



# Das Beurteilen von Lernprodukten als Facette diagnostischer Kompetenz fördern

Andreas Eichler · Elisabeth Rathgeb-Schnierer ·  
Jan Philipp Volkmer

Eingegangen: 30. Dezember 2020 / Angenommen: 3. November 2022 / Online publiziert: 2. Dezember 2022  
© Der/die Autor(en) 2022

**Zusammenfassung** Diagnostische Kompetenz ist eine zentrale Komponente der professionellen Kompetenzen von Lehrkräften, die die Qualität von Unterricht und somit das Lernen von Schülerinnen und Schülern beeinflusst. Aufgrund dieser zentralen Bedeutung rückt die systematische Schulung der diagnostischen Kompetenz zunehmend in den Forschungsfokus. Diese Arbeit zielt auf die Erforschung der Förderung diagnostischer Kompetenz im Bereich der Mathematik ab. Wir stellen eine quasi-experimentelle Studie mit Treatment- und Kontrollgruppe vor, die den Effekt einer Intervention bei Studierenden des Grundschullehramts ( $n = 74$ ) untersucht. Die Intervention erfolgt im Rahmen eines Seminars, das die als effektiv bekannten Bestandteile der Förderung diagnostischer Kompetenz aufgreift. Diagnostische Kompetenz modellieren wir als Fähigkeit, Lernprodukte von Schülerinnen adäquat und multiperspektivisch beurteilen zu können und messen deren Entwicklung in einem Pre-Post-Design. Die Ergebnisse zeigen, dass unsere Schulung insbesondere auf einen Teil sogenannter epistemischer Aktivitäten wirkt: die Entwicklung und Stützung von Hypothesen zu Fähigkeiten von Schülerinnen und Schülern. Hier unterscheidet sich die Treatmentgruppe im Post-Test signifikant von der Kontrollgruppe. Die Arbeit leistet insgesamt einen Beitrag zur Entwicklung und multiperspektivischen Messung diagnostischer Kompetenz bezogen auf epistemische Aktivitäten im diagnostischen Prozess und der multiperspektivischen Beurteilung der Lernprodukte von Schülerinnen und Schülern zu offenen Lernangebote zur Arithmetik.

---

Dieser Beitrag wurde ursprünglich für das im März 2022 erschienene Heft mit Themenschwerpunkt „Diagnostisches Denken und Handeln von Mathematiklehrkräften – Theoretische Fundierungen, empirische Befunde und Implikationen für die Praxis“ eingereicht. Aufgrund von Verzögerungen erfolgt die Veröffentlichung nun in einer regulären Ausgabe des *Journals für Mathematik-Didaktik*.

✉ Andreas Eichler · Elisabeth Rathgeb-Schnierer · Jan Philipp Volkmer  
Institut für Mathematik, Universität Kassel, Heinrich-Plett-Straße 40, 34132 Kassel, Deutschland  
E-Mail: [eichler@mathematik.uni-kassel.de](mailto:eichler@mathematik.uni-kassel.de)

**Schlüsselwörter** Diagnostische Kompetenz · Entwicklung diagnostischer Kompetenz · Messung diagnostischer Kompetenz · Kontrastieren und Vergleichen · Offene Lernangebote

**MESC Codes** C70 · D70 · F30

## **Promoting Diagnostic Competence when Assessing Written Solutions of School Students**

**Abstract** Diagnostic competence is a significant part of teachers' professional competencies that influences the quality of teaching and students' learning. Therefore, developing teachers' diagnostic competence has become a relevant topic of educational research. In this regard, we present a quasi-experimental study that investigates the effect of an intervention on the promotion of diagnostic competence in prospective elementary teachers. The study follows a pre-post design. It comprises a treatment group and a control group ( $N=75$ ). The intervention takes place as part of a university seminar and includes components that are known to be effective for promoting diagnostic competence. We measure prospective teachers' diagnostic competence regarding so-called epistemic activities when assessing students' solutions to problem-oriented tasks. The results reveal our treatment to be effective, particularly regarding generating and reasoning hypotheses about students' abilities. With the specific focus on epistemic activities in the diagnostic process, our research contributes to developing and measuring diagnostic competence in a multi-perspective way.

**Keywords** Diagnostic competence · Developing diagnostic competence · Measuring diagnostic competence · Comparing and contrasting · Inquiry-based learning

### **1 Einführung**

Diagnostische Kompetenz wird als eine von vier Schlüsselkompetenzen von Lehrkräften bezeichnet (Weinert 1998). Baumert und Kunter (2006, S. 489) stellen fest, dass „kaum eine Komponente des professionellen Wissens und Könnens so prominent [ist] wie die so genannte diagnostische Kompetenz“. Diagnostische Kompetenz als Fähigkeit, sowohl Schülerinnen und Schüler als auch Aufgaben- und Lernanforderungen zutreffend einzuschätzen und zu beurteilen (Artelt und Gräsel 2009; Schrader 2013), wird als wichtige Voraussetzung für die effektive Planung und Durchführung von Unterricht (McElvany et al. 2009) und als Schlüssel für eine adäquate Steuerung von Unterrichtsprozessen verstanden (Schrader 2013). Chernikova et al. (2020, S. 158) kommen zum generellen Schluss: „Making efficient decisions in professional fields is impossible without being able to identify, understand, and even predict situations and events relevant to the profession“. Anders et al. (2010) sehen in der diagnostischen Kompetenz die Grundlage von Unterrichtsqualität (vgl. Busch et al. 2015; Praetorius et al. 2020), die zudem einen positiven Effekt auf das

Lernen von Schülerinnen und Schülern hat (Anders et al. 2010; Hattie 2009). Obwohl es zum letztgenannten Aspekt auch Studien mit divergenten Ergebnissen gibt (z. B. Karing et al. 2011), halten Leuders et al. (2018, S. 4) fest: „there is no doubt about the impact of diagnostic competences of teachers on the learning outcome of students“.

Aufgrund der zentralen Bedeutung der diagnostischen Kompetenz von Lehrkräften für die Unterrichtsqualität und das Lernen von Schülerinnen und Schülern rückte die systematische Entwicklung dieser Kompetenz in den letzten Jahren zunehmend in den Forschungsfokus (Busch et al. 2015; Chernikova et al. 2020; Schrader 2013; v. Aufschnaiter et al. 2015). Bei diesem Forschungsfokus wird davon ausgegangen, dass diagnostische Kompetenz ohne Schulung nicht angemessen ausgeprägt ist (vgl. Krauss und Brunner 2011; van Ophuysen und Behrmann 2015), Kompetenzen aber generell „erlernbare kontextspezifische Leistungsdispositionen“ darstellen (Klieme und Hartig 2007, S. 17). Diagnostische Kompetenz gilt damit als lern-, trainier- und erweiterbar (vgl. Herppich et al. 2018; Klug et al. 2016; Schäfer und Seidel 2015).

In diesem Beitrag stellen wir eine an bisherigen empirischen Ergebnissen orientierte Entwicklung und Evaluation einer Intervention zur Förderung diagnostischer Kompetenz von Studierenden des Grundschullehramts vor. Die Intervention greift ein Bündel von Eigenschaften von Lehrkonzepten auf, dessen Bestandteile sich bisher als erfolgreich erwiesenen haben (Chernikova et al. 2020), und orientiert sich zudem am stetigen Vergleichen (Alferi et al. 2013) als Bestandteil des diagnostischen Prozesses (Chernikova et al. 2020; Philipp 2018). Die Förderung der diagnostischen Kompetenz der Studierenden wird unter dem Aspekt des Beurteilens von Lernprodukten untersucht. In der bisherigen Forschung zur diagnostischen Kompetenz lag und liegt ein besonderer Schwerpunkt auf der Messung von Urteilsgenauigkeit. Bei diesem Ansatz wird untersucht „wie gut Lehrpersonenurteile mit den bei den Schülerinnen und Schülern gemessenen ‚tatsächlichen‘ Merkmalsausprägungen übereinstimmen“ (Schrader 2013, S. 157), d. h. es wird eindimensional die akkurate Beurteilung zukünftiger Lösungen von Lernenden adressiert (z. B. Schrader 2013; Loibl et al. 2020). Wir beziehen uns auf eine alternative Untersuchungsrichtung und erfassen als Facette diagnostischer Kompetenz die Fähigkeit, existierende Lernprodukte von Schülerinnen und Schülern adäquat und multiperspektivisch beurteilen zu können (Busch et al. 2015; Chernikova et al. 2020; Philipp 2018; Weinsheimer 2016). Für die multiperspektivische Messung dieser Facette der diagnostischen Kompetenz angehender Lehrkräfte differenzieren wir drei Bereiche mathematischer Kompetenzen von Schülerinnen und Schülern (Rathgeb-Schnierer und Schütte 2011), zu denen jeweils sogenannte epistemische Aktivitäten (Chernikova et al. 2020; Fischer et al. 2014) bei der Diagnose der Lernprodukte untersucht werden. Der spezifische Fokus dieser Arbeit liegt insbesondere auf der differenzierten Beschreibung der epistemischen Aktivitäten als Wirkung der Intervention zur Förderung diagnostischer Kompetenzen. Dabei steht die Frage im Zentrum, welche Aspekte der adäquaten Beurteilung von Lernprodukten von Schülerinnen und Schülern als Bestandteile diagnostischer Kompetenz von Lehrkräften sich durch eine spezifische Intervention fördern lassen. Dieser Frage wird im Rahmen einer quasi-experimentellen Studie mit Treatment- und Kontrollgruppe bei Studierenden des Grundschullehramts ( $n=74$ ) nachgegangen. Die Studie selbst ist innerhalb ei-

nes Verbundprojekts zum Kontrastieren und Vergleichen in der Lehramtsausbildung eingebettet (KoVeLa, [www.uni-kassel.de/go/kovela](http://www.uni-kassel.de/go/kovela)).

## 2 Diagnostische Kompetenz

### 2.1 Begriffsbestimmung

Über die Bedeutung der diagnostischen Kompetenz für das professionelle Handeln von Lehrkräften herrscht Konsens, nicht aber über deren Definition und die Frage, welche Facetten sie umfasst (Dübbelde 2013). Die Vielschichtigkeit des Konstrukts (McElvany et al. 2009; Südkamp und Praetorius 2017) wird in vielen Übersichtsbeiträgen zur diagnostischen Kompetenz betont: „Although many authors use the term ‚diagnostic competence‘, there is a wealth of understandings, of varying notions and of differing emphases with respect to its definition and constituents“ (Leuders et al. 2018, S. 6).

Wir verstehen diagnostische Kompetenz in Anlehnung an die Kompetenzdefinition von Klieme und Leutner (2006, S. 879) als „kontextspezifische kognitive Leistungsdisposition“, die es Lehrkräften ermöglicht, diagnostische Anforderungen in ihrem Handlungsfeld zu bewältigen (Herppich et al. 2018). Konkret lässt sich diagnostische Kompetenz beschreiben 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“ (Weinert 2000, S. 19). Wenngleich diagnostische Kompetenz von Lehrkräften auch die Beurteilung von Lernaufgaben umfasst (vgl. auch Artelt und Gräsel 2009), fokussieren wir hier auf die Fähigkeit, Schülerinnen und Schüler adäquat einzuschätzen. Diese wird in der Forschung häufig über die Urteilsgenauigkeit erfasst, also die Fähigkeit von Lehrkräften, die Performanz von Lernenden eindimensional (z. B. bezogen auf die zu erwartende Punktzahl bei einem vorgegebenen Test) einschätzen zu können (z. B. Anders et al. 2010; Helmke et al. 2004; Loibl et al. 2020; Rieu et al. 2020; Schrader 2013). Praetorius et al. (2012) haben an diesem Ansatz allerdings die Ferne zu Unterrichtsprozessen kritisiert. Alternativ wird diagnostische Kompetenz mit Fokus auf die Schülerinnen und Schüler so modelliert, dass sie über die Einschätzung von deren Lösungen im Hinblick auf ein richtiges oder falsches Ergebnis hinausgeht. Diesem Ansatz folgen wir und erfassen in Anlehnung an Busch et al. (2015), Chernikova et al. (2020) und Philipp (2018) die Fähigkeit von Lehrkräften, existierende Lernprodukte von Schülerinnen und Schülern bezogen auf die Kompetenzen von Schülerinnen und Schülern einschätzen zu können als eine Facette diagnostischer Kompetenz.

Bei diagnostischer Kompetenz wird eine Kontextabhängigkeit angenommen, d. h. sie ist nicht allgemein vorhanden, sondern bezieht sich auf ein Fach, wie beispielsweise Mathematik, bzw. auf spezifische Inhaltsbereiche (z. B. Lorenz und Artelt 2009). Chernikova et al. (2020) betonen zudem, dass die diagnostische Kompetenz vom bereichsspezifischen Wissen abhängig ist. Der inhaltliche Kontext der vorliegenden Arbeit ist die Arithmetik, also eine Subdomäne der Mathematik. Diese Wahl begründet sich dadurch, dass Arithmetik in der Grundschule, der zukünftigen Schul-

form für die angehenden Lehrkräfte in dieser Studie, ein zentraler Bestandteil des Mathematikunterrichts ist.

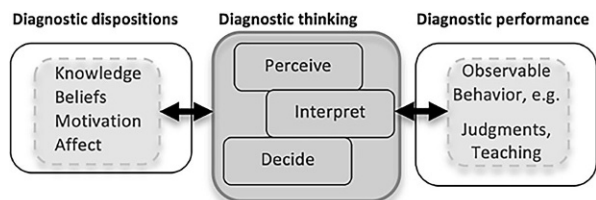
## 2.2 Verortung des Forschungsansatzes in theoretischen Modellen zur Beschreibung diagnostischer Kompetenz

Die Einbettung der diagnostischen Kompetenz von Lehrkräften in einen allgemeinen Kompetenzbegriff ist weit verbreitet (Herppich et al. 2018). Es existieren verschiedene Modelle, die sich im Detail hinsichtlich der Verankerung der diagnostischen Kompetenz in weiteren Aspekten der professionellen Kompetenz von Lehrkräften unterscheiden (Ball et al. 2008; Brunner et al. 2011; Krauss und Brunner 2011; Dübeldel 2013). In dieser Arbeit verstehen wir mit Brunner et al. (2011) das fachdidaktische und das pädagogisch psychologische Wissen als grundlegende Disposition der diagnostischen Kompetenz (vgl. auch Chernikova et al. 2020).

Für die spezifische Verortung unseres Forschungsansatzes greifen wir auf die Adaption des von Blömeke et al. (2015) entwickelten Kompetenzmodells in der Arbeit von Leuders et al. (2018) zurück. Dort wird diagnostische Kompetenz als Kontinuum beschrieben, das die Bereiche „diagnostic dispositions“ (diagnostische Disposition), „diagnostic thinking“ (diagnostisches Denken) und „diagnostic performance“ (diagnostische Performanz) einschließt (Leuders et al. 2018, S. 8). Der Fokus auf einen oder mehrere dieser Bereiche wird durch die spezifischen Forschungsfragen bestimmt. In unserer Forschung fokussieren wir auf den Bereich des diagnostischen Denkens (Abb. 1; Leuders et al. 2018, S. 15).

Leuders et al. (2018) stellen fest, dass es für die kognitiven Prozesse im Bereich des diagnostischen Denkens (wahrnehmen, interpretieren und entscheiden) verschiedene Modellierung gibt, zu denen beispielsweise in der Bildungsforschung das Noticing (van Es und Sherin 2002) oder das Decision-Making (Schoenfeld 2011) gehören. Leuders et al. (2018) verweisen als mögliche alternative Modellierung dieser kognitiven Prozesse zusätzlich auf das aus dem medizinischen Bereich stammende „clinical reasoning“. In unserer Arbeit orientieren wir uns explizit an dem zuletzt genannten Ansatz und nutzen die von Chernikova et al. (2020) vorgestellte Synthese sogenannter diagnostischer Prozesse als gemeinsamer Bestandteil diagnostischer Kompetenz von medizinischem Personal und Lehrkräften. Die diagnostischen Prozesse gehen auf grundlegende kognitive Prozesse im wissenschaftlichen Denken zurück (Fischer et al. 2014) und können als spezifische Modellierung der bei Leuders et al. (2018) beschriebenen kognitive Prozesse des diagnostischen Denkens aufgefasst werden, die sich auf die Beurteilung von Lernprodukten von Schülerinnen und Schülern beziehen.

**Abb. 1** Adaption des Kompetenzmodells von Blömeke et al. (2015) durch Leuders et al. (2018, S. 15) mit dem Fokus auf „diagnostic thinking“



Nach Chernikova et al. (2020) können die diagnostischen Prozesse durch epistemische Aktivitäten beschrieben werden. Zu diesen epistemischen Aktivitäten gehören „(a) identifying a problem, (b) questioning, (c) generating hypotheses, (d) constructing artefacts, (e) generalizing evidence, (f) evaluating evidence, (g) drawing conclusions, and (h) communicating process and results“ (Chernikova et al. 2020, S. 161; vgl. auch Fischer et al. 2014). In dieser Arbeit werden die kognitiven Prozesse des diagnostischen Denkens angehender Lehrkräfte innerhalb der Kompetenzfacette der Beurteilung von Lernprodukten von Schülerinnen und Schülern durch die genannten epistemischen Aktivitäten modelliert.

### 2.3 Förderung diagnostischer Kompetenz

Chernikova et al. (2020) fassen in ihrer Meta-Studie Forschungsarbeiten zusammen, die die Förderung diagnostischer Kompetenz von Lehrkräften adressieren. Als wesentliche Elemente einer wirksamen Schulung gelten die Problemorientierung und das Scaffolding (vgl. Belland et al. 2017). Die Metapher des Scaffolding wurde erstmals von Wood et al. (1976, S. 90) im Zusammenhang mit der Unterstützung von Problemlöseaktivitäten durch gezielte Motivation und adressatenbezogene Auswahl geeigneter Bestandteile genutzt. Generell wird unter Scaffolding verstanden, dass die Lehrkraft die Lernenden dann bei der Bearbeitung von Aufgaben unterstützt, wenn diese ohne Unterstützung nicht gelöst werden können. Die drei zentralen Merkmale hierbei sind „contingency“ (van de Pol et al. 2010, S. 274) „fading“ und „transfer of responsibility“ (van de Pol et al. 2010, S. 275). Chernikova et al. (2020, S. 162) beziehen das Scaffolding beim Erlernen diagnostischer Kompetenz spezifisch auf die Steuerung des diagnostischen Prozesses durch „(a) providing examples [...], (b) providing prompts [...], (c) assigning roles and (d) including reflection phases“. In der Metaanalyse von Chernikova et al. (2020) haben sich alle vier Aspekte des Scaffolding auch bei der Förderung diagnostischer Kompetenz im mathematischen Zusammenhang als wirksam erwiesen. Dabei sind in die dort betrachteten Studien wie auch die im Folgenden diskutierten Studien jeweils mehrere der vier Aspekte des Scaffolding eingegangen, während eine differenzierte experimentelle Analyse einer der Aspekte kein explizites Ziel der Studien war.

Hinsichtlich der mathematischen Aus- und Fortbildung hat sich das Bearbeiten konkreter, fallbasierter Aufgaben (providing examples), wie beispielsweise die „Analyse schriftlicher und mündlicher Schülerlösungsprozesse“ (Besser et al. 2015, S. 114), in diversen schulbezogenen Studien als wirksam erwiesen (Besser et al. 2015; Busch et al. 2015; Gold et al. 2013; Klug et al. 2016). In zwei der Studien wurde wie in dieser Arbeit auf Lernprodukte fokussiert (Besser et al. 2015; Busch et al. 2015), allerdings die Kompetenzentwicklung bei Lehrkräften untersucht. Da Chernikova et al. (2020) darauf hinweisen, dass insbesondere wenig fortgeschrittene Lernende von fallbasierten Diagnoseaufgaben profitieren, nehmen wir an, dass auch Lehramtsstudierende von dieser Art von Scaffolding profitieren. Heitzmann et al. (2018) zeigen weiterhin, dass ein Prompt zur Fehlererklärung einen positiven Effekt auf die diagnostische Kompetenz von Lehramtsstudierenden hat (providing prompts). Die Einnahme der Rolle als diagnostisch tätige Lehrkraft (assigning roles) ist in nahezu allen Studien zur Entwicklung der diagnostischen Kompetenz von Ma-

thematiklehrkräften enthalten (vgl. Besser et al. 2015; Busch et al. 2015; Gold et al. 2013; Klug et al. 2016). In wenigen Studien ist auch die Übernahme der Rolle als Lehrkraft im realen Unterricht vorhanden (z. B. Besser et al. 2015). Ebenso ist die Reflexion als angeleitete kritische Rückschau (reflection phases) Bestandteil nahezu aller Interventionen zur Förderung der diagnostischen Kompetenz von Mathematiklehrkräften (vgl. Besser et al. 2015; Busch et al. 2015; Gold et al. 2013; Klug et al. 2016). Chernikova et al. (2020) stellen fest, dass auch die Wirkung von Reflexionsphasen vom Vorwissen der geschulten Lehrkräfte abhängt, wobei von diesen Phasen insbesondere fortgeschrittene Lernende zu profitieren scheinen.

Als weiteren Aspekt einer effektiven Schulung diagnostischer Kompetenz betonen Chernikova et al. (2020) den konkreten Bezug zu den oben genannten epistemischen Aktivitäten, die auch in einer größeren Variante als diagnostische Mikro-Prozesse bei Reinhold (2018) dargestellt sind. Fischer et al. (2014) beschreiben die epistemischen Aktivitäten auf der Basis dreier Modi des wissenschaftlichen Denkens: einem Generieren und Testen von Hypothesen, einer wissenschaftlich begründeten Entwicklung von Lösungen und einem systematischen, iterativen Entwickeln und Prüfen von Prototypen (artefacts). Da sich nicht alle epistemischen Aktivitäten (s. oben, Abschn. 2.2) auf die Beurteilung von Lernprodukten von Schülerinnen und Schülern übertragen lassen, haben wir in dieser Arbeit folgende Anpassungen vorgenommen:

- Als „identifying a problem“ verstehen wir im Sinne von Fischer et al. (2014, S. 33) das Erkennen manifester Merkmale in einem Lernprodukt von Schülerinnen und Schülern sowie die Entscheidung, welche dieser Merkmale für den diagnostischen Prozess relevant sind. Ein Lernprodukt weist beispielsweise direkt beobachtbare, also manifeste Merkmale auf, wie ein richtiges oder falsches Ergebnis bei einer Additionsaufgabe.
- Als „questioning“ (Fischer et al. 2014, S. 33) bezeichnen wir den Abgleich von wahrgenommenen Merkmalen mit möglichen Ursachen.
- Bei der „hypothesis generation“ (Fischer et al. 2014, S. 33) werden idealerweise aufbauend auf der Wahrnehmung manifester Merkmale Hypothesen zu Kompetenzen einer Schülerin bzw. eines Schülers entwickelt.
- Die „evidence evaluation“ (Fischer et al. 2014, S. 34) umfasst die Prüfung des Gehalts bzw. der Evidenz einer entwickelten Hypothese.
- Mit dem „drawing conclusions“ (Fischer et al. 2014, S. 35) schließt der diagnostische Prozess ab. Dazu gehört zunächst die Festlegung auf eine Diagnose. Aufbauend auf dieser Diagnose zu einem Lernprodukt sowie der Evaluation der Evidenz des diagnostischen Urteils wird nun die Entscheidung für das weitere diagnostische Vorgehen getroffen.

Die epistemischen Aktivitäten „construction and redesign of artefacts“ sowie „evidence generation“ (Fischer et al. 2014, S. 34) beziehen sich auf mehrere Zyklen einer Entwicklung, Erhebung und Prüfung und sind daher auf die punktuelle Beurteilung von Lernprodukten von Schülerinnen und Schülern nicht anwendbar.

Alle epistemischen Aktivitäten, die diagnostischen Prozessen zugrunde liegen, bauen explizit oder implizit auf einem Vergleich auf. Ein Problem kann etwa nur identifiziert werden, wenn es ein Vergleichsmuster zu einer Lösung gibt. Chernikova

et al. (2020) nennen den Vergleich mit Lernzielen als zentralen Bestandteil eines diagnostischen Prozesses: „More generally, diagnosing first focuses on comparing the current state of learners’ knowledge and skills to predefined learning objectives.“ Philipp (2018) bezeichnet auf der Basis von Nickerson (1999) den fortwährenden Vergleich eines Lernprodukts mit relevanten Vergleichsobjekten ebenfalls als zentralen Bestandteil des diagnostischen Prozesses.

### 3 Materialien, Hypothesen und Methode

#### 3.1 Training zur Förderung diagnostischer Kompetenz

Die Konzeption des Seminars zur Entwicklung diagnostischer Kompetenz (14 Semesterwochen) orientiert sich an den strukturellen Merkmalen, die Chernikova et al. (2020) im Rahmen ihrer Metastudie für die Förderung diagnostischer Kompetenz beschreiben. Konkret sind dies das Bereitstellen von Beispielen (providing examples), das Anbieten von prompts (providing prompts), die Zuweisung von Rollen (assigning roles) sowie die Anregung von Reflexion (inducing reflection phases). Zusätzlich ist die Anregung des stetigen Vergleichs von Lernprodukten mit Vergleichsobjekten Teil der Seminarkonzeption.

Das Kernanliegen in der konkreten Studie besteht darin, die Veränderung der epistemischen Aktivitäten durch das Seminar insgesamt zu untersuchen. Es war explizit kein Anliegen, die Wirkung einzelner Elemente der Seminarkonzeption zu evaluieren.

Im Folgenden beschreiben wir die Ausgestaltung der einzelnen Elemente der Seminarkonzeption und begründen diese anhand der empirischen Hinweise bezüglich ihrer Wirksamkeit zur Förderung diagnostischer Kompetenz (vgl. Chernikova et al. 2020).

##### 3.1.1 Beispiele bereitstellen (providing examples)

Die Arbeit im Seminar ist fallbasiert gestaltet. Grundlage sind vier verschiedene offene Lernangebote (Schütte 2008) aus dem Bereich Arithmetik, die bezogen auf zwei Perspektiven bearbeitet werden sollen: Im ersten Teil des Seminars werden die vier offenen Lernangebote durch die Studierenden selbst bearbeitet. Im zweiten Teil des Seminars sollen die Studierenden reale Bearbeitungen (Lernprodukte) von Schülerinnen und Schülern der Primarstufe zu den gleichen vier offenen Lernangeboten (vgl. dazu auch Besser et al. 2015) analysieren.

Offene Lernangebote sind nach Schütte (2008) mathematisch ergiebige Problem- und Aufgabenstellungen, die im Sinne natürlicher Differenzierung (Wittmann 1990; Krauthausen und Scherer 2014) auf unterschiedlichen Schwierigkeitsniveaus bearbeitet werden können. Bei der Bearbeitung offener Lernangebote werden drei verschiedene Kompetenzbereiche angeregt, da sie durch ihre Charakteristik nicht nur fachliches Grundwissen erfordern, sondern auch mathematische Handlungskompetenzen sowie kommunikative Kompetenzen (Rathgeb-Schnierer und Schütte 2011).



Zahlenketten werden gebildet indem die erste und die zweite Zahl addiert die dritte Zahl ergeben. Die zweite und die dritte Zahl ergeben zusammen die vierte Zahl usw.

Suchen Sie viele verschiedene Zahlenketten mit fünf Gliedern, die die Zielzahl 100 ergeben.

(Birnstengel-Höft und Feldhaus 2006)

**Abb. 2** Offenes Lernangebot Zahlenketten (mit der Aufgabenformulierung für Studierende)

Ein Beispiel der insgesamt vier offenen Lernangebote ist in Abb. 2 aufgeführt. Weitere offene Lernangebote waren sogenannte Reihenzahlen (Scherer und Steinbring 2004), Abbauzahlen (Hengartner et al. 2006) und Minustürme (Wittmann und Müller 1992). Eine Bearbeitung eines Schülers zu einem analogen offenen Lernangebot in Abb. 3 zu sehen.

### 3.1.2 Prompts anbieten (providing prompts)

Die erste Hälfte des Seminars (sieben Semesterwochen) ist für die eigene Beschäftigung der Studierenden mit den offenen Lernangeboten (wie in Abb. 2) vorgesehen. Die Studierenden sind explizit aufgefordert:

<p>Zielzahl 20:</p> <p>Aufgabenstellung für Schülerinnen und Schüler:</p> <p>So werden Zahlenketten gebildet: Wenn du die erste und die zweite Zahl addierst, bekommst du die dritte. Die zweite und die dritte Zahl ergeben zusammen die vierte Zahl.</p> <p>Suche möglichst viele Zahlenketten mit vier Kettenglieder. Die letzte Zahl der Kette soll die 20 sein.</p> <p>Aufgabenstellung im Test: Bitte analysieren Sie das Schülerdokument</p>	<p>Schülerdokument (3. Klasse):</p>
---	-------------------------------------

**Abb. 3** Beispiel für ein Fragebogenitem

- eine eigene Lösung zu den offenen Lernangeboten zu erzeugen, wobei die Lernangebote nacheinander behandelt werden. Die Lernangebote sind so komplex, dass auch Studierende des Lehramts in aller Regel keine umfassende Lösung erzeugen können. In dem in Abb. 2 gezeigten Lernangebot führen die Zahlenketten mit der Bildungsregel der Fibonacci-Folge mit beliebigen Startwerten auf stets in den ganzen Zahlen lösbare, lineare diophantische Gleichungen.
- bei der eigenen Lösung sowohl die Herangehensweise als auch die Ergebnisse zu reflektieren und zu dokumentieren sowie mögliche Variationen (Schupp 2002) der Problemstellung zu formulieren und diese zu untersuchen.
- die eigenen Lösungen im Tandem nach jeweils einer Woche der individuellen Beschäftigung zu vergleichen und schließlich auf der Basis entsprechender Dokumentationen in der Gesamtgruppe Herangehensweisen, Entdeckungen und Variationen vergleichend zu diskutieren. Die Ergebnisse der Bearbeitungen im Tandem in einer non-Präsenz-Phase sind wiederum Grundlage der Reflexion innerhalb der Seminarsitzungen in größeren Gruppen.

In der zweiten Hälfte des Semesters (sieben Semesterwochen) liegt der Schwerpunkt auf der Analyse von Lernprodukten von Schülerinnen und Schülern zu den vier offenen Lernangeboten (ein Beispiel ist in Abb. 3 enthalten). Die Schülerinnen und Schüler haben die gleiche Aufgabe wie die Studierenden bearbeitet, lediglich der Zahlenraum ist teilweise eingeschränkt wie es in Abb. 3 zu sehen ist. Für die multiperspektivische Analyse der Lernprodukte von Schülerinnen und Schülern sind zwei Modelle die Grundlage: Erstens, das von Rathgeb-Schnierer und Schütte (2011) vorgestellte Modell zur Erfassung mathematischer Kompetenzen bei Grundschülerinnen und Grundschulern, das vergleichbar mit Niss and Højgaard (2019) verschiedene mathematikspezifische Kompetenzbereiche ausdifferenziert und diese an die typischen Herausforderungen des Mathematiklernens von Kindern adaptiert. Zweitens, ein an den epistemischen Aktivitäten nach Chernikova et al. (2020) orientiertes Modell zur Analyse der Lernprodukte von Schülerinnen und Schülern. Beide Modelle werden den Studierenden als Prompt zur Verfügung gestellt. Zudem wird im Seminar ein Best-Practice-Beispiel für die Analyse einer Schülerlösung anhand beider Modelle vorgestellt.

Ein Prompt in diesem Teil des Seminars besteht konkret darin, in Anlehnung an Rathgeb-Schnierer und Schütte (2011) drei Bereiche mathematischer Kompetenzen von Grundschülerinnen und Grundschulern zu unterscheiden, und zwar

1. fachliches Grundwissen, das bei den von uns genutzten arithmetischen Lernangeboten die inhaltlichen Kompetenzen umfasst (z. B. Symbole verstehen, mathematische Hilfsmittel nutzen, Rechenoperationen ausführen, Strategien anwenden)
2. mathematische Handlungskompetenz, die sich auf den Lösungsprozess bezieht, z. B. auf das Vorgehen beim Problemlösen und
3. kommunikative Kompetenz, im Rahmen derer der Schwerpunkt auf dem Aufgabenverständnis, der Darstellung der Lösungswege und dem mathematischen Argumentieren liegt.

**Tab. 1** Modell zur Beurteilung von Lernprodukten

Kognitive Prozesse diagnostischen Denkens	Epistemische Aktivitäten
Wahrnehmen	<i>Manifeste Merkmale in einem Lernprodukt erkennen</i> (identifying a problem)
Interpretieren	<i>Manifeste Merkmale hinterfragen und Ursachen ergründen</i> (questioning) <i>Annahmen zu den Gedanken und Kompetenzen der Schülerinnen und Schüler bilden</i> (generating hypotheses) <i>Annahmen prüfen und anreichern mit weiteren Beispielen und Wissen zu Schülerprodukten</i> (questioning) <i>Sicherheit der Annahmen evaluieren und Maßnahmen zur Erhöhung der Sicherheit entwickeln</i> (evaluating hypotheses)

Ein weiterer Prompt (Tab. 1) besteht in der Anregung von kognitiven Prozessen des diagnostischen Denkens, also dem Wahrnehmen („perceiving“) und dem Interpretieren („interpreting“) (Leuders et al. 2018), im Hinblick auf das Beurteilen von Lernprodukten von Schülerinnen und Schülern zu unterschiedlichen epistemischen Aktivitäten (vgl. Chernikova et al. 2020). Das Wahrnehmen (perceiving) zielt auf die sichtbare Performanz in Lernprodukten ab. Es geht es darum, aufgabenbezogene manifeste Merkmale vor dem Hintergrund des fachdidaktischen Wissens zu erkennen und sachadäquat zu beschreiben (identifying a problem). Beim Interpretieren geht es darum, aus den manifesten Merkmalen Aussagen über die Kompetenzen der Schülerinnen und Schüler im Sinne kognitiver Leistungsdispositionen abzuleiten. Hierfür sind verschiedene Aktivitäten notwendig, wie das Ergründen der Ursachen für die manifesten Merkmale (questioning), die Entwicklung von Hypothesen (generating hypotheses) und deren Stützung (als Ausdruck eines questioning) sowie die Beurteilung der Evidenz der Hypothese (evaluating hypotheses). Die Entscheidungsfindung (drawing conclusions) wird nicht explizit thematisiert. Sie findet im Analyseprozess implizit dann statt, wenn Studierende auf Grundlage der wahrgenommenen manifesten Merkmale Hypothesen bezüglich vorhandener Kompetenzen formulieren und sich auf eine Diagnose festlegen.

Die Vorgehensweise im zweiten Teil des Seminars entspricht der Vorgehensweise im ersten Teil. Am Anfang steht die individuelle Auseinandersetzung der Studierenden mit einem selbst ausgewählten Lernprodukt von Schülerinnen und Schülern. Daran anschließend erfolgt der Vergleich der Analysen zu den Lernprodukten anhand der Kompetenzbereiche und der epistemischen Aktivitäten in unterschiedlichen Settings: erst im Tandem, dann in größeren Gruppen im Seminar und abschließend im Plenum.

### 3.1.3 Rollen zuweisen (assigning roles)

Im Verlauf des Seminars nehmen die Studierenden sowohl die Rolle der Lernenden ein, die ein offenes Lernangebot bearbeiten, als auch die Rolle der Lehrenden,

die zu den gleichen offenen Lernangeboten Lernprodukte von Schülerinnen und Schülern diagnostizieren. Damit gibt es innerhalb des Seminars die Möglichkeit, beide für die Lehramtsausbildung maßgeblichen Rollen einzunehmen (Chernikova et al. 2020). Die Übernahme der Rolle der Lehrenden ist weitgehend identisch zu den auf Mathematik bezogenen Studien zur Entwicklung der diagnostischen Kompetenz von Lehrkräften (vgl. Besser et al. 2015; Busch et al. 2015; Gold et al. 2013; Klug et al. 2016).

### 3.1.4 *Reflexion anregen (inducing reflection phases)*

Im gesamten Seminar spielen Reflexionsphasen im Sinne einer „guided reflection“ (Chernikova et al. 2020, S. 164) eine zentrale Rolle. Reflexionen beziehen sich dabei auf beide Rollen (Lernende, Lehrende) und werden individuell, im Tandem und in der Großgruppe initiiert. Bezogen auf die Analyse der Lernprodukte von Schülerinnen und Schülern wird insbesondere die Reflexion der epistemischen Aktivitäten im diagnostischen Prozess angeregt.

### 3.1.5 *Vergleichen anregen*

Ein wesentlicher Bestandteil des diagnostischen Prozesses ist das Vergleichen (Chernikova et al. 2020; Philipp 2018), das im Seminar auf verschiedene Arten initiiert wird. So werden die Lernprodukte der Schülerinnen und Schüler auf der Grundlage der eigenen Lösungen von Studierenden analysiert sowie die verschiedenen Vorgehensweisen verglichen und reflektiert. Zudem steuern die Prompts die Studierenden, die Lernprodukte der Schülerinnen und Schüler anhand eines festen Modells systematisch zu analysieren. Der fortwährende Vergleich als zentraler Bestandteil der Diagnostik lässt sich auch in das Lehr-Lernkonzept des Kontrastierens und Vergleichens einordnen. Dieses hat sich bereits in verschiedenen Bereichen als lernwirksam herausgestellt (Alfieri et al. 2013), wurde bisher aber kaum in der Lehramtsausbildung eingesetzt (Lipowsky et al. 2019). Das Kontrastieren und Vergleichen als Methode ist in dieser Studie allerdings dem diagnostischen Denken innewohnenden Vergleich untergeordnet.

## 3.2 **Fragestellung und Hypothesen**

Ausgehend von der theoretischen Konzeptualisierung diagnostischer Kompetenzen, den empirischen Erkenntnissen zu deren Förderung bei Lehrkräften sowie den empirischen Erkenntnissen zur konkreten Umsetzung eines Treatments zur Förderung diagnostischer Kompetenzen liegt unserer Untersuchung folgende zentrale Fragestellung zugrunde:

„Welche Bestandteile der diagnostischen Kompetenz von Lehrkräften – bezogen auf das adäquate Beurteilen von Lernprodukten von Schülerinnen und Schülern in drei Kompetenzbereichen sowie die damit verbundenen epistemischen Aktivitäten – lassen sich durch eine spezifische Intervention verändern?“

Obwohl die Fragestellung insbesondere auf die differenzierte Betrachtung der Förderung verschiedener Bestandteile der diagnostischen Kompetenz bezogen auf das Beurteilen von Lernprodukten zielt, lässt sich zunächst folgende allgemeine Hypothese aufstellen:

**H<sub>1</sub>** Das spezifische Treatment, das wesentliche Eigenschaften einer wirksamen Intervention zur Förderung diagnostischer Kompetenzen umfasst, ist für die Förderung der diagnostischen Kompetenzen angehender Lehrkräfte in Bezug auf die Beurteilung der Lernprodukte von Schülerinnen und Schülern zu offenen Lernangeboten der Arithmetik wirksam.

Mit dieser Hypothese gehen wir davon aus, dass sich insgesamt ein positiver Effekt in den epistemischen Aktivitäten im diagnostischen Prozess für die Treatmentgruppe im Vergleich zur Kontrollgruppe zeigt. Ein möglicher positiver Effekt hinsichtlich der Anzahl von Nennungen manifester Merkmale ist in Hypothese 2 formuliert.

**H<sub>2</sub>** Das spezifische Treatment hat einen positiven Effekt auf die Anzahl wahrgenommener manifester Merkmale beim Beurteilen der Lernprodukte von Schülerinnen und Schülern.

Neben der Anzahl der Nennungen manifester Merkmale gehen wir insbesondere davon aus, dass das Treatment auch auf die Vielfalt in der Nennung manifester Merkmale wirkt, da innerhalb des Seminars etwa verschiedene Kompetenzbereiche und deren unterschiedliche Facetten in Lernprodukten thematisiert sowie Unterschiede zu dieser Vielfalt in verschiedenen Lernprodukten reflektiert wurden. Diese Vielfalt bezeichnen wir in dieser Studie als „Breite“ der epistemischen Aktivitäten (vgl. Abschn. 3.4) und formulieren folgende Hypothese hinsichtlich dieser Breite (H<sub>3</sub>):

**H<sub>3</sub>** Das spezifische Treatment hat einen positiven Effekt auf die Breite wahrgenommener manifester Merkmale beim Beurteilen der Lernprodukte von Schülerinnen und Schülern.

Sowohl für die Anzahl und Breite der Hypothesen, die von den Studierenden möglicherweise auch unabhängig von der Nennung manifester Merkmale formuliert wurden, stellen wir für unsere Untersuchung folgende Hypothesen auf:

**H<sub>4.1</sub>** Das spezifische Treatment hat einen positiven Effekt auf die Anzahl der generierten Hypothesen zu Kompetenzen beim Beurteilen der Lernprodukte von Schülerinnen und Schülern.

**H<sub>4.2</sub>** Das spezifische Treatment hat einen positiven Effekt auf die Breite der generierten Hypothesen zu Kompetenzen beim Beurteilen der Lernprodukte von Schülerinnen und Schülern.

**Tab. 2** Übersicht zum Design

	Pre-Test: April 2020	Treatment: Sommersemester 2020	Post-Test: Juli 2020
Treatment-Gruppe ( $n = 57$ )	Fragebogen (1) zur Erfassung diagnostischer Kompetenz	Lernangebote im Bereich Arithmetik erkunden; Diagnose der Lernprodukte von Schülerinnen und Schülern	Fragebogen (2) zur Erfassung diagnostischer Kompetenz
Kontroll-Gruppe ( $n = 18$ )	Fragebogen (1) zur Erfassung diagnostischer Kompetenz	Interaktionsanalyse von ausgewählten Sequenzen aus dem Mathematikunterricht ohne spezifischen Bezug zur Arithmetik	Fragebogen (2) zur Erfassung diagnostischer Kompetenz

Die Formulierung von gestützten Hypothesen ist abhängig von der Nennung manifester Merkmale und der Nennung von Hypothesen. Für die Anzahl wie auch die Breite formulieren wir daher zwei von den von den Hypothesen  $H_2$  bzw.  $H_3$  und  $H_{4,1}$  bzw.  $H_{4,2}$  abhängige, weitere Hypothesen:

**H<sub>5,1</sub>** Das spezifische Treatment hat einen positiven Effekt auf die Anzahl der generierten gestützten Hypothesen zu Kompetenzen beim Beurteilen der Lernprodukte von Schülerinnen und Schülern.

**H<sub>5,2</sub>** Das spezifische Treatment hat einen positiven Effekt auf die Breite der generierten gestützten Hypothesen zu Kompetenzen beim Beurteilen der Lernprodukte von Schülerinnen und Schülern.

Über diese gerichteten Hypothesen hinaus untersuchen wir ergänzend und explorativ den beispielsweise von Heitzmann et al. (2018) postulierten Einfluss des Wissens der diagnostizierenden Personen auf die Diagnostik. Dieses Wissen fassen wir als fachwissenschaftliches und fachdidaktisches Wissen im Bereich der Arithmetik auf.

### 3.3 Design und Stichprobe

Im Sommersemester 2020 wurde eine Interventionsstudie mit Treatment- und Kontrollgruppe im Pre-Posttest-Design durchgeführt (Tab. 2). Die Treatment-Gruppe umfasste insgesamt 57 Studierende ( $n = 57$ ), die im Mittel als Gesamtgruppe 4,7 Semester absolviert haben. Sie hat inhaltlich in Seminaren in der in Abschn. 3.1. erläuterten Konzeption gearbeitet. Um den Einflussfaktor der Lehrperson zu minimieren, wurden die Seminare von drei Lehrenden im wöchentlich rollierenden Wechsel gehalten. Die Kontrollgruppe wurde aus einem inhaltsfremden didaktischen Seminar rekrutiert, das nicht im gleichen Modul angeboten wird, wie das der Treatment-Gruppe. Diese Festlegung verhinderte Überschneidungen von Treatmentgruppe und Kontrollgruppe. Dieses Seminar wurde von einem nicht am Projekt beteiligten Dozenten durchgeführt. Die Kontrollgruppe setzte sich aus 18 Studierenden ( $n = 18$ ) zusammen. Die Studierenden der Kontrollgruppe haben im Mittel 6,1 Semester absolviert.

Die Gruppengrößen wurden vorab mit G\*Power aus der Perspektive der Kontrollgruppe in einem Seminar mit ca. 25 Studierenden geprüft. Unter der Annahme einer geringen Korrelation von Pre- und Posttest ( $r = 0,3$ ) und dem üblichen  $\alpha$ -Fehler

von 0,05 sind bei einem mittleren (partielles  $\eta^2 > 0,06$ ) bis großen Effekt (partielles  $\eta^2 > 0,14$ ), bei zwei Gruppen und zwei Messzeitpunkten Stichprobengrößen ab 20 Studierenden vorzusehen, um mit einer Wahrscheinlichkeit von 80 % einen Effekt des Treatments zu entdecken (Rasch et al. 2010). Durch Drop-outs wurde der Richtwert von 20 Studierenden in der Kontrollgruppe leicht unterschritten.

Alle Studierenden der (Gesamt)-Stichprobe haben vor der Teilnahme am fachdidaktischen Seminar drei Vorlesungen besucht: je eine Fach- und Fachdidaktik verbindende Vorlesung zur Arithmetik und Geometrie sowie eine Vorlesung zur Diagnostik in der Grundschule. Ebenso haben sie ein Praxissemester absolviert. Mit der Verortung der Studie im vierten Semester ist sichergestellt, dass die Studierenden bereits die Möglichkeit hatten, grundlegende fachliche, fachdidaktische und diagnostische Kenntnisse zu erwerben. Diese stellen eine Voraussetzung für die Entwicklung diagnostischer Kompetenzen dar (Loibl et al. 2020; Heitzmann et al. 2018).

### 3.4 Messung diagnostischer Kompetenz

In dieser Studie wird die diagnostische Kompetenz angehender Lehrkräfte anhand ihrer schriftlichen Analyse der Lernprodukte von Schülerinnen und Schülern zu offenen Lernangeboten gemessen. Als Instrumente wurden für den Pre- und Posttest zwei verschiedene Fragebögen eingesetzt, die jeweils drei Items umfassen. Jedes Item enthält ein Dokument von Schülerinnen und Schülern zur Bearbeitung eines offenen Lernangebots zur Arithmetik und die Aufforderung, dieses zu beurteilen (vgl. Abb. 3).

Als Indikatoren für die diagnostische Kompetenz der Studierenden wurden die epistemischen Aktivitäten im Diagnoseprozess (Chernikova et al. 2020; Fischer et al. 2014) in der in Tab. 1 genannten Form verwendet und dabei explizit zwischen den genannten manifesten Merkmalen, den entwickelten Hypothesen und der Evaluation dieser Hypothesen unterschieden. Bei den Hypothesen wurde zudem differenziert, ob diese mit manifesten Merkmalen aus den Lernprodukten gestützt wurden. Weitere erfasste Möglichkeiten der Stützung waren die mehrfache Stützung durch manifeste Merkmale, ein Bezug zu didaktischer Literatur oder zu curricularem Wissen.

Ein Analysefokus in dieser Arbeit bezieht sich also auf die absolut skalierten Anzahlen (1) der manifesten Merkmale, (2) der gebildeten Hypothesen, (3) der durch Informationen gestützten Hypothesen, und (4) der Aussagen zur Evidenz der identifizierten Informationen. Um die Anzahlen epistemischer Aktivitäten zu messen, wurden die schriftlichen Analysen der Studierenden zunächst in Sinnabschnitte eingeteilt und anschließend bezogen auf die epistemischen Aktivitäten kodiert. Der Kodierung lag ein vorab entwickelter Leitfaden mit Ankerbeispielen zugrunde.

Weiterhin wurde kodiert, welchen Kompetenzbereich die Studierenden in ihren Analysen der Lernprodukte der Schülerinnen und Schüler adressierten, also fachliches Grundwissen, mathematische Handlungskompetenz oder kommunikative Kompetenz. Innerhalb dieser deduktiven Kodierung wurden zudem induktive Codes (vgl. Kuckartz 2012) entwickelt, die Facetten innerhalb der drei Kompetenzbereiche beschreiben: Etwa wurde innerhalb des fachlichen Grundwissens kodiert, ob in der Analyse der Studierenden Bezug auf korrekte Rechnungen, fehlerhafte Rechnungen,

**Tab. 3** Kompetenzbereiche, Kompetenzfacetten sowie beobachtete Anzahl epistemischer Aktivitäten

Kompetenzbereich	Fachliches Grundwissen	Handlungskompetenzen	Kommunikative Kompetenzen
Kompetenzfacetten	Korrektes Rechnen Rechenfehler Stellenwertverständnis Übergänge Zahlbegriff Rechentricks Zahldarstellung	Lösungsstrategie Aufgabenbeziehung Erfinden	Äußere Form Kommentare Aufgabenverständnis

den Umgang mit Stellenwerten oder auf Übergänge genommen wurde. Insgesamt konnten so sieben verschiedene induktive Kodes als Facetten des fachlichen Grundwissens und je drei Kodes als Facetten der mathematischen Handlungskompetenzen und der kommunikativen Kompetenzen entwickelt werden. Die tatsächlich entwickelten Kodes beschreiben wir im Abschnitt zu den Ergebnissen (Abschn. 4, vgl. auch Tab. 3). Die Kodierungen nahmen zwei Personen vor, die Reliabilitätsprüfung der Kodierungen erfolgte mit dem ICC. Für die Kodierungen der manifesten Merkmale und Hypothesen, die für jeden Kompetenzbereich zusammengefasst wurden, ergaben sich für alle Kompetenzbereiche sehr gute Reliabilitäten von mindestens 0,8 (Bortz und Döring 2016). Für die gestützten Hypothesen ergeben sich weiterhin hohe Reliabilitäten von mindestens 0,7 (Bortz und Döring 2016).

Aufgrund der induktiven Kodierung innerhalb der Kompetenzbereiche war es möglich, den zweiten Analysefokus zu verfolgen, nämlich neben der Anzahl der epistemischen Aktivitäten auch die Vielfalt bzw. Breite der epistemischen Aktivitäten zu messen. Die Breite einer epistemischen Aktivität haben wir durch die Anzahl der unterschiedlichen Facetten gemessen, die innerhalb der drei Kompetenzbereiche erkannt und genannt wurden. Bei der Analyse gehen wir dabei insbesondere auf die durchschnittliche Breite der epistemischen Aktivitäten in den jeweils drei Aufgaben des Pre- und Posttests ein.

Da auch die Abhängigkeit der diagnostischen Kompetenz vom fachlichen wie fachdidaktischen Wissen angenommen wird (z.B. Heitzmann et al. 2018), haben wir für unsere Studie einen einfachen Test zum arithmetischen Fachwissen und zum fachdidaktischen Wissen als explorativen Zusatz eingesetzt. Der Test zum Fachwissen ist in Kolter et al. (2018) beschrieben, der fachdidaktische Test ist eine primarstufenspezifische Adaption aus Baumert et al. (2009). Die Entwicklung der Tests ist nicht Gegenstand dieser Arbeit. Die Tests sind in einem open repository zugänglich (<https://osf.io/mzbgj/>). Der Fachwissenstest hat sich als akzeptabel reliabel erwiesen. Die 14 Items zeigen ein Cronbachs alpha von größer 0,7. Im Gegensatz dazu hat unser Test zum fachdidaktischen Wissen weniger gute Reliabilitätswerte. Daher haben wir eine Reduktion auf die Items vorgenommen, die den Bereich des Wissens über das mathematische Denken von Schülerinnen und Schülern abdecken (vgl. Baumert et al. 2009) und die einen Wert von Cronbachs alpha von ungefähr 0,6 gewährleistet haben. Aufgrund der geringen Reliabilitäten wird daher der Einfluss des Wissens auf die diagnostische Kompetenz in der Treatmentgruppe allein in einem explorativen Sinne betrachtet. Die Kontrollgruppe hat keinen Wissenstest ausgeführt.



Die Auswertungen basieren überwiegend auf Varianzanalysen mit Messwiederholung.

## 4 Ergebnisse

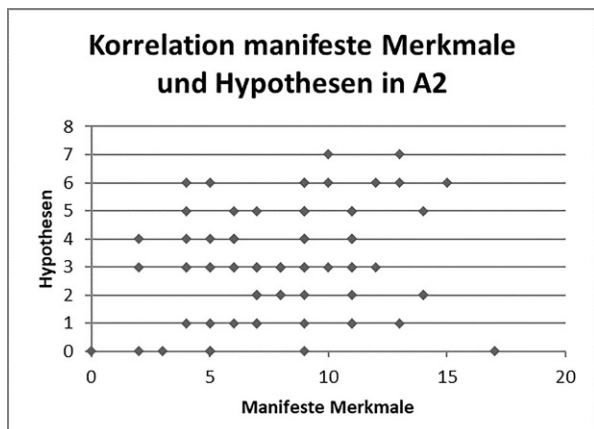
Wir gliedern die Ergebnisse in vier Abschnitte. Im ersten Abschnitt betrachten wir vorbereitend zwei deskriptive Ergebnisse. Im zweiten und dritten Abschnitt beziehen wir uns zunächst auf den Fokus der Arbeit, nämlich den Aspekt des diagnostischen Denkens (Leuders et al. 2018) und hier auf die Anzahlen und die Breite der gezeigten epistemischen Aktivitäten der Studierenden im Pre- und Posttest (Diagnoseprodukte der Studierenden). Weiterhin betrachten wir im vierten Abschnitt ergänzend und explorativ den Einfluss des Wissens der Studierenden auf die Anzahl und Breite der epistemischen Aktivitäten.

### 4.1 Deskriptive Ergebnisse

Die Kodierung der Diagnoseprodukte der Studierenden hat ergeben, dass sich die Studierenden innerhalb der drei Kompetenzbereiche – fachliches Grundwissen, mathematische Handlungskompetenzen und kommunikative Kompetenzen (vgl. Rathgeb-Schnierer und Schütte 2011 und Abschn. 3.1) – auf die in Tab. 3 dargestellten Facetten beziehen.

Bei der induktiven Kodierung wurde erfasst, ob die epistemischen Aktivitäten adäquat waren, also etwa die Nennung von manifesten Merkmalen und Hypothesen bezogen auf das analysierte Lernprodukt aus Forschertsicht passend sind. Als nicht adäquat wurden beispielsweise inhaltlich falsche Aussagen sowie Wiederholungen von bereits genannten Aspekten kodiert. Tatsächlich sind 6% der genannten manifesten Merkmale und Hypothesen als nicht adäquat eingeschätzt worden, wobei der Großteil der nicht adäquaten epistemischen Aktivitäten auf den Pretest entfallen. Auf deskriptiver Ebene zeigt sich wie erwartet, dass mehr manifeste Merkmale als Hypothesen und mehr Hypothesen als gestützte Hypothesen identifiziert werden

**Abb. 4** Zusammenhang der Anzahlen manifester Merkmale und Hypothesen



konnten. Dabei sind deutliche Unterschiede zwischen den Aufgaben (vgl. <https://osf.io/mzbgj/>) vorhanden.

In Abschn. 3.2 wurden die Hypothesen 4.1 und 4.2 hinsichtlich der Anzahl und Breite der Nennungen von Hypothesen als unabhängig von der Nennung manifester Merkmale aufgeführt. Tatsächlich zeigt sich deskriptiv, dass die Anzahlen von adäquat erkannten manifesten Merkmalen und geäußerten Hypothesen nur schwach zusammenhängen. Die Korrelationen liegen für die insgesamt sechs Aufgaben zwischen 0,03 (Aufgabe 1, Posttest) und 0,434 (Aufgabe 2, Pretest). Abb. 4 gibt für den Fall der höchsten Korrelation einen Eindruck für den statistisch zwar signifikant von 0 verschiedenen, aber dennoch diffusen Zusammenhang der Anzahl von Nennungen manifester Merkmale und Hypothesen.

## 4.2 Anzahl epistemischer Aktivitäten

Betrachtet man die Entwicklung des diagnostischen Denkens (Leuders et al. 2018) bezogen auf die Anzahl der epistemischen Aktivitäten in Treatment- und Kontrollgruppe, so ist die Anzahl der in den Lernprodukten wahrgenommenen manifesten Merkmale (alle drei Kompetenzbereiche) nicht signifikant durch das Treatment beeinflusst worden. Hier hat die ANOVA mit Messwertwiederholung, die auch bei den folgenden Untersuchungen eingesetzt wurde, einen knapp nicht-signifikanten Interaktionseffekt von Zeit \* Gruppe ergeben ( $F(1,73) = 3,52$ ;  $p = 0,065$ ;  $\eta^2 = 0,046$ ). Damit könnte Hypothese 2 höchstens hinsichtlich des Niveaus  $0,05 < p < 0,1$  bestätigt werden.

Die Entwicklung der Anzahl der Hypothesen zu Kompetenzen von Schülerinnen und Schülern in den drei Kompetenzbereichen ist dagegen in Treatmentgruppe und Kontrollgruppe unterschiedlich (Abb. 5). Der Interaktionseffekt von Zeit \* Gruppe ist hochsignifikant ( $F(1,73) = 11,43$ ;  $p < 0,01$ ). Die entsprechende Hypothese  $H_{4.1}$  ist also hinsichtlich des Ergebnisses der ANOVA bestätigt worden. Der Effekt (partielles  $\eta^2 = 0,138$ ) ist auf der Grenze zu einem großen Effekt (Cohen 2013). Die zugehörigen Werte für diese Untersuchung sind in Abb. 5 enthalten. In beiden deskriptiven Darstellungen ist die Überlegenheit der Kontrollgruppe im Pretest sichtbar, die im Posttest umgedreht ist. Allerdings ist der Unterschied zwischen Treatmentgruppe und Kontrollgruppe allein im Posttest nicht signifikant. Insofern scheint das Treatment zunächst nur ein Aufholen gegenüber der Kontrollgruppe zu ermöglichen.

	Gruppe	Mittelw.	Stdabw.	N
Pre	Kontroll	12,33	5,36	18
	Treatment	7,26	3,79	57
	Gesamt	8,48	4,71	75
Post	Kontroll	11,22	8,14	18
	Treatment	12,53	5,82	57
	Gesamt	12,21	6,42	75

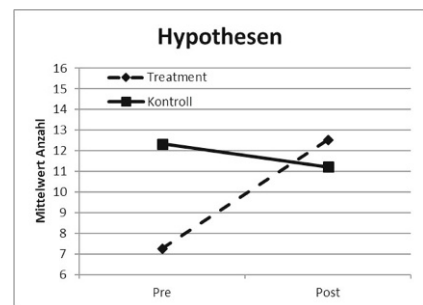
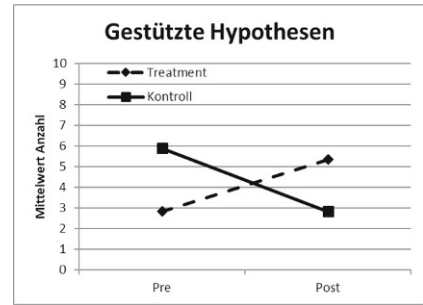


Abb. 5 Anzahl der Hypothesen in Pre- und Posttest

	Gruppe	Mittelw.	Stdabw.	N
Pre	Kontroll	5,89	4,31	18
	Treatment	2,83	2,18	57
	Gesamt	3,56	3,10	75
Post	Kontroll	2,83	1,38	18
	Treatment	5,35	4,61	57
	Gesamt	4,75	4,41	75



**Abb. 6** Anzahl der gestützten Hypothesen

Den Unterschied von Kontrollgruppe und Treatmentgruppe im Pretest, der in Abb. 5 wie auch einem Teil der weiteren Abbildungen sichtbar ist, nehmen wir weiter unten wieder auf.

Die gestützten Hypothesen, also jene, die beispielweise mit manifesten Merkmalen begründet wurden, sind stark vom Treatment beeinflusst worden (Abb. 6). Der Interaktionseffekt von Zeit \* Gruppe ist hochsignifikant ( $F(1,73) = 18,77$ ;  $p < 0,001$ ). Die zugehörige Hypothese  $H_{5,1}$  ist also bestätigt worden. Der Effekt ist groß ( $\eta^2 = 0,205$ ). Die zugehörigen Werte für diese Untersuchung sind in Abb. 6 enthalten. Im Vergleich zur vorangegangenen Untersuchung (Anzahl der Hypothesen) wird deutlich, dass in beiden Gruppen nur ein Teil der Hypothesen gestützt wird. Der Unterschied im Posttest von Treatmentgruppe und Kontrollgruppe ist signifikant ( $t\text{-Test}, t(72,94) = 3,642$ ;  $p < 0,001$ ,  $d = 0,6$ ).

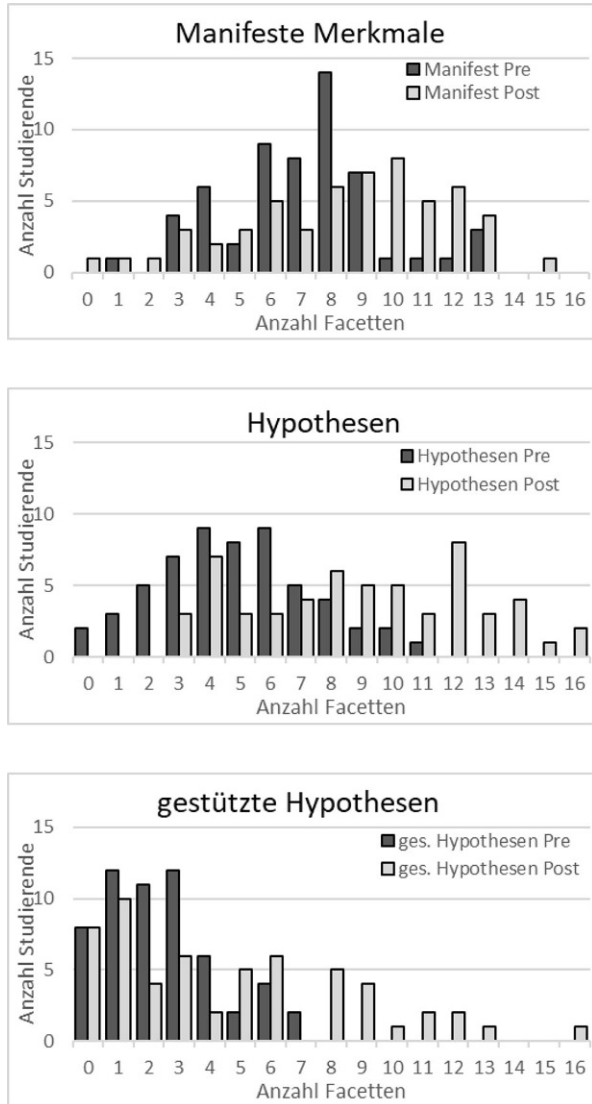
Weitere epistemische Aktivitäten, wie etwa die Evaluation der Belastbarkeit der Hypothesen, sind weder in der Treatmentgruppe noch der Kontrollgruppe in nennenswertem Maße vorhanden gewesen.

Die Pretestergebnisse der Kontrollgruppe sind für die hier betrachteten epistemischen Aktivitäten deutlich höher als diejenigen der Treatmentgruppe: Bezogen auf die Nennung von manifesten Merkmalen und Hypothesen werden durch das Treatment Unterschiede ausgeglichen. Bei der Nennung der gestützten Hypothesen holt die Treatmentgruppe dagegen nicht nur auf, sondern zeigt im Posttest eine signifikante Überlegenheit.

### 4.3 Breite epistemischer Aktivitäten in Pre- und Posttest

Als Breite der epistemischen Aktivitäten haben wir die Anzahl der verschiedenen, von den Studierenden genannten Kompetenz-Facetten pro epistemischer Aktivität modelliert (vgl. Abschn. 3). In Tab. 3 sind diese Kompetenzfacetten für die drei Kompetenzbereiche fachliches Grundwissen (7), Handlungskompetenzen (3) und kommunikative Kompetenzen (3) aufgeführt. Im Folgenden betrachten wir zunächst deskriptiv die Wirkung des Treatments auf die Breite der diagnostizierten Kompetenzfacetten in den Lernprodukten von Schülerinnen und Schülern für die jeweils drei verschiedenen Aufgaben im Pre- und Posttest. Abb. 7 fokussiert dabei auf die Verteilungen der Anzahlen von den Studierenden genannten verschiedenen Facetten im Pre- und Posttest jeweils für die epistemischen Aktivitäten der Nennung ma-

**Abb. 7** Verteilung der Breite der von den Studierenden diagnostizierten Kompetenzfacetten



nifester Merkmale, Hypothesen und gestützte Hypothesen. Die horizontale Achse beschreibt dabei die Anzahl verschiedener Facetten, die von den Studierenden mindestens einmal adressiert wurden. Durch die Darstellung der Verteilung wird hier sichtbar, dass sowohl bei den manifesten Merkmalen als auch bei den Hypothesen und gestützten Hypothesen die Breite, also die Anzahl der verschiedenen genannten Kompetenzfacetten, vom Pre- zum Posttest in der Treatmentgruppe zunimmt, ohne dass sich die Form der Verteilung substantiell zu ändern scheint. Dabei wird auch die Heterogenität der Studierenden zu beiden Testzeitpunkten sichtbar.

**Tab. 4** Explorative Analyse der Entwicklung der Breite von epistemischen Aktivitäten im Treatment

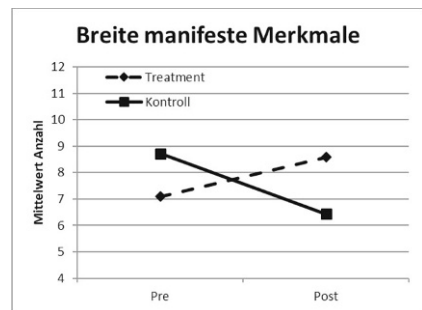
	Manifeste Merkmale	Hypothesen	Gestützte Hypothesen
Pretest: <i>m (sd)</i>	4,44 (1,25)	4,47 (1,71)	1,86 (1,44)
Posttest: <i>m (sd)</i>	5,91 (1,75)	5,47 (1,86)	2,71 (2,20)
Teststatistik	<i>t</i> (112)= 5,18	<i>t</i> (112)= 2,98	<i>t</i> (112)= 2,46
<i>p</i> und Cohens <i>d</i>	<i>p</i> <0,001; <i>d</i> =0,9	<i>p</i> =0,004; <i>d</i> =0,6	<i>p</i> =0,015; <i>d</i> =0,5

Obwohl die Aufgaben von Pre- und Posttest verschieden sind, führen wir die Unterschiede hinsichtlich der epistemischen Aktivitäten in der Breite auf das Treatment zurück. Explorativ haben wir die Unterschiede in der Breite von Pre- zu Posttest in der Treatmentgruppe mit einem t-Test analysiert. Dieser ist hinsichtlich der manifesten Merkmale, der Hypothesen und auch der gestützten Hypothesen hochsignifikant mit mittleren bis großen Effekten (Tab. 4).

Bezieht man die Breite der epistemischen Aktivitäten auf Treatment und Kontrollgruppe, so zeigt sich hinsichtlich der manifesten Merkmale ein hochsignifikanter Interaktionseffekt von Zeit \* Gruppe ( $F(1,73) = 17,61; p < 0,001$ ). Die zugehörige Hypothese  $H_3$  ist also bestätigt worden. Der Effekt ist groß ( $\eta^2 = 0,194$ ). Wiederum sind die Ergebnisse tabellarisch und grafisch in Abb. 8 dargestellt. Betrachtet man nur den Posttest, so ist der Unterschied der Gruppen signifikant (t-Test:  $t(31,78) = 2,55; p = 0,016, d = 0,6$ ).

Auch bei den nicht gestützten sowie den gestützten Hypothesen (Abb. 9) zeigen sich hochsignifikante Interaktionseffekte von Zeit \* Gruppe (Hypothesen:  $F(1,73) = 16,99; p < 0,001$ ; gestützte Hypothesen:  $F(1,73) = 10,74; p < 0,01$ ). Die zugehörigen Hypothese  $H_{4,2}$  und  $H_{5,2}$  sind also bestätigt worden. Der Effekt ist groß bzw. mittel ( $\eta^2 = 0,189$  bzw.  $\eta^2 = 0,128$ ). Betrachtet man allein den Posttest, so ist wiederum in beiden Fällen der Unterschied von Treatmentgruppe zur Kontrollgruppe signifikant. Für die Hypothesen ergibt sich mit einem t-Test  $t(35,84) = 2,82; p = 0,006; d = 0,7$ . Im Fall der gestützten Hypothesen ergibt sich  $t(69,28) = 3,13; p = 0,002; d = 0,6$ .

	Gruppe	Mittelw.	Stdabw.	N
Pre	Kontroll	8,72	2,40	18
	Treatment	7,02	2,39	57
	Gesamt	7,43	2,49	75
Post	Kontroll	6,44	3,01	18
	Treatment	8,58	3,39	57
	Gesamt	8,07	3,41	75



**Abb. 8** Breite der manifesten Merkmale

	Gruppe	Mittelw.	Stdabw.	N
Pre	Kontroll	6,56	2,75	18
	Treatment	4,86	2,49	57
	Gesamt	5,27	2,64	75
Post	Kontroll	6,61	2,87	18
	Treatment	8,95	3,64	57
	Gesamt	8,38	3,59	75

	Gruppe	Mittelw.	Stdabw.	N
Pre	Kontroll	2,83	1,25	18
	Treatment	1,72	1,46	57
	Gesamt	1,99	1,48	75
Post	Kontroll	1,72	1,23	18
	Treatment	3,32	3,18	57
	Gesamt	2,93	2,91	75

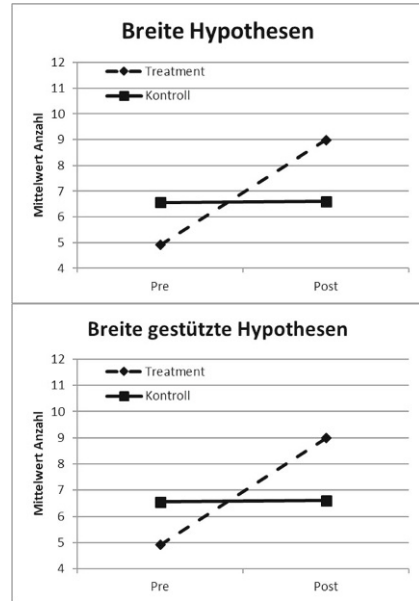


Abb. 9 Breite nicht gestützter und gestützter Hypothesen

#### 4.4 Explorative Analyse des Einflusses von Wissen auf die Veränderung der epistemischen Aktivitäten

Obwohl der primäre Fokus auf der Entwicklung des diagnostischen Denkens (Leuders et al. 2018) liegt, sollte der Einfluss des Fachwissens und des didaktischen Wissens auf die epistemischen Aktivitäten der Studierenden exploriert werden. Die folgenden Ergebnisse sind auf der Basis der insbesondere für den fachdidaktischen Test noch nicht zufriedenstellenden Charakteristiken zu interpretieren.

Eine bivariate Korrelation der summierten Testscores hat keinerlei signifikante Korrelationen zu den epistemischen Aktivitäten im Pretest ergeben. Diese ergeben sich erst zu den epistemischen Aktivitäten, die die Studierenden der Treatmentgruppe im Posttest gezeigt haben. In Tab. 5 sind nur die signifikanten Korrelationen der Ergebnisse im Fachwissenstest (links) und im fachdidaktischen Wissenstest (rechts) zu epistemischen Aktivitäten bei der Beurteilung von Schülerlösungen aufgeführt. Am fachwissenschaftlichen und am fachdidaktischen Test haben nicht alle Studierenden

Tab. 5 Korrelationen zwischen fachwissenschaftlichem und fachdidaktischem Wissen und den epistemischen Aktivitäten

(n = 40)	Fachwissen	(n = 47)	Fachdid. Wissen
Manifeste Merkmale (post)	0,343*	Hypothesen (post)	0,303*
Breite manifeste Merkmale (post)	0,361*	Gestützte Hypothesen (post)	0,321*
–	–	Breite Hypothesen (post)	0,288*

\* $p < 0.05$

der drei Treatment-Seminare teilgenommen. Die eingeschränkten Stichproben haben eine Größe von  $n=40$  Studierenden (Test zum Fachwissen) und  $n=47$  Studierenden (Test zum fachdidaktischen Wissen).

## 5 Diskussion

Die Förderung diagnostischer Kompetenz ist ein zentrales Ziel der Lehramtsausbildung, und die Konzeption und Evaluierung domänenspezifischer Interventionen stellt eine herausfordernde Aufgabe für die Forschung dar (Chernikova et al. 2020; Südkamp und Praetorius 2017). In diesem Kontext steht die dargestellte Studie. Diese fokussiert auf die Facette diagnostischer Kompetenz, die sich spezifisch auf das Beurteilen von Lernprodukten von Schülerinnen und Schülern bezieht. Zu dieser Facette wurde die Wirkung einer spezifischen Intervention auf das diagnostische Denken (Leuders et al. 2018) mit einem besonderen Augenmerk auf die Differenzierung von epistemischen Aktivitäten untersucht. Die Intervention selbst baut auf einem Bündel von Merkmalen auf, dessen Bestandteile sich in der Forschung zur Förderung diagnostischer Kompetenz bisher als wirksam zeigten. Ein zusätzlicher Schwerpunkt dieser Intervention lag auf dem Vergleich von Lernprodukten von Schülerinnen und Schülern.

Untersucht wurden die epistemischen Aktivitäten von Studierenden beim Analysieren der Lernprodukte von Schülerinnen und Schülern, die als Diagnoseprodukte in schriftlicher Form vorlagen. Für die Anzahl der genannten manifesten Merkmale hat sich durch das spezifische Treatment kein signifikanter Effekt ergeben. Wir erklären dieses Ergebnis damit, dass das Identifizieren von manifesten Merkmalen in Lernprodukten von Schülerinnen und Schülern auch über die hier analysierte Intervention hinaus Inhalt des Lehramtsstudiums ist. Insofern ist anzunehmen, dass Studierende auch ohne spezifische Schulung zur Diagnostik manifeste Merkmale in den Lösungen von Schülerinnen und Schülern in gleicher oder ähnlicher Weise wahrnehmen können wie geschulte Studierende. Das Erkennen von Oberflächenmerkmalen setzt (z. B. richtige oder falsche Lösungen von arithmetischen Aufgaben oder einfache Strategien) fachliches und fachdidaktisches Wissen voraus, das im Rahmen der Intervention nicht explizit geschult wurde, sondern vermutlich zu den Vorkenntnissen der Treatment- und Kontrollgruppe gehört. Bezogen auf das fachliche Wissen wird diese Hypothese auch durch die explorativen Ergebnisse dieser Studie unterstützt. So scheint fachliches Wissen mit der Anzahl der manifesten Merkmale zusammenzuhängen, die in Lernprodukten von Schülerinnen und Schülern erkannt werden. Dennoch ist bei der Anzahl der wahrgenommenen manifesten Merkmale ein nachweisbarer Effekt des Treatments auf dem Niveau  $0,05 < p < 0,1$  vorhanden. Dieser kann in der vorliegenden Studie auch als Aufholen von Treatmentgruppe zur im Pretest stärkeren Kontrollgruppe gewertet werden. Ob und inwieweit sich die Anzahl der in Lernprodukten von Schülerinnen und Schüler erkannter manifesten Merkmale durch eine Intervention zur Förderung diagnostischer Kompetenzen auch statistisch bedeutsam fördern lässt, müsste ein Folgeuntersuchung zeigen.

Sehr deutlich lässt sich dagegen erkennen, dass das Treatment mit einem großen Effekt auf die Breite der wahrgenommenen manifesten Merkmale wirkt. Gerade

die Breite der genannten manifesten Merkmale bezogen auf die hier als Maßstab verwendeten drei Kompetenzbereiche und deren Kompetenzfacetten (vgl. Tab. 3) ist ein Beleg für die Reichhaltigkeit eines Diagnoseprodukts. Wir interpretieren diese Breite als ein zentrales Kriterium von Diagnosekompetenz bezogen auf das diagnostische Denken (Leuders et al. 2018).

Das spezifische Treatment hat weiterhin erhebliche Auswirkung auf die epistemischen Aktivitäten der Interpretation. Bei der Anzahl der geäußerten Hypothesen können die Ergebnisse allerdings auch als ein Aufholen der Treatmentgruppe gegenüber der Kontrollgruppe gewertet werden (vgl. die Ergebnisse zu Hypothese 4.1 und Abb. 5). Dagegen unterscheiden sich die Anzahl der gestützten Hypothesen wie auch die Breite sowohl der Hypothesen als auch der gestützten Hypothesen bei einem überwiegend starken Effekt signifikant von der Kontrollgruppe. Auch bezogen auf die Anzahl gestützter Hypothesen sowie die Breite von Hypothesen und gestützten Hypothesen offenbart die Intervention in der Treatmentgruppe vom Prezum Posttest ohne Einschränkungen einen substantiellen Effekt. Wir führen die Entwicklung innerhalb der Treatmentgruppe und den Unterschied zur Kontrollgruppe auf das spezifische Treatment zurück. Die Fähigkeit, Hypothesen und insbesondere gestützte Hypothesen anhand von Lernprodukten von Schülerinnen und Schülern zu formulieren, betrachten wir als elaborierte Form der epistemischen Aktivitäten. Damit interpretieren wir diese Fähigkeit auch als wesentlichen Bestandteil diagnostischer Kompetenz bezogen auf das diagnostische Denken, der durch das spezifische Treatment gefördert wurde.

Aus diesen Haupt-Ergebnissen der Arbeit ergeben sich Antworten auf die eingangs gestellte Forschungsfrage nach den Bestandteilen diagnostischer Kompetenz von Lehrkräften bezogen auf die Beurteilung der Lernprodukte von Schülerinnen und Schülern, die sich durch eine spezifische Intervention fördern lassen: Die diagnostische Kompetenz von angehenden Lehrkräften im Fach Mathematik lässt sich zum Teil bezogen auf die Anzahl elaborierter epistemischer Aktivitäten und durchweg bezogen auf die Breite epistemischer Aktivitäten substantiell fördern. Epistemische Aktivitäten sind hier das Identifizieren manifesten Merkmale, die Hypothesenbildung und die Stützung der gebildeten Hypothesen.

Die Ergebnisse der Kontrollgruppe geben einen Hinweis darauf, dass auch das Studienalter einen Einfluss auf die diagnostische Kompetenz zu haben scheint, und zwar insbesondere bezogen auf die Anzahl identifizierter manifesten Merkmale in den Lernprodukten von Schülerinnen und Schülern. Unsere Ergebnisse zeigen aber, dass das Treatment Unterschiede ausgleicht und auch unabhängig von anfänglichen Unterschieden wirksam ist. Damit hat sich ein Bündel bereits vorab als vorteilhaft bekannter Eigenschaften einer Intervention zur Förderung diagnostischer Kompetenz bezogen auf das Beurteilen von Lernprodukten von Schülerinnen und Schülern als wirksam erwiesen (vgl. Besser et al. 2015; Busch et al. 2015; Gold et al. 2013; zusammenfassend Chernikova et al. 2020). Insbesondere ermöglicht dabei die Differenzierung des diagnostischen Denkens in epistemische Aktivitäten den vertieften Einblick in die Facetten der geförderten diagnostischen Kompetenzen.

Während die Bildung von Hypothesen in der Breite und die Stützung von Hypothesen in Anzahl und Breite durch die hier vorgestellte Intervention wirksam entwickelt werden konnte, ist dies bei der Evaluation der Hypothesen im Sinne ihrer



Belastbarkeit nicht der Fall gewesen. Entsprechende epistemische Aktivitäten konnten in unserer Stichprobe nur äußerst selten identifiziert werden. Wir interpretieren dieses Ergebnis so, dass die reflektierende Betrachtung nur durch eine deutlich spezifischere Thematisierung in einer Intervention entwickelt werden kann, wie sie im vorliegenden Fall nicht vorhanden war.

Als einen zur Analyse des diagnostischen Denkens nachgeordneten Aspekt haben wir auch den Einfluss des fachlichen und fachdidaktischen Wissens der Studierenden zur Arithmetik auf ihre diagnostische Kompetenz untersucht (z. B. Heitzmann et al. 2018). Die explorativ erzeugten Ergebnisse unterstützen die Annahme dieses Einflusses. Unerwartet war dabei, dass im Pretest die diagnostische Kompetenz in unserer Stichprobe nahezu unabhängig vom Wissen gewesen ist. An den Ergebnissen dieser Studie fällt zudem auf, dass die Facetten des fachlichen und fachdidaktischen Wissens offenbar auf unterschiedliche epistemische Aktivitäten wirken. Während das fachliche Wissen, das in dieser Arbeit reliabel gemessen werden konnte, mit der Anzahl und Breite der erkannten manifesten Merkmale in den Lernprodukten der Schülerinnen und Schüler zusammenhängt, korreliert das fachdidaktische Wissen eher mit der Entwicklung von Hypothesen. Hier ist allerdings als deutliche Limitation zu bemerken, dass einerseits die Messung des fachdidaktischen Wissens mit dem eigens entwickelten Test, zumindest in der vorliegenden Stichprobe, noch nicht hinreichend reliabel ist und andererseits das fachdidaktische Wissen auf das Wissen zu Schülerinnen und Schülern beschränkt ist. Dennoch geben unsere Ergebnisse in einem heuristischen Sinne einen Eindruck davon, dass eine Differenzierung der diagnostischen Kompetenz bezogen auf epistemische Aktivitäten auch eine Differenzierung des Einflusses des fachlichen und fachdidaktischen Wissens und damit eine Erweiterung der bisherigen Analysen zum Zusammenhang fachlichen und fachdidaktischen Wissens mit der diagnostischen Kompetenz ermöglichen kann (vgl. Chernikova et al. 2020; Heitzmann et al. 2018; Loibl et al. 2020).

Als Beitrag zur Schulung der diagnostischen Kompetenz von Lehrkräften wurde ein sehr spezifisches Interventionskonzept mit zwei inhaltlichen Schwerpunkten als Ganzes evaluiert. Als Folgefrage ergibt sich aus unserer Arbeit dabei, welche Bestandteile der Intervention maßgeblich für die Effektivität des Treatments verantwortlich waren. Ein Schwerpunkt lag auf der intensiven fachlichen Auseinandersetzung mit dem Lernstoff, indem die individuelle Erkundung der offenen Lernangebote angeregt und diese individuellen Erfahrungen durch Kontrastieren und Vergleichen mit den Lernprodukten anderer Studierender angereichert wurde. Diese Rollenzuweisung als Lernende (assigning roles) haben wir in unserer Intervention stark betont (vgl. Chernikova et al. 2020). Ein weiterer Schwerpunkt lag auf der Beurteilung ausgewählter Lernprodukte von Schülerinnen und Schülern, die von uns bereitgestellt wurden (providing examples), hierbei wurde den Studierenden die Rolle als Lehrende (assigning roles) zugewiesen (vgl. Chernikova et al. 2020). Innerhalb beider Schwerpunkte zeichnete sich die Intervention durch die Bereitstellung von Prompts (providing prompts) und die Anregung von Reflexionsphasen (inducing reflection phases) aus (vgl. Chernikova et al. 2020). Bei der Reflexion stand generell das Kontrastieren und Vergleichen von Lösungen im Mittelpunkt (Alfieri et al. 2013; Philipp 2018). Beide Bestandteile der Intervention führten schließlich dazu, auch Lernprodukte der Studierenden mit denen der Schülerinnen und Schüler zu vergleichen. Die

bewusste Anlage dieser Studie umfasst allerdings die Beschränkung, dass sich nicht differenziert ermitteln lässt, welchen Einfluss die einzelnen Bestandteile bzw. deren Kombination auf die Entwicklung diagnostischer Kompetenz hatten. Daher sollen in einer nachfolgenden Studie die Bestandteile systematisch variiert werden, in denen Lernprodukte von Studierenden, Lernprodukte von Schülerinnen und Schülern oder beide Sorten von Lernprodukten kontrastiert oder verglichen werden. Die oben genannten Aspekte zur wirksamen Entwicklung diagnostischer Kompetenz werden dabei konstant gehalten.

Im Gegensatz zu einer Differenzierung von Eigenschaften einer Intervention zur Förderung diagnostischer Kompetenz war es ein Kernanliegen unserer Arbeit, einen Beitrag zur differenzierten, multiperspektivischen Messung des Stands und der Entwicklung diagnostischer Kompetenz zu leisten. Die Messung nimmt dabei die epistemischen Aktivitäten im diagnostischen Prozess nach Chernikova et al. (2020) auf, die eine Erweiterung der Mikro-Prozesse nach Reinhold (2018) darstellen. Ebenso ermöglichen die Kompetenzbereiche nach Rathgeb-Schnierer und Schütte (2011) eine Differenzierung. Spezifisch sind hier die induktiv entwickelten Kategorien (Kompetenzfacetten) zur Subdomäne Arithmetik zu sehen. Da diese ähnlich zu den kognitiven Elementen nach Reinhold (2018) sind, gehen wir von einer Allgemeingültigkeit für die Analyse diagnostischer Prozesse beim Beurteilen von Lernprodukten in der Arithmetik aus. Über die Arithmetik hinaus sind dagegen andere bereichsspezifische Kategorien denkbar. Hier wäre demnach die Übertragbarkeit von der Arithmetik auf andere Bereiche der Mathematik in Folgestudien zu prüfen. Allgemeiner scheinen die Kompetenzbereiche für die Domäne Mathematik nutzbar zu sein, und zwar sowohl bezogen auf verschiedene Subdomänen (Arithmetik, Geometrie usw.) als auch auf das Alter der Schülerinnen und Schüler. So sind die Kompetenzbereiche beispielsweise auch für die Analysis anwendbar. Über die Mathematik hinaus ist schließlich der Bezug zu den epistemischen Aktivitäten bei der Beurteilung der Lösungen von Schülerinnen und Schülern möglich. Die differenzierende Messung ist zu der Anzahl epistemischer Aktivitäten, insbesondere aber zur Breite der epistemischen Aktivitäten möglich gewesen.

Eine Limitation der Studie ist die geringe Größe der Kontrollgruppe. Insofern sind die Ergebnisse und ihre Reliabilität auf der Basis der Stichprobe zu beurteilen. Tatsächlich wäre eine Replikation in einer vergrößerten Stichprobe sinnvoll. Eine weitere zukünftige Anforderung besteht in der Elaboration der Tests zum fachwissenschaftlichen und insbesondere zum fachdidaktischen Wissen, um den Einfluss des Wissens auf die diagnostische Kompetenz über eine Exploration hinaus differenziert messen zu können. Zudem besteht die auf der Basis der Daten nicht vollständig zu klärende Frage, warum die Leistungen der Kontrollgruppe vom Pre- zum Posttest unterschiedlich sind. Hier haben die unterschiedlichen Aufgaben im Vor- und im Nachtest trotz theoretisch-inhaltlicher Parallelisierung offensichtlich unterschiedliche Leistungen ergeben und insbesondere in der Kontrollgruppe einen scheinbaren Abfall der Leistungen bewirkt. Insofern ist weniger die Entwicklung einer Einzelgruppe als vielmehr das Verhältnis der beiden Gruppen zueinander interpretierbar. Neben der Unterschiedlichkeit der Aufgaben, ist der Bekanntheitsgrad der Aufgaben ein weiterer möglicher Grund für die Unterschiede in Vor- und Nachtest auf der Basis der deskriptiven Daten. So sind die geäußerten manifesten Merkmale zu

einem nachweislich bekannten offenen Lernangebot deutlich höher als zu anderen offenen Lernangeboten. Tatsächlich ist aber die Anzahl epistemischer Aktivitäten in allen offenen Lernangeboten sehr unterschiedlich und zudem sind die epistemischen Aktivitäten bezogen auf Wahrnehmung und Interpretation nur schwach korreliert. Der Einfluss der Aufgabenstellung auf die Art und den Umfang epistemischer Aktivitäten ist damit ein Desiderat dieser Studie. Ebenso könnten auch motivationale Variablen einen Einfluss auf das Verhalten der Studierenden im Vor- und Nachtest haben. Auch hier wäre es möglich, motivationale Variablen als Kovariaten zu erheben, insbesondere wenn ein Treatment zur Entwicklung diagnostischer Kompetenz langfristig angelegt ist.

Zusammengefasst zeigte sich, dass diagnostische Kompetenz bezogen auf die Fähigkeit, Lernprodukte von Schülerinnen und Schülern adäquat beurteilen zu können, bereits im Studium effektiv gefördert werden kann. Hier haben sich die bekannten effektiven Elemente einer Schulung, wie etwa Problemorientierung und Scaffolding (vgl. Chernikova et al. 2020) insbesondere bezogen auf die Breite epistemischer Aktivitäten aber auch auf die Anzahl elaborierter epistemischer Aktivitäten als wirksam erwiesen. Insbesondere hat diese Studie gezeigt, dass sich diagnostische Kompetenz mit einem multiperspektivischen Ansatz, der hier auf Kompetenzbereiche sowie epistemische Aktivitäten beim Prozess des Diagnostizierens fokussiert, differenziert messen lässt und dadurch einen vertieften Einblick in Unterschiede in der diagnostischen Kompetenz und deren Entwicklung zulässt.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**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

- Alfieri, L., Nokes-Malach, T.J., & Schunn, C.D. (2013). Learning through case comparisons: a meta-analytic review. *Educational Psychologist, 48*(2), 87–113. <https://doi.org/10.1080/00461520.2013.775712>.
- Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht, 57*(3), 175–193. <https://doi.org/10.2378/peu2010.art13d>.
- Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie, 23*(34), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>.

- 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*(5), 738–758.
- Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, *59*(5), 389–407. <https://doi.org/10.1177/0022487108324554>.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, *9*(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>.
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M., & Tsai, Y. M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente*. Materialien aus der Bildungsforschung, Bd. 83. Berlin: Max-Planck-Institut für Bildungsforschung.
- Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: a meta-analysis. *Review of Educational Research*, *87*(2), 309–344.
- Besser, M., Leiss, D., & Klieme, E. (2015). Wirkung von Lehrerfortbildungen auf Expertise von Lehrkräften zu formativem Assessment im kompetenzorientierten Mathematikunterricht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *47*(2), 110–122.
- Birnstengel-Höft, U., & Feldhaus, A. (2006). Gute Aufgaben in der Lehrerbildung und -weiterbildung. In S. Ruwisch & A. Peter-Koop (Hrsg.), *Gute Aufgaben im Mathematikunterricht der Grundschule* (S. 196–210). Offenburg: Miltenberger.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift für Psychologie*, *223*(1), 3–13.
- Bortz, J., & Döring, N. (2016). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- 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. Ergebnisse des Forschungsprogramms COACTIV* (S. 215–234). Münster: Waxmann.
- Busch, J., Barzel, B., & Leuders, T. (2015). Promoting secondary teachers' diagnostic competence with respect to functions: development of a scalable unit in continuous professional development. *ZDM*, *47*(1), 53–64. <https://doi.org/10.1007/s11858-014-0647-2>.
- 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*(1), 157–196. <https://doi.org/10.1007/s10648-019-09492-2>.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Hoboken: Taylor and Francis.
- Dübbelde, G. (2013). Diagnostische Kompetenzen angehender Biologie-Lehrkräfte im Bereich der naturwissenschaftlichen Erkenntnisgewinnung. Dissertation, Universität Kassel. <https://kobra.uni-kassel.de/handle/123456789/2013122044701>. Zugegriffen: 26.10.2022.
- van Es, E., & Sherin, M. (2002). Learning to notice: scaffolding new teachers' interpretations of classroom interactions. *Journal of Information Technology for Teacher Education*, *10*(4), 571–596.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Stijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, *4*, 28–45.
- Gold, B., Förster, S., & Holodyski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift für Pädagogische Psychologie*, *27*(3), 141–155. <https://doi.org/10.1024/1010-0652/a000100>.
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Heitzmann, N., Fischer, F., & Fischer, M. R. (2018). Worked examples with errors: When self-explanation prompts hinder learning of teachers' diagnostic competences on problem-based learning. *Instructional Science*, *46*(2), 245–271.
- Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2004). Vergleichsarbeiten als Werkzeug für die Verbesserung der diagnostischen Kompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119–144). Hohengehren: Schneider.
- Hengartner, E., Hirt, U., Wälti, B., & PRIMARSCHULTEAM (2006). *Lernumgebungen für Rechenschwache bis Hochbegabte*. Leipzig: Klett.
- Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approa-

- ches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>.
- Karing, C., Pfost, M., & Artelt, C. (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? *Journal for Educational Research Online*, 3(2), 119–147.
- Klieme, E., & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik: Zeitschrift für Erziehungswissenschaften (Sonderheft 8)* (S. 11–29).
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903.
- Klug, J., Gerich, M., & Schmitz, B. (2016). Can teachers' diagnostic competence be fostered through training and the use of a diary? *Journal for Educational Research Online*, 8(3), 184–206.
- Kolter, J., Blum, W., Bender, P., Biehler, R., Haase, J., Hochmuth, R., & Schukajlow, S. (2018). Zum Erwerb, zur Messung und zur Förderung studentischen (Fach-)Wissens in der Vorlesung „Arithmetik für die Grundschule“ – Ergebnisse aus dem KLIMAGS-Projekt. In R. Möller & R. Vogel (Hrsg.), *Konzepte und Studien zur Hochschuldidaktik und Lehrerbildung Mathematik. Innovative Konzepte für die Grundschullehrerausbildung im Fach Mathematik* (S. 95–121). Wiesbaden: Springer.
- Krauss, S., & Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittest für Mathematiklehrer/innen. *Journal für Mathematik-Didaktik*, 32(2), 233–251. <https://doi.org/10.1007/s13138-011-0029-z>.
- Krauthausen, G., & Scherer, P. (2014). *Natürliche Differenzierung im Mathematikunterricht. Konzepte und Praxisbeispiele aus der Grundschule*. Seelze: Kallmeyer.
- Kuckartz, U. (2012). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz Juventa.
- 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* (S. 3–31). Cham: Springer. [https://doi.org/10.1007/978-3-319-66327-2\\_1](https://doi.org/10.1007/978-3-319-66327-2_1).
- Lipowsky, F., Hess, M., Arend, J., Böhnert, A., Denn, A. K., Hirstein, A., & Rzejak, D. (2019). Lernen durch Kontrastieren und Vergleichen – Ein Forschungsüberblick zu wirkmächtigen Prinzipien eines verständnisorientierten und kognitiv aktivierenden Unterrichts. In U. Steffens & R. Messner (Hrsg.), *Grundlagen der Qualität von Schule. Unterrichtsqualität: Konzepte und Bilanzen gelingenden Lehrens und Lernens* (S. 373–402). Münster: Waxmann.
- Loibl, K., Leuders, T., & Dörfler, T. (2020). A framework for explaining teachers' diagnostic judgements by cognitive modeling (Diacom). *Teaching and Teacher Education*, 91, 103059. <https://doi.org/10.1016/j.tate.2020.103059>.
- Lorenz, C., & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 211–222.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 223–235.
- Nickerson, R. S. (1999). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5(2), 241–301. <https://doi.org/10.1037/1082-989x.5.2.241>.
- Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, 102(1), 9–28.
- 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.
- 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* (S. 109–127). Cham: Springer. [https://doi.org/10.1007/978-3-319-66327-2\\_6](https://doi.org/10.1007/978-3-319-66327-2_6).
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: a decade of research. *Educational Psychological Review*, 22, 271–296.
- Praetorius, A.-K., Rogh, W., & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkma-

- len der Unterrichtsqualität. *Unterrichtswissenschaft*, 48(3), 303–318. <https://doi.org/10.1007/s42010-020-00072-w>.
- Praetorius, A. K., Lipowsky, F., & Karst, K. (2012). Diagnostische Kompetenz von Lehrkräften. Aktueller Forschungsstand, unterrichtspraktische Umsetzbarkeit und Bedeutung für den Unterricht. In R. Lazarides & A. Ittel (Hrsg.), *Differenzierung im mathematisch-naturwissenschaftlichen Unterricht: Implikationen für Theorie und Praxis* (S. 115–146). Bad Heilbrunn: Julius Klinkhardt.
- Rasch, B., Frieze, M., Hofmann, W. J., & Naumann, E. (2010). *Quantitative Methoden: Einführung in die Statistik*. Berlin: Springer.
- Rathgeb-Schnierer, E., & Schütte, S. (2011). Mathematiklernen in der Grundschule. In G. Schönknecht (Hrsg.), *Lernen fördern: Deutsch, Mathematik, Englisch, Sachunterricht* (S. 143–208). Seelze: Kallmeyer/Klett.
- 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* (S. 129–148). Cham: Springer. [https://doi.org/10.1007/978-3-319-66327-2\\_7](https://doi.org/10.1007/978-3-319-66327-2_7).
- Rieu, A., Loibl, K., Leuders, T., & Herppich, S. (2020). Diagnostisches Urteilen als informationsverarbeitender Prozess – Wie nutzen Lehrkräfte ihr Wissen bei der Identifizierung und Gewichtung von Anforderungen in Aufgaben? *Unterrichtswissenschaft*, 61(5), 738. <https://doi.org/10.1007/s42010-020-00071-x>.
- Schäfer, S., & Seidel, T. (2015). Noticing und reasoning of teaching and learning components by pre-service teachers. *Journal for Educational Research Online*, 7(2), 34–58.
- Scherer, P., & Steinbring, H. (2004). Zahlen geschickt addieren. In G. N. Müller, H. Steinbring & E. C. Wittmann (Hrsg.), *Arithmetik als Prozess*. Seelze: Kallmeyer.
- Schoenfeld, A. H. (2011). Toward professional development for teachers grounded in a theory of decision making. *ZDM*, 43(4), 457–469.
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerinnen und Lehrerbildung*, 31(2), 154–165.
- Schupp, H. (2002). *Thema mit Variationen oder Aufgabenvariation im Mathematikunterricht*. Hildesheim: Franzbecker.
- Schütte, S. (2008). *Qualität im Mathematikunterricht der Grundschule sichern. Für eine zeitgemäße Unterrichts- und Aufgabenkultur*. München: Oldenbourg.
- Südkamp, A., & Praetorius, A.-K. (Hrsg.). (2017). *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen*. Münster, New York: Waxmann.
- Weinert, F. E. (1998). Vermittlung von Schlüsselqualifikationen. In S. Matalik & D. Schade (Hrsg.), *Entwicklungen in Aus- und Weiterbildung. Anforderungen, Ziele, Konzepte* (S. 23–43). Baden-Baden: Nomos.
- Weinert, F. E. (2000). Lehren und Lernen für die Zukunft-Ansprüche an das Lernen in der Schule. *Nachrichten der Gesellschaft zur Förderung Pädagogischer Forschung*, 2, 4–23.
- Weinsheimer, J. (2016). *Diagnostische Fähigkeiten von Mathematiklehrkräften bei der Begleitung von Lernprozessen im arithmetischen Anfangsunterricht. Theoretische Konzeptualisierung, empirische Erfassung und Analyse*. Hildesheim: Franzbecker.
- Wittmann, E. C. (1990). Wider die Flut der „bunten Hunde“ und der „grauen Päckchen“: Die Konzeption des aktiven-entdeckenden Lernens und des produktiven Übens. In E. C. Wittmann & G. N. Müller (Hrsg.), *Handbuch produktiver Rechenübungen* (Bd. 1, S. 152–166). Stuttgart, Düsseldorf: Klett.
- Wittmann, E. C., & Müller, G. N. (1992). *Vom halbschriftlichen zum schriftlichen Rechnen. Handbuch produktiver Rechenübungen, Bd. 2*. Stuttgart: Klett.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>.