die Grundschule: Mathematik konkret* (S. 87–115). Berlin: Cornelsen Scriptor.
- Philipp, K. (2018). Diagnostic competences of mathematics teachers with a view to processes and knowledge resources. In T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic competence of mathematics teachers: unpacking a complex construct in teacher education and teacher practice* (Bd. 11, S. 109–127). Cham: Springer.
- Plessner, H. (2011). Urteilen. In T. Betsch, J. Funke & H. Plessner (Hrsg.), *Denken – Urteilen, Entscheiden, Problemlösen*. Berlin, Heidelberg: Springer.
- Pott, A. (2019). *Diagnostische Deutungen Im Lernbereich Mathematik*. Wiesbaden: Springer.
- Prediger, S., Selter, C., Hußmann, S., & Nührenbörger, M. (Hrsg.). (2017). *Mathe sicher können. Handreichungen für ein Diagnose- und Förderkonzept zur Sicherung mathematischer Basiskompetenzen. Sachrechnen*. Berlin: Cornelsen.
- Prediger, S., Tschierschky, K., Wessel, L., & Seipp, B. (2012). Professionalisierung für fach- und sprachintegrierte Diagnose und Förderung im Mathematikunterricht. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 17(1), 40–58.
- Rädiker, S., & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA*. Wiesbaden: Springer.
- Reinhold, S. (2018). Revealing and promoting pre-service teachers' diagnostic strategies in mathematical interviews with first-graders. In T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic competence*

- of mathematics teachers: unpacking a complex construct in teacher education and teacher practice* (Bd. 11, S. 129–148). Cham: Springer.
- Reuter, D. (2011). *Kindliche Konzepte zur Größe Gewicht und ihre Entwicklung. Theoretische Modellierung und zwei Einzelfallstudien mit Drittklässlern*. Dissertation. Berlin: Humboldt-Universität.
- Rösike, K. A., & Schnell, S. (2017). Do math! – Lehrkräfte professionalisieren für das Erkennen und Fördern von Potenzialen. In J. Leuders, T. Leuders, S. Prediger & S. Ruwisch (Hrsg.), *Mit Heterogenität im Mathematikunterricht umgehen lernen* (S. 223–233). Wiesbaden: Springer Spektrum.
- Scherer, P., & Opitz, M. E. M. (2010). *Fördern im Mathematikunterricht der Primarstufe*. Heidelberg: Springer Spektrum.
- Schipper, W. (2009). *Handbuch für den Mathematikunterricht an Grundschulen*. Braunschweig: Schroedel.
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 154–165.
- Schwarz, B., Wissmach, B., & Kaiser, G. (2008). “Last curves not quite correct”: diagnostic competence of future teachers with regard to modelling and graphical representations. *ZDM—The International Journal on Mathematics Education*, 40(5), 777–790.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 85–95.
- Stahnke, R., Schueler, S., & Roesken-Winter, B. (2016). Teachers’ perception, interpretation, and decision-making: a systematic review of empirical mathematics education research. *ZDM—The International Journal on Mathematics Education*, 48(1), 1–27.
- Streit, C., Rüede, C., Weber, C., & Graf, B. (2019). Zur Verknüpfung von Lernstandeinschätzung und Weiterarbeit im Arithmetikunterricht: Ein kontrastiver Vergleich zur Charakterisierung diagnostischer Expertise. *Journal für Mathematik-Didaktik*, 40(1), 37–62.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers’ judgments of students’ academic achievement: a meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762.
- Tversky, A., & Kahneman, D. (1974). Judgments under uncertainty. Heuristic and biases. *Science*, 185, 1124–1131.
- Ufer, S., & Leutner, D. (2017). Kompetenzen als Dispositionen-Begriffsklärungen und Herausforderungen. In A. Südkamp & A. K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 67–74). Münster: Waxmann.
- Weinert, F. E. (2000). Lehren und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule. *Pädagogische Nachrichten Rheinland-Pfalz*, 2, 1–16.
- Wember, F. B. (1998). Zweimal Dialektik: Diagnose und Intervention, Wissen und Intuition. *Sonderpädagogik*, 28, 106–120.
- Wollring, B., Peter-Koop, A., & Grüßing, M. (2013). Das ElementarMathematische BasisInterview EMBI. In M. Hasselhorn, A. Heinze, W. Schneider & U. Trautwein (Hrsg.), *Diagnostik mathematischer Kompetenzen* (Bd. 11, S. 81–96). Göttingen: Hogrefe.