Diagnostisches Urteilen als informationsverarbeitender Prozess – Wie nutzen Lehrkräfte ihr Wissen bei der Identifizierung und Gewichtung von Anforderungen in Aufgaben?

Andreas Rieu · Katharina Loibl · Timo Leuders · Stephanie Herppich

Online publiziert: 27. März 2020
© Der/die Autor(en) 2020

Zusammenfassung Bei der Auswahl von Aufgaben für einen adaptiven Unterricht schätzen Lehrkräfte deren fachliche Anforderungen ein. Im Rahmen eines Modells der Informationsverarbeitung wird angenommen, dass bei solchen diagnostischen Urteilen über Aufgaben auf Basis fachdidaktischen Wissens schwierigkeitsgenerierende Merkmale identifiziert und diese hinsichtlich ihres Einflusses auf die Aufgabenschwierigkeit gewichtet werden. Dabei erfolgt die Verarbeitung von Oberflächen- und Tiefenmerkmalen unterschiedlich schnell und ist daher abhängig von der verfügbaren Zeit. Anliegen der Studie ist es, diese Annahmen über die kognitiven Prozesse bei der Urteilsbildung zu prüfen, indem Aufgabenmerkmale systematisch variiert und fachdidaktisches Wissen als Personenmerkmal sowie Zeitdruck als Situationsmerkmal experimentell variiert werden. Zur Prüfung der Modellannahmen werden bei Lehramtsstudierenden ($N=175$) zwei Bedingungen verglichen: Einer Experimentalgruppe wird spezifisches fachdidaktisches Wissen über schwierigkeitsgenerierende Aufgabenmerkmale vermittelt. Ihre aufgabendiagnostischen Urteile werden durch paarweise Schwierigkeitsvergleiche erfasst und mit einer Kontrollgruppe verglichen – jeweils mit hoher und mit geringer Zeitrestriktion. Es zeigt sich, dass fachdidaktisches Wissen dazu führt, dass schwierigkeitsgenerierende Aufgabenmerkmale signifikant besser identifiziert und gewichtet werden, Zeitdruck hingegen hat signifikant negative Auswirkungen auf die Urteilsgüte. Die beschriebene Merkmalsvariation schlägt sich demnach hypothesenkonform in der

A. Rieu (✉)

Institut für Mathematische Bildung Freiburg, Pädagogische Hochschule Freiburg,
Kunzenweg 21, 79117 Freiburg, Deutschland
E-Mail: andreas.rieu@ph-freiburg.de

K. Loibl · T. Leuders

Pädagogische Hochschule Freiburg, Freiburg, Deutschland

S. Herppich

Universität Freiburg, Freiburg, Deutschland

Urteilsakkuratheit nieder, wobei die Prozesse der Identifizierung und der Gewichtung von schwierigkeitsgenerierenden Aufgabenmerkmalen unterschieden werden. Diese Ergebnisse erlauben Rückschlüsse auf die Bedeutung von spezifischem fachdidaktischem Wissen für diagnostische Urteile und geben damit Impulse für die Lehrkräfteaus- und -fortbildung. Die Unterscheidung der kognitiven Prozesse beim Einschätzen von Aufgabenschwierigkeit mit und ohne Zeitdruck legt nahe, dass sich Lehrkräfte hierüber während der Unterrichtsplanung (ohne Zeitdruck) und im Unterrichtsgeschehen (mit Zeitdruck) bewusst sein sollten, um reflektiert damit umgehen zu können.

Schlüsselwörter Diagnostische Kompetenz · Aufgabenschwierigkeit · Aufgabenmerkmale · Urteilsprozesse

Judging task difficulty: Effects of PCK and time pressure on cognitive processes

Abstract When choosing an assignment for an adaptive lesson, teachers judge the demands of tasks by activating their pedagogical content knowledge (*pck*). Based on a model of information processing, it is assumed that in such diagnostic judgments, teachers have to identify the relevant difficulty generating features of the task and weight them with regard to their influence on task difficulty. The processing of surface and structural features takes place at different speeds and therefore, depends on the available processing time. The study aims to examine these assumptions about the cognitive processes by experimentally varying *pck* and time pressure. In order to test the model assumptions, students in the experimental condition learn specific *pck* about difficulty generating task features. The accuracy of their judgments is compared with the judgments of a control group. In both groups ($N=175$), the participants judge pairs of tasks with regard to their difficulty twice—with high and with low time pressure. The results show that *pck* leads to significantly more accurate processing of the features. In addition, time pressure has a significant negative impact on the judgment accuracy. The described variation of the situation and the person characteristics, thus, influences differently the cognitive processes of identification and weighting of task features. These findings may have an impact on teacher education as they underline the importance of transmission of specific *pck* and of reflecting the effect of time pressure on the two cognitive processes in judging the task difficulty.

Keywords Assessment methods and tools · Mathematics task difficulty · Task features · Judgment processes

1 Einleitung

Bei der Auswahl von Aufgaben für einen adaptiven Unterricht schätzen Lehrkräfte die deren Anforderungen mittels ihrer diagnostischen Kompetenzen ein (McElvany et al. 2009; Ostermann et al. 2018; Schrader 1989; Schult und Lindner 2018).

Diagnostische Kompetenzen von Lehrkräften spielen eine grundlegende Rolle bei der Anpassung des Unterrichts an die Voraussetzungen der Lernenden und somit für den Lern- und Unterrichtserfolg (Black und Wiliam 1998; Schrader 2013). Sie sind daher auch Gegenstand von Forschung, z. B. mit Blick auf die Güte der abgegebenen Urteile (z. B. Schrader und Helmke 1987), auf die Lernendenförderung im Unterricht (z. B. Aufschnaiter et al. 2015) oder auf die Modellierung diagnostischen Handelns (z. B. Herppich et al. 2017).

Im Folgenden liegt der Fokus auf den kognitiven Prozessen bei der Genese der diagnostischen Urteile (Loibl et al. 2020). Diagnostische Urteilsprozesse wurden bislang zumeist in Fallstudien untersucht, die mit der Analysemethode des lauten Denkens Hinweise auf die ablaufenden Prozesse der Wahrnehmung, der Interpretation und des Entscheidens beim diagnostischen Urteilen liefern (z. B. Krolak-Schwerdt et al. 2009; Krolak-Schwerdt und Rummer 2005; Philipp 2018; Reinhold 2018; van Ophuysen 2006). Dabei liegt die Annahme zugrunde, dass die diagnostizierenden Personen Situationsmerkmale aufgreifen und diese zusammen mit ihrem Wissen bezüglich des Diagnosegegenstands und ihrem Wissen über die Schülerinnen und Schüler zu einem Urteil verarbeiten (Philipp 2018; Schrader 2013).

Es fehlen jedoch Studien, die die Prozesse der Informationsverarbeitung bei der Urteilsbildung theoretisch modellieren und experimentell prüfen (Loibl et al. 2020). Dazu sind laut Loibl et al. zu spezifizieren: 1. Die Charakteristik der *Diagnosesituation* (Aufgaben, Schülerlösungen, etc.), deren Informationsgehalt über Situationsmerkmale (z. B. die verwendeten Zahlen in Aufgaben) experimentell variierbar ist, 2. die Personencharakteristik als *Dispositionen* bei der diagnostizierenden Person, insbesondere deren erfassbares oder manipulierbares Wissen über die in den Situationen enthaltenen potenziell diagnoserelevanten Informationen und deren Interpretation, und 3. das diagnostische Denken als theoretische Beschreibung der angenommenen *Urteilsprozesse* in einem Modell der Informationsverarbeitung. Oft untersucht wurden bereits die korrelativen Zusammenhänge von Situations- oder Personencharakteristika mit der Urteilsakkuratheit (s. dazu die Metastudie von Südkamp et al. 2012). Untersuchungen, in welchen die Situations- und Personencharakteristik systematisch variiert werden, und bei denen alle drei genannten Bereiche in einem gemeinsamen Modell systematisch aufeinander bezogen sind und damit als Erklärung für diagnostische Urteile über informationsverarbeitende Prozesse dienen könnten, fehlen hingegen bislang.

Das Erkenntnisinteresse der vorliegenden Studie ist es, das Zusammenspiel von Situations- und Personencharakteristika im Rahmen des Urteilsprozesses theoretisch im Sinne einer Informationsverarbeitung zu modellieren und durch experimentelle Prüfung zu validieren. Hierzu wurde mit dem Thema „Textaufgaben zur Addition und Subtraktion von Brüchen“ ein Diagnosegegenstand ausgewählt, der eine systematische Spezifikationen von Merkmalen von Aufgaben als Konstituenten der Situationscharakteristik ermöglicht, der hinreichend reichhaltig für eine experimentelle Variation ist (s. Abschn. 4.1) und zu dem die Forschung zu Schülerkognitionen fundiert genug ist, um objektive diagnostische Urteile als Vergleich heranziehen zu können.

Die oben genannten drei Spezifikationsbereiche werden in der vorliegenden Studie folgendermaßen realisiert (vgl. Abb. 1):

Abb. 1 Einordnung der theoretischen Annahmen in das DiaCoM-Rahmenmodell nach Loibl et al. (2020)

1. **Situationscharakteristik:** Die Diagnosesituation umfasst textlich präsentierte Aufgaben zu denen die diagnostizierenden Personen die relative Schwierigkeit einschätzen müssen. Die Aufgaben fokussieren den Bereich Addition und Subtraktion von Brüchen. Relevante Aufgabenmerkmale sind verschiedene sprachliche und mathematische Anforderungen.
2. **Personencharakteristik:** Das relevante Wissen der Probandinnen und Probanden besteht in der Kenntnis der in der Situation variierenden schwierigkeitsgenerierenden Merkmale und ihren Einflüssen auf die Schwierigkeit der Aufgabe.
3. **Diagnostisches Denken:** Die diagnostizierenden Personen verarbeiten die Informationen der Situation und ihr Wissen über schwierigkeitsgenerierende Merkmale für deren Identifizierung und Verarbeitung und kommen so zu einem Urteil über die relative Schwierigkeit von Aufgabenpaaren.

Das hier angenommene Modell für einen Urteilsprozess kann durch systematische Variation des präsentierten Aufgabenmaterials und experimentelle Manipulation des Wissens, sowie durch Zeitrestriktion geprüft werden; die Trennung der beiden Urteilsprozesse der Identifizierung und der Verarbeitung von Aufgabenmerkmalen wird durch ein entsprechendes Versuchsdesign ermöglicht (s. Abschn. 4 Methode). Dabei wird im weiteren Verlauf unter *Identifizierung* das Erkennen von Aufgabenmerkmalen als schwierigkeitsgenerierend verstanden, während die Verarbeitung als *Gewichtung* unterschiedlicher schwierigkeitsgenerierender Aufgabenmerkmale operationalisiert wird. Abb. 1 ordnet die vorliegende Studie in das DiaCoM-Rahmenmodell (Explaining Teachers' Diagnostic Judgements by Cognitive Modeling, Loibl et al. 2020) ein.

2 Stand der Forschung

In den letzten Jahrzehnten wurden vielfältige Aspekte diagnostischer Kompetenzen untersucht und im Rahmen von Meta-Studien zusammengefasst (vgl. Hoge und Co-

ladarci 1989; Südkamp et al. 2012). Einzelne Studien untersuchten den Einfluss von situativen Informationen auf das Urteil von Lehrkräften (Helmke et al. 2003; Dünnebieer et al. 2009). So zeigen beispielsweise Dünnebieer et al. (2009) in ihrer Studie zu Ankereffekten den Einfluss von einzelnen Informationen (z. B. einer einzigen Klassenarbeitsnote) auf die situative Merkmalsverwertung im Urteilsprozess. Urteile von (angehenden) Lehrkräften werden demnach durch Situationscharakteristika beeinflusst.

Ebenso gibt es Befunde bezüglich der Auswirkungen der Personencharakteristika (z. B. unterschiedliches Wissen der befragten Personen) auf die gegebenen Urteile. Hierzu untersuchten McElvany et al. (2009) die Lehrkräfteeinschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern und berichten, dass das Wissen über schwierigkeitsgenerierende Aufgabenmerkmale in Zusammenhang mit der Unterschätzung der Schülerkompetenzen beim Lösen von Aufgaben steht. Im Bereich des mathematischen Problemlösens mit graphischen und numerischen Darstellungen von Funktionen weist die Studie von Ostermann et al. (2018) den Einfluss von zuvor vermitteltem fachdidaktischem Wissen zu relevanten Schülerfehlkonzepten auf die Urteilsakkuratheit bei der Einschätzung von Aufgabenschwierigkeiten nach.

Oudman et al. (2018) untersuchten die Effekte der Verfügbarkeit von Hinweisreizen über Lernende auf die Urteilsgenauigkeit von Lehrkräften, die einschätzen mussten, ob ihre eigenen Schülerinnen und Schüler Aufgaben zu Dezimalzahlen richtig oder falsch lösen. Dabei wurden die verfügbaren Informationen nach drei Bedingungen unterschieden: die Lehrkräfte erhielten entweder die Namen der Lernenden, die vorausgegangene Antwort einer Schülerin oder eines Schülers auf eine ähnliche Frage oder beide Informationen zusammen. Die Autorinnen und Autoren zeigen, dass die höchste Urteilsakkuratheit bei korrekten Antworten erzielt wird, wenn den Lehrkräften die Namen der Schülerinnen und Schülern zur Verfügung standen. Die besten Ergebnisse bei falschen Antworten gaben Lehrkräfte sobald sie nur die vorausgegangene Antwort der Lernenden erfuhren. Diese Studie liefert erste Einblicke in die Informationsverarbeitung von Lehrkräften beim Erstellen von pädagogischen Diagnosen. Dabei wurde allerdings nur die Situationscharakteristik experimentell variiert.

Zusammenfassend lässt sich feststellen, dass bisherige Studien meist *entweder* den Einfluss von Situationscharakteristika *oder* den Personencharakteristika (z. B. Wissen) auf diagnostische Urteile untersuchen. Da bisher nicht zugleich das Wissen der beteiligten Personen als auch situative Merkmale variiert oder manipuliert werden, geben diese Studien nur erste Hinweise auf zugrundeliegende informationsverarbeitende Prozesse.

Im Gegensatz dazu befasst sich die Studie von Hammer (2016) mit dem Zusammenspiel von Situationscharakteristika (Aufgaben) und Personencharakteristika (Wissensfacetten) im Rahmen von Diagnose. Sie legt den Schwerpunkt allerdings auf die Folgerungen der Lehrkräfte aus der Diagnose und weniger auf die Urteilsprozesse. So konnte die Autorin zeigen, dass das professionelle Wissen einer Lehrkraft und ihre Wahrnehmung von Aufgabenpotential wichtige Einflussfaktoren für die Aufgabenauswahl im Rahmen der Unterrichtsplanung sind.

Die nachfolgend vorgestellte Studie soll einen Beitrag zur Modellierung und Überprüfung kognitiver Urteilsprozesse bei der Diagnose leisten und befasst sich mit diagnostischen Urteilen zur Schwierigkeit von Aufgaben. Dabei werden die Einflüsse von Aufgabenmerkmalen, Zeitrestriktion und themenspezifischem fachdidaktischem Wissen bei einer Beschreibung des Urteilsprozesses als Informationsverarbeitung modellbezogen vorhergesagt und experimentell überprüft.

2.1 Diagnostisches Urteilen über Schülerinnen und Schüler und Aufgaben: die Rolle von Situationscharakteristik und Wissen

2.1.1 Situationscharakteristik

Einflüsse auf Informationsverarbeitung: Zeitdruck Der Einfluss von Zeitrestriktion auf gegebene Urteile wird von Müsseler und Rieger (2017) zusammengefasst. Demnach erhöht Zeitdruck die relativen kognitiven Kosten jedes Informationsverarbeitungsschrittes und führt dazu, dass weniger Informationen berücksichtigt werden. Der Zeitfaktor hat auch eine praktische Relevanz: Die Auswahl von Aufgaben für eine Lerngruppe oder einzelne Lernende wird als zentrales Element des unterrichtlichen Handelns betrachtet (Neubrand et al. 2011). Dabei kann dieser Auswahlprozess sowohl während der Unterrichtsvorbereitung ohne Zeitdruck als auch während des Unterrichtsgeschehens unter Zeitdruck erfolgen. Es ist anzunehmen, dass sich die dabei ablaufenden informationsverarbeitenden Prozesse unterscheiden und sich dies im Urteil niederschlägt. Es scheint daher sinnvoll, zwei Urteilsmodi abzugrenzen (Evans 2008; Müsseler und Rieger 2017): Die Auswahl einer in ihren Anforderungen passenden Aufgabe findet entweder ohne Zeitdruck in reflektierter Form statt (Böhmer et al. 2017, 2015) oder die Lehrkraft fällt ihr Urteil unter Zeitdruck und aktionsbezogen, also spontan und situationsabhängig (Lindmeier 2011).

Die kognitiven Prozesse bei der Beurteilung von Aufgabenschwierigkeit umfassen im hier angenommenen Modell die Identifizierung schwierigkeitsgenerierender Merkmale und die Gewichtung dieser Merkmale. Vor dem Hintergrund der von Müsseler und Rieger (2017) beschriebenen Studien ist davon auszugehen, dass die zur Verfügung stehende Zeit beide Prozesse der Informationsverarbeitung beeinflusst, da unter Zeitdruck weniger Informationen verarbeitet werden. Gleichzeitig werden die Urteilenden die von ihnen subjektiv als wichtig eingeschätzten Merkmale fokussieren und diese in ihrer Relevanz überschätzen (s. auch Rieskamp und Hoffrage 2008).

Einflüsse auf Informationsverarbeitung: schwierigkeitsgenerierende Merkmale

Ein Kriterium für die Einschätzung der Eignung einer Aufgabe für eine Lerngruppe ist die Identifizierung der zu ihrer Bearbeitung nötigen Fähigkeiten, welche sich an Merkmalen der Aufgabe erkennen lassen. In vereinfachter Form bildet sich dies in der eingeschätzten Schwierigkeit einer Aufgabe bzw. der relativen Schwierigkeit verschiedener Aufgaben ab (Ostermann et al. 2018). Die objektive Schwierigkeit einer Aufgabe kann anhand ihrer empirischen Lösungshäufigkeit bei Schülerinnen und Schülern beschrieben werden. Neubrand et al. (2002), sowie Krauss et al. (2008) gehen der Frage nach, welche Aufgabenmerkmale zur Schwierigkeit von Aufgaben aus den PISA-Erhebungen beitragen. Im curricular wichtigen Bereich der

Abb. 2 Beispiele für Bruchaufgaben mit bestimmten schwierigkeitsgenerierenden Merkmalen und deren Verortung an der Oberfläche (Bruchschwierigkeit und sprachlogische Komplexität) bzw. in der Tiefenstruktur (Operationsverständnis und Anzahl benötigter Grundvorstellung). Oberflächlich abzulesen sind die Bruchschwierigkeit (*links* gleichnamige und *rechts* ungleichnamige Brüche) und die sprachlogische Komplexität (*links*: höhere sprachliche Anforderungen durch Relativsätze). In der Tiefenstruktur der Aufgaben verbergen sich das Operationsverständnis (*rechts*: verborgene Subtraktion) und die Anzahl benötigter Grundvorstellung (*links*: zur Lösung muss zusätzlich ein ganzer Liter als Bruch dargestellt werden)

Bruchrechnung gibt es bereits Befunde zu schwierigkeitsgenerierenden Aufgabenmerkmalen (Padberg und Wartha 2017). Auf dieser Grundlage und in Anlehnung an das Klassifikationsschema von Jordan et al. (2006) werden im weiteren Verlauf vier Aufgabenmerkmale genauer betrachtet: Bruchschwierigkeit, Komplexität der Grundvorstellung (d.h. der zur Lösung nötigen mentalen Modelle), Operationsverständnis und sprachlogische Komplexität (für Details siehe Abschn. 4.2).

Bei der Einschätzung der Aufgabenschwierigkeit werden die Merkmale als Hinweisreize erkannt und verarbeitet (Ostermann et al. 2015; Rösler 2011). Dabei befinden sich diese Merkmale entweder leicht erkennbar an der Oberfläche des Textes oder müssen erst aus der Tiefenstruktur der Aufgabe erschlossen werden. So sind die in einer Aufgabe verwendeten Brüche (z.B. gleichnamige oder ungleichnamige Brüche) für die beurteilende Person auf den ersten Blick zu erkennen; die für die Lösung benötigte Grundvorstellung kann allerdings nicht direkt abgelesen werden. Die Relevanz dieser Unterscheidung wird durch Zeitdruck möglicherweise verstärkt: Während Oberflächenmerkmale schnell auf den ersten Blick erkannt werden, können Tiefenstrukturmerkmale nur beim intensiven Lesen des Textes und höchstens von geübten oder geschulten Beobachtern auf die Schnelle erkannt werden (Betsch et al. 2011; Krolak-Schwerdt et al. 2009). Dadurch werden Tiefenstrukturmerkmale unter Zeitdruck gegebenenfalls weniger berücksichtigt. Die konkrete Operationalisierung der Strukturebenen wird in Abb. 2 veranschaulicht.

2.1.2 Wissen als Personencharakteristik

Neben den genannten situativen Einflüssen auf die Informationsverarbeitung beeinflussen Persönlichkeitsmerkmale die kognitiven Urteilsprozesse. Um die relevanten Merkmale hinsichtlich der Aufgabenschwierigkeit zu erfassen, benötigen die diagnostizierenden Personen fachdidaktisches Wissen (*pedagogical content knowledge, pck*, Shulman 1986), genauer fachdidaktisches diagnostisches Wissen (*knowledge about content and students, kcs*, Ball et al. 2008). Die Vermittlung dieser Wissenskategorien ist hinsichtlich der Vorbereitung auf die zukünftigen beruflichen Herausfor-

derungen ein zentrales Anliegen der Ausbildung angehender Lehrkräfte (Ostermann et al. 2019; Thames und Ball 2010).

In den im Abschn. 2 genannten Studien (Dünnebier et al. 2009; Hammer 2016; McElvany et al. 2009; Oudman et al. 2018) zeigt sich, dass die Bereitstellung von Informationen oder die Vermittlung von fachdidaktischem Wissen, welches auf die Inhalte der Urteilsituation bezogen ist, die Urteilsakkuratheit verbessern. In Analysen, in denen fachdidaktisches Wissen eher breit und nicht bezogen auf die konkrete Diagnosesituation erfasst wird (Binder et al. 2018), ist dieser Zusammenhang weniger ausgeprägt. Diese Ergebnisse unterstützen die plausible Annahme, dass spezifisches fachdidaktisches Wissen als Personencharakteristik einen Einfluss auf die kognitiven Urteilsprozesse hat.

3 Fragestellung und Hypothesen

In der vorliegenden Studie soll untersucht werden, wie sich Situations- und Personencharakteristika auf die kognitiven Prozesse bei der Beurteilung der Aufgabenschwierigkeit auswirken. Basierend auf den dargestellten Befunden wird postuliert, dass Personen bei der Einschätzung von Aufgaben unter Zuhilfenahme ihres fachdidaktischen Wissens (1) deren schwierigkeitsgenerierende Merkmale identifizieren und (2) gewichten, in welchem Ausmaß die identifizierten Merkmale die Aufgabenschwierigkeit beeinflussen. Experimentell soll über die Manipulation der Situationscharakteristika – verschiedene Aufgabenmerkmale und Zeitrestriktion – und der Personencharakteristika – spezifisches fachdidaktisches Wissen – die Auswirkungen auf das diagnostische Verhalten untersucht und auf die kognitiven Prozesse des diagnostischen Denkens zurückgeführt werden.

Aus der obigen Annahme lassen sich die folgenden Voraussagen ableiten: Spezifisches fachdidaktisches Wissen über schwierigkeitsgenerierende Merkmale fördert die Identifizierung aller Merkmale und insbesondere jener in der Tiefenstruktur der Aufgabe. Für eine akkurate Gewichtung von Aufgabenmerkmalen wird zusätzlich Wissen darüber benötigt, wie relevant der Einfluss des jeweiligen Merkmals auf die Lösungshäufigkeit ist. Somit wird erwartet, dass

spezifisches fachdidaktisches Wissen dazu führt, dass schwierigkeitsgenerierende Merkmale besser identifiziert und gewichtet werden (Hypothese 1).

Zur Überprüfung dieser Hypothese wird einer Experimentalgruppe spezifisches fachdidaktisches Wissen vermittelt und die Urteile dieser Probanden über relative Aufgabenschwierigkeiten mit einer Kontrollgruppe verglichen. Durch diese Veränderung der Personencharakteristik kann deren Einfluss auf den Diagnoseprozess und das diagnostische Verhalten untersucht werden.

Darüber hinaus wird erwartet, dass

Zeitdruck zu einer Verarbeitungsdominanz von Oberflächenmerkmalen führt (vgl. Müsseler und Rieger 2017) und somit in beiden Gruppen bessere Urteilergebnisse ohne Zeitdruck erzielt werden können (Hypothese 2).

Dabei gilt den Tiefenstrukturmerkmalen – sie tragen zwar zur Aufgabenschwierigkeit bei, werden jedoch nicht leicht identifiziert und verarbeitet – ein besonderes Augenmerk. Es wird davon ausgegangen, dass die kognitive Verarbeitung von Tiefenstrukturmerkmalen sowohl spezifisches fachdidaktisches Wissen als auch ausreichend Zeit benötigt.

4 Methode

4.1 Spezifizierung der Diagnosesituation

Die Diagnosesituation wurde operationalisiert als Entscheidung zwischen zwei dargebotenen Mathematikaufgaben. Die teilnehmenden Personen sollten die ihrer Meinung nach für Schülerinnen und Schüler der 6. Klasse schwierigere Aufgaben auswählen. Dazu wurden zunächst Bruchrechenaufgaben mit den Rechenoperationen „Addieren“ und „Subtrahieren“ konzipiert. Bei der Auswahl der Aufgaben wurde Wert darauf gelegt, praxisrelevante, vorhersagestarke und weitgehend voneinander unabhängige Merkmale zu verwenden (Hartig und Frey 2012).

Inhaltlicher Rahmen Die Verortung des Forschungsprojektes im Stoffgebiet der Bruchrechnung beruht auf zweierlei Überlegungen: Aus didaktischer Sicht stellt die Bruchrechnung einen anspruchsvollen Lernstoff (Winter 1999) dar, welcher sowohl für das weitere mathematische Curriculum wie auch für die Lebenswelt bedeutsam ist (Padberg und Wartha 2017). Aus wissenschaftlicher Perspektive wurden im Bereich der Bruchrechnung bereits eine Vielzahl von Studien durchgeführt, die sich mit den besonderen Schwierigkeiten dieses Stoffgebietes befassen (Eichelmann et al. 2012; Padberg und Wartha 2017; Wartha 2007). Der Bereich der schwierigkeitsgenerierenden Merkmale und deren Verarbeitung im diagnostischen Urteilsprozess durch (angehende) Lehrkräfte ist in der Literatur allerdings noch lückenhaft.

4.2 Vorstudie: Entwicklung eines Aufgabensatzes und empirische Fundierung der schwierigkeitsgenerierenden Merkmale

In der oben genannten Literatur findet sich bereits eine Vielzahl von schwierigkeitsgenerierenden Merkmale und deren Erklärung (s. Abschn. 4.1). Für die Einschätzung der Aufgabenschwierigkeit benötigen (angehende) Lehrkräfte allerdings Informationen hinsichtlich der relativen Bedeutung der Aufgabenmerkmale. Um empirische Aussagen über die relative Bedeutung der unterschiedlichen Aufgabenmerkmale treffen zu können, wurde ein Aufgabenpool konzipiert, welcher systematisch die bereits erwähnten und im Folgenden detailliert beschriebenen Merkmale „Bruchschwierigkeit“, „Operationsverständnis“, „zusätzliche Grundvorstellung“ und „sprachlogische Komplexität“ variiert. Die relative Schwierigkeit der Aufgaben wurde mit Schülerinnen und Schüler der 6. und 7. Klasse erfasst.

Die verwendeten Aufgaben wurden in Anlehnung an häufig in Mathematikschulbüchern vorkommende Übungsaufgaben konzipiert. Die Aufgaben erfordern Übersetzungen zwischen dem realitätsentnommenen Kontext und der Mathematik.

<p>Saras Kuchen, dessen Teig sich aus $1\frac{2}{7}$ kg Butter und Mehl zusammensetzt, wiegt nun $2\frac{5}{7}$ kg. Wieviel Mehl hat sie verwendet?</p>	<p>Saras Kuchen, dessen Teig sich aus $1\frac{1}{5}$ kg Butter und Mehl zusammensetzt, wiegt nun $2\frac{1}{4}$ kg. Wieviel Mehl hat sie verwendet?</p>
---	---

Abb. 3 Variationen der Bruchschwierigkeit in gleichnamige und ungleichnamige Brüche

Diesbezüglich verorten sich die konzipierten Aufgaben auf niedrigem Niveau, also denjenigen „Übersetzungen, die unmittelbar ausgeführt werden können, da das Modell explizit gegeben ist [...]“ (Jordan et al. 2006, S. 34). Die Ausprägungen der genannten schwierigkeitsgenerierenden Merkmale werden im Folgenden detailliert dargestellt.

1. Bruchschwierigkeit Die Beziehung der Nenner der verwendeten Brüche sind nach Padberg und Wartha (2017) ein Faktor des Schwierigkeitsgrades bei der Lösung von Additions- und Subtraktionsaufgaben. Dabei werden im Folgenden gleichnamige und ungleichnamige Brüche unterschieden. In Einigen der konzipierten Aufgaben wurde die Bruchschwierigkeit kontrolliert variiert, während alle anderen Merkmale mit gleichbleibender Ausprägung gehalten wurden (s. Abb. 3).

2. Operationsverständnis Laut Padberg (1986) spielt die Verwendung der Rechenoperation Plus oder Minus nur eine geringe Rolle für die Lösungshäufigkeit einer Aufgabe. So können durchschnittlich 86 % aller Additionsaufgaben mit gleichnamigen Brüchen gelöst werden, während bei der Subtraktion 82 % richtig gelöst werden. Alleine die Verwendung der Rechenoperation „Minus“ stellt für Schülerinnen und Schüler demnach kaum eine zusätzliche Herausforderung dar. In Textaufgaben allerdings entsteht durch die notwendige Übersetzung vom Text der dargestellten Situation in eine mathematische Operation eine zusätzliche Schwierigkeit, welche sich in den semantischen Sprachstrukturen versteckt und von den Lernern entdeckt und anschließend als Rechenausdruck angewendet werden muss. Dabei verbirgt sich die Subtraktion oft hinter Begrifflichkeiten wie „auffüllen“ oder „ergänzen“ und ist nicht offensichtlich als Differenz zu erkennen. Die hier relevante Fähigkeit, Situationen oder Texte in passende Rechenoperationen zu übersetzen wird auch als Operationsverständnis bezeichnet (Schulz et al. 2019). Im Sinne des Operationsverständnisses wird die Addition explizit über Vorstellungen des Hinzufügens oder Zusammenfassens ausgedrückt, die Subtraktion beschreibt sowohl das Wegnehmen oder Vergleichen als auch stärker implizit das Ergänzen oder Auffüllen. Das für die Lösung der Aufgaben notwendige Operationsverständnis (explizite Addition, implizite Subtraktion) ist eine weitere Variable in den konzipierten Aufgaben und wird in Abb. 4 dargestellt.

3. Zusätzliche Grundvorstellung Auch bei der Anzahl zur Lösung benötigten Grundvorstellungen, also den Beziehungen zwischen Mathematik, Realität und in-

<p>Mehmet feiert Geburtstag und lädt dazu seine besten Freunde ein. Als besonderes Getränk möchte er $2\frac{1}{5}$ / Kirschsaft mit $1\frac{3}{5}$ / Bananensaft mischen. Welche Gesamtmenge an Kirsch-Bananensaft erhält er?</p>	<p>Mehmet feiert Geburtstag und lädt dazu seine besten Freunde ein. Als besonderes Getränk möchte er $2\frac{1}{5}$ / Kirsch-Bananensaft mischen. Er hat bereits $1\frac{3}{5}$ / Bananensaft verwendet, welche Menge an Kirschsaft benötigt er?</p>
--	--

Abb. 4 Veranschaulichung des sprachlichen Operationsverständnisses (*linke* Aufgaben „hinzufügen“ und *rechte* Aufgabe „auffüllen“)

dividuellen mentalen Strukturen (vgl. vom Hofe 1995), wurde eine Variation der schwierigkeitsgenerierenden Merkmale im untersuchten Aufgabensatz vorgenommen. So wurden sowohl Aufgaben konzipiert, die anhand einer elementaren Grundvorstellung gelöst werden können als auch Aufgaben, die eine „Kombination zweier nichtverwandter elementarer Grundvorstellungen“ (Jordan et al. 2006) benötigen. Konkret wird zum grundlegenden Kontext des Mischens von zwei Mengen und der Berechnung der Gesamtmenge in manchen Aufgaben die Verwendung einer zusätzlichen Grundvorstellung – die Differenz zum nächsten vollen Liter – zur Lösung notwendig (s. Abb. 5).

4. Sprachlogische Komplexität Auch rein sprachliche Merkmale können die Aufgabenschwierigkeit beeinflussen. So zeigte eine Analyse der Sprachverwendung in Prüfungsaufgaben (Gürsoy et al. 2013), dass sich sprachliche Hürden auf Wort-, Satz- und Textebene befinden und Einfluss auf die Lösungshäufigkeit der Aufgaben haben.

Im hier vorliegenden Projekt wurden die schwierigkeitsgenerierenden Merkmale der rein sprachlichen Art auf der Satzebene variiert und ausschließlich durch die Verwendung von Relativsätzen operationalisiert, die Bezüge zwischen den im Satz verwendeten Begriffen herstellen und damit den linearen Lesefluss unterbrechen (s. Abb. 6).

<p>Mehmet feiert Geburtstag und lädt dazu seine besten Freunde ein. Als besonderes Getränk möchte er $\frac{3}{5}$ / Kirschsaft mit $\frac{1}{5}$ / Bananensaft mischen. Welche Gesamtmenge an Kirsch-Bananensaft erhält er?</p>	<p>Mehmet feiert Geburtstag und lädt dazu seine besten Freunde ein. Als besonderes Getränk möchte er $\frac{3}{5}$ / Kirschsaft mit $\frac{1}{5}$ / Bananensaft mischen. Wieviel Wasser muss er noch hinzugeben, um einen vollen Liter zu erhalten?</p>
--	---

Abb. 5 Darstellung einer zusätzlichen Grundvorstellung (in der *linken* Aufgabe wird zur Lösung eine Grundvorstellung benötigt, *rechts* kommt eine zusätzliche Grundvorstellung hinzu)

<p>Daniels Crêpes-Teig setzt sich zunächst aus $\frac{3}{8}$ kg Mehl und $\frac{1}{8}$ kg Butter zusammen.</p> <p>Welches Gesamtgewicht stellt er her, bevor er dann noch Eier und Milch hinzugibt?</p>	<p>Daniel backt Crêpes, deren Teig sich zunächst aus $\frac{3}{8}$ kg Mehl und $\frac{1}{8}$ kg Butter zusammensetzt.</p> <p>Welches Gesamtgewicht stellt er her, bevor er dann noch Eier und Milch hinzugibt?</p>
---	--

Abb. 6 Syntaktische Schwierigkeit als Merkmal der sprachlogischen Komplexität (*rechte Aufgabe mit Verwendung eines Relativsatzes*)

Somit verorten sich in den vorliegenden Textaufgaben zwei Merkmale auf der sprachlichen Ebene: die sprachlogische Komplexität und das Operationsverständnis (s. oben). Während das Operationsverständnis durch die Durchdringung des Kontextes auf der semantischen Ebene der Aufgabe gekennzeichnet wird, sind die Schwierigkeiten der sprachlogischen Komplexität auf syntaktischer Ebene zu finden. Die für die Berechnung notwendige Codierung der Merkmalsausprägungen werden in Tab. 1 dargestellt, wobei die einfachere Variante mit „0“, die schwierigere Ausprägung mit „1“ codiert wurde.

4.2.1 Ergebnisse der Vorstudie

Die erstellten Aufgaben wurden von Schülerinnen und Schülern der 6. und 7. Schulstufe ($N=203$) an verschiedenen Realschulen in Baden-Württemberg gelöst. Aufgrund der Verwendung zweier sprachlicher Merkmale (Operationsverständnis und sprachlogische Komplexität) erscheint eine detaillierte Angabe zur Stichprobe hinsichtlich ihrer sprachlichen Voraussetzungen relevant: 37 % der für die Studie befragten Schülerinnen und Schüler geben an, im Elternhaus eine andere Sprache als Deutsch zu sprechen. Diese Ergebnisse decken sich mit Daten des statistischen Bundesamtes (Destatis) im Mikrozensus 2018, nach welchem 41 % der Schülerinnen und Schüler an allgemeinbildenden Realschulen einen Migrationshintergrund haben (Statistisches Bundesamt 2019). Die erhobenen Sprachleistungen sind daher repräsentativ für eine durchschnittliche Lerngruppe einer Realschule.

Anhand der Lösungshäufigkeiten der einzelnen Aufgaben konnte die Bedeutung der verwendeten und unterschiedlich ausgeprägten schwierigkeitsgenerierenden Merkmale für die Aufgabenschwierigkeit bestimmt werden. Die Lösungshäufigkeit wurde als Kriterium auf die vier unabhängigen Merkmalsausprägungen regrediert (s. Tab. 2). In den verwendeten Aufgaben wird die empirische Aufgabenschwierigkeit hauptsächlich von der zusätzlichen Grundvorstellung ($\beta = -0,720$) und der Bruchschwierigkeit ($\beta = -0,663$) bestimmt, während das Operationsverständnis ($\beta = -0,512$) und die sprachlogische Komplexität ($\beta = -0,161$) einen geringeren bzw. keinen gerichteten Zusammenhang mit der Aufgabenschwierigkeit aufweisen. Bei den hier verwendeten Aufgaben hängen also vor allem das Tiefenstrukturmerkmal der zusätzlichen Grundvorstellung und das Oberflächenmerkmal der Bruchschwie-

Tab. 1 Merkmalsausprägung und Codierung für Regression bei verwendeten Aufgaben

Merkmalsausprägungen		
Beispielaufgabe	Merkmalsausprägung	Code
<p> Meimet feiert Geburtstag und lädt dazu seine besten Freunde ein. Als besonderes Getränk möchte er $\frac{3}{5}$ l Kirschsaf mit $\frac{1}{5}$ l Bananensaft mischen. Welche Gesamtmenge an Kirschs-Bananensaft erhält er? </p>	Bruchschwierigkeit: „gleichnamig“	0
	Operationsverständnis: „hinzufragen“	0
	Grundvorstellung: „einfach“	0
	Sprachlogische Komplexität: „einfach“	0
<p> Saras Kuchen, dessen Teig sich aus $\frac{2}{5}$ kg Butter und Mehl zusammensetzt, wiegt nun $\frac{5}{7}$ kg. Wieviel Mehl hat sie verwendet? </p>	Bruchschwierigkeit: „ungleichnamig“	1
	Operationsverständnis: „auffüllen“	1
	Grundvorstellung: „einfach“	0
	Sprachlogische Komplexität: „schwierig“	1
<p> Silke bereitet sich nach der Schule ein Mischgetränk zu, welches sich aus $2\frac{1}{6}$ l Orangen- und $1\frac{2}{5}$ l Zitronensaft zusammensetzt. Wieviel Wasser benötigt sie, um die fehlende Differenz zum nächsten vollen Liter mit Wasser aufzufüllen? </p>	Bruchschwierigkeit: „ungleichnamig“	1
	Operationsverständnis: „hinzufragen“	0
	Grundvorstellung: „kombiniert“	1
	Sprachlogische Komplexität: „schwierig“	1

rigkeit mit der Lösungshäufigkeit zusammen. Ebenfalls gibt es einen Zusammenhang mit dem Tiefenstrukturmerkmal des Operationsverständnisses.

Die Modellaufklärung ist mit einem korrigierten $R^2=0,84$ sehr zufriedenstellend.

Wie bereits erwähnt, werden sowohl das Operationsverständnis wie auch die sprachlogische Komplexität durch die verwendete Sprache ausgedrückt. Schwierigkeiten, die bei Schülerinnen und Schülern auftreten und auf sprachliche Merkmale zurückzuführen sind, manifestieren sich sowohl auf syntaktischer wie auch

Tab. 2 Regressionsgewichte für abhängige Variable Lösungshäufigkeit

Koeffizienten			
Modell	Standardisierter Koeffizient β	T-Wert	Signifikanz
Zusätzliche Grundvorstellung	-0,720	-6,431	$\leq 0,001$
Bruchschwierigkeit	-0,663	-6,838	$\leq 0,001$
Operationsverständnis	-0,512	-4,575	0,001
Sprachlogische Komplexität	-0,161	-1,661	0,121

auf semantischer Ebene. Eine getrennte Betrachtung der Ergebnisse von mehrsprachigen Kindern bestätigt die Annahme, dass der sprachliche Hintergrund Einfluss auf die mathematischen Leistungen hat: Die Schülerinnen und Schüler, die eine andere Sprache als Deutsch als ihre Erstsprache angaben, lösten alle vorgelegten Aufgaben weniger häufiger ($M=0,302$, $SD=0,227$) als die Kinder mit Deutsch als Erstsprache ($M=0,421$, $SD=0,227$). Die relative Relevanz der schwierigkeitsgenerierenden Merkmale ändert sich dabei nicht. Auch bei dieser Gruppe wird die empirische Aufgabenschwierigkeit hauptsächlich von der zusätzlichen Grundvorstellung ($\beta=-0,763$) und der Bruchschwierigkeit ($\beta=-0,562$) bestimmt, während die sprachlichen Merkmale Operationsverständnis ($\beta=-0,418$) und sprachlogische Komplexität ($\beta=-0,201$) einen geringeren Zusammenhang mit der Aufgabenschwierigkeit aufweisen.

4.3 Hauptstudie: Untersuchung der Urteilsprozesse anhand der Einflüsse auf die Diagnosegenauigkeit

Teilnehmende Personen Die teilnehmenden Personen waren Studierende am Anfang ihres Lehramtsstudiums für das Fach Mathematik. Die Datenerhebung wurde in zwei Parallelkursen derselben, aber zu unterschiedlichen Zeiten angebotenen Lehrveranstaltung durchgeführt. Im Rahmen des stark strukturierten modularisierten Aufbaus des Lehramtsstudiums am Standort Freiburg, ist davon auszugehen, dass die Studierenden vor der Erhebung noch keine fachdidaktischen Veranstaltungen besucht haben, in denen sie ein Vorwissen zur Anforderungen der Bruchrechnung erworben haben könnten. Ein gesamter Kurs wurde zufällig der Experimentalgruppe zugeordnet, wobei die Studierenden vor der Erhebung über die Studie informiert wurde und der Verwendung ihrer Daten widersprechen konnten. Abb. 7 verdeutlicht schematisch den Ablauf der Datenerhebung.

In die Analysen mit Messwiederholung gingen nur die Daten der Probandinnen und Probanden ein, die an beiden Messzeitpunkten anwesend waren. Die sich ergebenden Gruppenzusammensetzungen zeigt Tab. 3. Zur Erfassung möglicher didaktischer Lerngelegenheiten wurden die persönlichen Praxis- und Nachhilfeeferfahrungen der Teilnehmenden erhoben. Die Inhalte des in den ersten Studiensemestern abge-

Abb. 7 Ablaufplan der Intervention mit Warte-Kontrollgruppen-Design

Tab. 3 Übersicht über Vergleich der Teilnehmergruppen und Ergebnisse der χ^2 -, T- und Mann-Whitney-U-Tests

Vergleich der teilnehmenden Personen nach Messzeitpunkten			
	Kontrollgruppe	Experimentalgruppe	Gruppenvergleich KG – EG
N	18	49	–
Anteil weiblich	94,4 %	91,8 %	$\chi^2(1) = 0,130, p = 0,719$
Mittleres Alter	23,2 Jahre (SD = 3,5)	22,4 Jahre (SD = 3,7)	$T(65) = 0,731, p = 0,467$
Praxiserfahrung	0,2 Wochen (SD = 0,5)	0,6 Wochen (SD = 1,2)	$U = 1189,50, Z = -0,494, p = 0,621$
Nachhilfeerfahrung	19,7 h (SD = 36,5)	6,3 h (SD = 24,4)	$U = 1205,50, Z = -0,323, p = 0,746$

legte Praktikums dient der allgemeindidaktisch orientierten Unterrichtsbeobachtung und vermittelt dabei kaum fachdidaktisches Wissen. Die erhobenen Daten können daher nur distale Hinweise auf fachdidaktische Lerngelegenheiten geben. Von einer expliziten Messung des fachdidaktischen Wissens wurde abgesehen, da die einschlägig bekannten Tests breit angelegt sind und nur indirekte Informationen über die Kenntnis von schwierigkeitsgenerierenden Aufgabenmerkmalen in der Bruchrechnung liefern. Der Diagnosetest mit den Vergleichen zweier Aufgaben konnte wiederum nicht als Prätest eingesetzt werden, da diese Lerngelegenheit die Unterschiede zwischen den Gruppen möglicherweise reduziert hätte.

Die teilnehmenden Personen der beiden Gruppen unterschieden sich nicht signifikant bezüglich der soziodemographischen Daten. Im Bereich ihrer Vorerfahrungen verfügte die Experimentalgruppe deskriptiv über mehr Praxiserfahrung, die Kontrollgruppe mehr Nachhilfeerfahrung (siehe Tab. 4). Da die vorliegenden Daten der Praxis- und Nachhilfeerfahrung nicht normalverteilt sind (Kolmogorov-Smirnov-Test, $p < 0,001$), wurde ein Mann-Whitney-U-Test durchgeführt, um zu überprüfen, ob dieser Gruppenunterschied signifikant wird. Hierbei ergaben sich keine signifikanten Unterschiede zwischen den Gruppen.

Design Der Studie liegt ein zweifaktorielles (2×2) Design mit dem Zwischensubjektfaktor spezifisches fachdidaktisches Wissen (Experimentalgruppe vs. Kontrollgruppe) und dem Innersubjektfaktor verfügbare Zeit (mit Zeitdruck vs. ohne Zeitdruck) zugrunde. Um eine möglichst typische und realitätsnahe Situation hinsichtlich der Vorbereitung bzw. während des Unterrichts abzubilden, wurde den Teilnehmenden mehrere Vergleiche von zwei ähnlichen Aufgaben vorgelegt. Dabei wurden sie gebeten zu entscheiden, welche der beiden Aufgaben das höhere Schwierigkeitsniveau für durchschnittliche Realschülerinnen und -schüler der 6. Klasse besitzt. Dieser Aufgabenvergleich vermeidet die in diesem Kontext den teilnehmenden Personen häufig gestellte (vgl. Ostermann et al. 2015), aber sehr schwierige Frage nach der prozentualen Lösungshäufigkeit einer Aufgabe oder der Erstellung einer Rangordnung der Aufgaben. Somit kann die Antwortvalidität gesteigert werden.

Erfassung der Urteile zur Trennung der kognitiven Prozesse Zur getrennten Erfassung der beiden kognitiven Urteilsprozesse der Identifizierung und der Gewich-

Tab. 4 Übersicht über die durchschnittlichen prozentualen Anteile richtiger Entscheidungen der Teilnehmenden an beiden Messzeitpunkten (beide Gruppen in beiden Urteilsmodi aufgelistet nach Identifizierung und Gewichtung)

	Gesamtübersicht Gruppenvergleich nach Zeitrestriktion und kognitiven Prozessen			
	M (SD) Identifizierung – unter Zeitdruck	M (SD) Identifizierung – ohne Zeitdruck	M (SD) Gewichtung – unter Zeitdruck	M (SD) Gewichtung – ohne Zeitdruck
Kontrollgruppe	0,590 (0,233)	0,667 (0,250)	0,512 (0,210)	0,635 (0,165)
Experimentalgruppe	0,725 (0,182)	0,776 (0,155)	0,633 (0,179)	0,717 (0,178)

<p>Mehmet feiert Geburtstag und lädt dazu seine besten Freunde ein. Als besonderes Getränk möchte er $\frac{3}{5}$ / Kirschsaft mit $\frac{1}{5}$ / Bananensaft mischen. Welche Gesamtmenge an Kirsch-Bananensaft erhält er?</p>	<p>An einem heißen Sommertag plant die Klasse einen Ausflug ins Museum. Tom stellt dafür $\frac{3}{5}$ / Apfelsaftschorle her. $\frac{1}{5}$ / davon ist Mineralwasser. Wieviel Apfelsaft hat er verwendet?</p>
--	---

Abb. 8 Beispiel für konsistenten Aufgabenvergleich, Unterschied nur beim Operationsverständnis (*links* „hinzufügen“, *rechts* „auffüllen“)

tung von Aufgabenmerkmalen, wurden den Teilnehmenden zwei Typen von Aufgabenvergleichen vorgelegt. Anhand des Urteilsverhaltens kann zurückgeschlossen werden, welche Prozesse stattfanden. Diese werden in den Abb. 8 und 9 veranschaulicht:

1. Es wurden konsistente Paarvergleiche vorgelegt, die sich in ihrer Schwierigkeit nur hinsichtlich eines Merkmals unterscheiden. Sobald die Teilnehmenden dieses Merkmal identifizieren, können sie ihre Entscheidung treffen.
2. Darüber hinaus wurden inkonsistente Aufgabenpaare gebildet, welche sich in zwei Aufgabenmerkmalen unterscheiden, wobei die Variation der Aufgabenmerkmale gegenläufig ist. Das bedeutet, wenn die Ausprägung des ersten Merkmals bei der rechten Aufgabe schwierigkeitsgenerierend ist, so ist die Ausprägung des zweiten Merkmals bei der linken Aufgabe schwierigkeitsgenerierend. Bei diesen Paaren müssen beide Merkmale identifiziert werden und dann die Gewichtung zugunsten des relevanteren Merkmals vorgenommen werden. Somit werden die kognitiven Prozesse der Identifizierung der Aufgabenmerkmale über konsistente Aufgabenpaare und die der Gewichtung über inkonsistente Paare getrennt erfasst (Sherman et al. 2014).

Variation der Personencharakteristik Auf personaler Ebene soll spezifisches Wissen das akkurate Identifizieren und Gewichten der Aufgabenmerkmale ermöglichen. In einer ca. 30-minütigen Intervention wurde eine Woche vor der

<p>Mehmet feiert Geburtstag und lädt dazu seine besten Freunde ein. Als besonderes Getränk möchte er $\frac{3}{5}$ / Kirschsaft mit $\frac{1}{4}$ / Bananensaft mischen. Welche Gesamtmenge an Kirsch-Bananensaft erhält er?</p>	<p>An einem heißen Sommertag plant die Klasse einen Ausflug ins Museum. Tom stellt dafür $\frac{3}{5}$ / Apfelsaftschorle her. $\frac{1}{5}$ / davon ist Mineralwasser. Wieviel Apfelsaft hat er verwendet?</p>
--	---

Abb. 9 Beispiel für inkonsistenten Aufgabenvergleich, Unterschiede gegenläufig in Bruchschwierigkeit (*links* ungleichnamige Brüche, *rechts* gleichnamige Brüche) und Operationsverständnis (*links* „hinzufügen“, *rechts* „auffüllen“)

ersten Erhebung spezifisches, fachdidaktisches Wissen an die Experimentalgruppe vermittelt: Zunächst wurde das Konzept der schwierigkeitsgenerierenden Merkmale (Aufgabenmerkmale, die Auswirkungen auf die Lösungshäufigkeit einer Aufgabe haben) erklärt und anhand der COACTIV-Klassifizierung einige Merkmale aufgezählt und definiert. Im zweiten Schritt wurden die Ergebnisse der empirischen Vorstudie bekannt gegeben (s. Abschn. 4.2.1 Ergebnisse der Vorstudie), wobei nicht die konkreten Regressionskoeffizienten, sondern lediglich die Merkmalshierarchie besprochen wurden (Grundvorstellung > Bruchschwierigkeit > Operationsverständnis > sprachlogische Komplexität). Die vermittelten Inhalte wurden im dritten Schritt konkretisiert, indem beispielhaft Aufgaben hinsichtlich ihrer Schwierigkeit beurteilt wurden. Dabei wurden nicht die Aufgaben der Erhebung verwendet. Die Intervention fokussierte somit spezifisches, fachdidaktisches Wissen, welches nahe an der Erhebung liegt, ohne diese identisch abzubilden.

Variation der Situationscharakteristik Die beiden Urteilsmodi wurden variiert, indem beiden Versuchsgruppen 15 Aufgabenvergleiche zur Einschätzung unter Zeitdruck und ohne Zeitdruck vorgelegt wurden. Unter Zeitdruck verschwand der Aufgabenvergleich nach der berechneten durchschnittlichen Lesezeit (Musch und Rösler 2011) von der Präsentationsfläche und die Teilnehmenden mussten anschließend schnell ihr Urteil fällen bevor der nächste Aufgabenvergleich präsentiert wurde. Nach einer weiteren Woche erfolgte die Einschätzung ohne Zeitdruck, zu welcher den teilnehmenden Personen dieselben Aufgabenvergleiche (in veränderter Reihenfolge) in schriftlicher Form vorlagen.

Auswertung Die Identifizierung bzw. Gewichtung der schwierigkeitsgenerierenden Merkmale wurde darüber operationalisiert, welche Aufgabe die Teilnehmenden im jeweiligen Paarvergleich als schwieriger auswählten. Zur Definition der richtigen Antwort wurde die vermittelte Rangreihe (Grundvorstellung > Bruchschwierigkeit > Operationsverständnis > sprachlogische Komplexität) zu Grunde gelegt. Nicht gegebene Antworten wurden als falsch gewertet. Anhand der durchschnittlichen richtigen Antworten der Probandinnen und Probanden bei den sechs konsistenten Aufgabenvergleichen (s. oben) kann auf die Identifizierung der Aufgabenmerkmale rückgeschlossen werden. Die durchschnittlichen richtigen Antworten bei den inkonsistenten Aufgabenvergleichen erlauben Rückschlüsse auf die Gewichtung der schwierigkeitsgenerierenden Merkmale.

Zur Beantwortung der ersten Forschungsfrage

Beeinflusst fachdidaktisches Wissen zu Aufgabenmerkmalen und deren Einfluss auf die Lösungshäufigkeit die Identifizierung und Gewichtung von Aufgabenmerkmalen im Beurteilungsprozess?

werden die Mittelwerte der Experimental- und der Kontrollgruppe in den Bereichen der Identifizierung und der Gewichtung von Aufgabenmerkmalen mit und ohne Zeitdruck untersucht. Darüber hinaus wird die unterschiedliche Identifizierung von Tiefenstrukturmerkmalen betrachtet. Bezogen auf die Gewichtung können die Tie-

fenstrukturmerkmale nicht isoliert werden, da in den inkonsistenten Aufgabenpaaren Oberflächen- und Tiefenstrukturmerkmale gemeinsam variiert werden.

Bezüglich der zweiten Forschungsfrage

Wirkt sich Zeitdruck auf die Identifizierung und Gewichtung von Aufgabenmerkmalen aus?

wird der Innersubjekteffekt (mit und ohne Zeitdruck) der Teilnehmenden untersucht. Der Effekt wird über alle Teilnehmenden und für die beiden Gruppen getrennt bestimmt. Auch hier wird die Identifizierung von Tiefenstrukturmerkmalen getrennt betrachtet.

Das verwendete Signifikanzniveau ist 5 %, die gerichteten Hypothesen wurden einseitig getestet und die Effektstärke wurde über das partielle Eta-quadrat ermittelt, wobei Werte $>0,01$ einem kleinen, Werte zwischen 0,06 und 0,14 einem mittleren und Werte $>0,14$ einem großen Effekt entsprechen (Cohen 2013). Alle Analysen wurden mit dem Programm SPSS (Version 24) durchgeführt.

5 Ergebnisse

Tab. 4 gibt eine Übersicht der richtig eingeschätzten Aufgabenvergleiche aller Personen, die an beiden Messzeitpunkten teilgenommen haben. Dabei zeigt sich, dass die Experimentalgruppe in allen Bereichen akkuratere Ergebnisse bei der Schwierigkeitseinschätzung der Aufgaben erzielte:

5.1 Effekte von Zeitdruck und fachdidaktischem Wissen auf die Identifizierung schwierigkeitsgenerierender Merkmale

Zur Untersuchung der Beeinflussung von spezifischem fachdidaktischem Wissen und der Zeitrestriktion hinsichtlich des Urteilsprozess „Identifizierung“ (bei konsistenten Paarvergleichen) wurde eine ANOVA mit Messwiederholung mit dem Zwischensubjektfaktor *Bedingung* (*spezifisches fachdidaktisches Wissen*) und dem Innersubjektfaktor *Zeitdruck* gerechnet. Beim Haupteffekt des Zwischensubjektfaktors *Bedingung* zeigte sich ein signifikanter Unterschied ($F(1,66)=7,177$, $p=0,009$, $\eta^2=0,098$), das vermittelte spezifische fachdidaktische Wissen wirkte sich (über beide Urteilsmodi hinweg) mit einer mittleren Effektstärke positiv auf den durchschnittlichen Anteil an identifizierten Aufgabenmerkmalen aus (Hypothese 1). Auch der Haupteffekt des Innersubjektfaktors *Zeitdruck* wurde signifikant ($F(1,66)=5,600$, $p=0,021$, $\eta^2=0,078$), *Zeitdruck* reduzierte mit mittlerer Effektstärke über beide Bedingungen hinweg die Identifizierung der schwierigkeitsgenerierenden Merkmale (Hypothese 2).

Die Interaktion zwischen *Zeitdruck* und *Bedingung* zeigte keinen signifikanten Unterschied ($F(1,66)=0,217$, $p=0,643$, $\eta^2=0,003$). Beide Gruppen wurden demnach gleichermaßen vom *Zeitdruck* beeinträchtigt.

Tab. 5 Durchschnittlicher prozentualer Anteil der akkuraten Identifizierung der Tiefenstrukturmerkmale bei konsistenten Aufgabenvergleichen

	M (SD)	M (SD)
	Identifizierung Tiefenstrukturmerkmale – unter Zeitdruck	Identifizierung Tiefenstrukturmerkmale – ohne Zeitdruck
Kontrollgruppe	0,570 (0,275)	0,717 (0,225)
Experimentalgruppe	0,614 (0,257)	0,651 (0,225)

Unterscheidung zwischen Tiefenstruktur- und Oberflächenmerkmalen Tab. 5 zeigt das Urteilsverhalten der Gruppen bei der Identifizierung der Tiefenstrukturmerkmale.

Der Haupteffekt des Innersubjektfaktors Zeitdrucks zeigte, dass sich der mittlere Anteil an identifizierten Tiefenstrukturmerkmalen bei ausreichend Zeit signifikant verbesserte ($F(1,66)=5,715$, $p=0,020$, $\eta^2=0,080$). Der Haupteffekt des Zwischen-subjektfaktors Bedingung, also das höhere spezifische Wissen, spielte bei der Identifizierung von Tiefenstrukturmerkmalen keine entscheidende Rolle ($F(1,66)=0,020$, $p=0,888$, $\eta^2=0,000$). Die Interaktion zwischen Zeitdruck und Bedingung wurde nicht signifikant ($F(1,66)=1,1946$, $p=0,168$, $\eta^2=0,029$). Die Hypothese, dass für das Identifizieren von Tiefenstrukturmerkmalen sowohl ausreichend Zeit als auch spezifisches, fachdidaktisches Wissen notwendig ist, kann damit nur teilweise bestätigt werden.

5.2 Effekte von Zeitdruck und fachdidaktischem Wissen auf die Gewichtung schwierigkeitsgenerierender Merkmale

Hinsichtlich der Untersuchung des Urteilsprozesses der Gewichtung von Aufgabenmerkmalen wurden die Ergebnisse der inkonsistenten Aufgabenpaare (d.h. gegenläufige Variation zweier Aufgabenmerkmale) verglichen. Auch bei der Gewichtung ist der Haupteffekt des Zwischensubjektfaktors des vermittelten Wissens signifikant ($F(1,66)=5,863$, $p=0,018$, $\eta^2=0,082$); die Experimentalgruppe erzielte mit mittlerem Effekt die besseren Ergebnisse (Hypothese 1). Ebenfalls wirkte sich der Innersubjektfaktor Zeitdruck signifikant auf die durchschnittliche Anzahl an richtig eingeschätzten Aufgabenvergleichen aus ($F(1,66)=12,874$, $p=0,001$, $\eta^2=0,163$) (Hypothese 2). Die Interaktion zwischen Zeitdruck und Bedingung zeigte keinen signifikanten Effekt ($F(1,66)=0,446$, $p=0,506$, $\eta^2=0,007$).

6 Diskussion

Das Hauptanliegen der vorliegenden Studie war es, die Modellannahmen über das Identifizieren und Gewichten von Aufgabenmerkmalen als zentrale kognitive Prozesse bei der Bildung diagnostischer Urteile über Aufgabenschwierigkeiten (Blömeke et al. 2015; Loibl et al. 2020.) empirisch zu untermauern. Dazu wurden Diagnose-situationen (hier: Aufgabenvergleiche) generiert, in denen systematisch untersucht werden konnte, inwiefern informationsverarbeitende Prozesse die getroffenen Urteile erklären können. Ganz konkret wurde dabei die Verknüpfung der in der Situation vorliegenden Informationen (hier: Aufgabenmerkmale) und deren Kontext (hier: zur Verfügung stehende Zeit) mit dem bei den Personen vorhandenem Wissen (hier: spezifisches, fachdidaktisches Wissen) systematisch variiert. Diese Variationen wurden auf der Basis von veränderbaren Merkmalen in den genannten Bereichen aufeinander abgestimmt vorgenommen und das resultierende Urteil vorhergesagt. Im Einzelnen lassen sich die Befunde folgendermaßen interpretieren.

6.1 Einfluss des spezifischen fachdidaktischen Wissens auf das Urteilsverhalten

Die Kurzvermittlung von spezifischem fachdidaktischen Wissen (als Veränderung der Personencharakteristik) führte zu einer akkurateren Identifizierung und Gewichtung vom schwierigkeitsgenerierenden Merkmalen und damit einer höheren Urteilsgenauigkeit. Diese Ergebnisse entsprechen Hypothese 1. Entgegen der Annahmen konnte bei der Identifizierung der Tiefenstrukturmerkmale ohne Zeitdruck jedoch kein signifikanter Effekt des fachdidaktischen Wissens gezeigt werden. Unter Zeitdruck ließen sich die Teilnehmenden vor allem von den offensichtlichen oberflächlichen Merkmalen leiten, bei ausreichender Zeit identifizierten aber auch Teilnehmende ohne Kurzintervention Merkmale der Tiefenstruktur. Zusammenfassend konnte der Einfluss von spezifischem fachdidaktischen Wissen auf die Verarbeitungsgüte relevanter Merkmale der Aufgabenschwierigkeit gezeigt werden. Ostermann et al. (2018) fanden vergleichbare Ergebnisse im Bereich der Funktionen. In beiden Studien zeigte sich, dass instruierte Merkmale in Diagnosesituationen auch verwendet werden. Die vorliegende Studie liefert zusätzlich die Erkenntnis, welche dieser Merkmale in welcher Gewichtung verarbeitet werden. Durch die systematische Variation und Verortung der Merkmale in der Strukturebene der Aufgabe, konnte eine Beziehung zwischen den Situations- und Personencharakteristika der Urteilsituation hergestellt werden (Loibl et al. 2020).

6.2 Einfluss der zur Verfügung stehenden Zeit auf das Urteilsverhalten

Außerdem wurde in der vorliegenden Studie gezeigt, dass Zeitdruck das Urteilsverhalten der befragten Personen in der vorausgesagten Weise beeinflusst: Entsprechend Hypothese 2, identifizierten alle teilnehmenden Studierende weniger Merkmale und gewichteten sie weniger akkurat unter Zeitdruck als mit ausreichend Zeit. Dieses Ergebnis reiht sich in die Befunde verschiedener von Müsseler und Rieger (2017) berichteter Studien ein.

Zusammenfassend beeinflussen sowohl Personen- als auch Situationscharakteristika die Urteilsprozesse Identifizierung und Gewichtung von Aufgabenmerkmalen. Durch das Studiendesign mit paarweisen Vergleichen konsistenter und inkonsistenter Aufgaben (vgl. Sherman et al. 2014) und die strukturierte Variation der verwendeten Aufgabenmerkmale konnten die angenommenen kognitiven Urteilsprozesse getrennt erfasst werden.

6.3 Grenzen der aktuellen Studie

In der Intervention wurde eine Hierarchie der Aufgabenmerkmale vermittelt. Die Varianz der Regressionskoeffizienten (siehe Ergebnisse der Vorstudie zur empirischen Fundierung der schwierigkeitsgenerierenden Aufgabenmerkmale, Tab. 2) impliziert jedoch, dass die relative Schwierigkeit von Aufgabentypen nicht immer eindeutig im Sinne einer Hierarchie der Merkmale zu formulieren ist. So legen die Daten der Vorstudie nahe, dass Aufgabenmerkmale mit kleinen Unterschieden in den Regressionskoeffizienten auch gleich relevant sein können. Somit ist die relative Schwierigkeit von Aufgabentypen in der Intervention nicht zufriedenstellend darge-

stellt. Zur Überprüfung der Modellannahmen zur wissensbasierten Identifizierung und Gewichtung der Merkmale in der vorliegenden Hauptstudie ist eine so präzise Darstellung des Diagnosegegenstandes aber auch nicht nötig, da unabhängig von der empirischen Fundierung der vermittelten Wissensinhalte, der Einfluss von Wissen auf die Diagnoseprozesse untersucht wurde.

Die Variation der Situationscharakteristik, also die Beurteilung der Aufgabenschwierigkeit unter *Zeitdruck*, wurde durch die durchschnittliche Lesedauer der verwendeten Aufgaben umgesetzt. Die Ergebnisse der Experimentalgruppe unter Zeitdruck zeigen allerdings, dass trotz der begrenzten Zeit Tiefenstrukturmerkmale verarbeitet wurden. Eine Möglichkeit, um noch genaueren Aufschluss über den Einfluss von Zeit zu bekommen, wäre eine Erhöhung der Zeitrestriktion oder das Bewerten der Aufgabenschwierigkeit unter anderweitiger kognitiver Auslastung. In der vorliegenden Studie ist zudem der Faktor Zeitdruck mit der wiederholten Bearbeitung der Aufgabenvergleiche konfundiert. Eine Verwendung von zwei unterschiedlichen Aufgabensätzen pro Messzeitpunkt würde dieser Konfundierung entgegenreten.

Ein besonderes Erkenntnisinteresse bezog sich auf die Identifizierung von Aufgabenmerkmalen in der Tiefenstruktur (Betsch et al. 2011), da diese auf den ersten Blick schwierig zu entdecken sind und dennoch – wie im Fall der zusätzlichen Grundvorstellung – entscheidend für die Aufgabenschwierigkeit sind. Entgegen unserer Hypothese, dass Tiefenstrukturmerkmale nur bei vorhandenem fachdidaktischem Wissen und ohne Zeitrestriktion verarbeitet werden können, zeigten die Ergebnisse, dass ohne Zeitrestriktion auch die Kontrollgruppe teilweise in der Lage war, Tiefenstrukturmerkmale wahrzunehmen. Vor dem Hintergrund dieser Ergebnisse gilt es die Verortung der Merkmale zwischen Oberfläche und Tiefenstruktur zu validieren sowie die Hypothese mit einer systematischen Variation weiterer Tiefenstruktur- und Oberflächenmerkmale erneut zu untersuchen.

Bezüglich der verwendeten und variierten *Aufgabenmerkmale* werden sowohl das Operationsverständnis wie auch die sprachlogische Komplexität durch die verwendete Sprache ausgedrückt. Schwierigkeiten, die bei Schülerinnen und Schülern auftreten und auf sprachliche Merkmale zurückzuführen sind, manifestieren sich dabei auf syntaktischer wie auch auf semantischer Ebene. Dies kann dazu führen, dass sprachliche Bezüge weder verstanden noch die dahintersteckende Operation richtig gedeutet wird. Dies könnte den hohen Einfluss des Merkmals Operationsverständnis auf die Aufgabenschwierigkeit erklären (sprachliche und inhaltliche Schwierigkeit). Ein Desiderat für künftige Untersuchungen schwierigkeitsgenerierender Merkmale von Aufgabe ist demnach die systematische Konzeption von Aufgaben unter Berücksichtigung der theoretischen Trennschärfe ihrer Merkmale. Diese Einschränkung bezieht sich auf die Vorstudie als empirische Grundlage der Intervention, nicht auf die untersuchten kognitiven Prozesse bei der Urteilsbildung.

Die Untersuchung der bei der Beurteilung zugrundeliegenden *kognitiven Prozesse* wurde durch die Gegenüberstellung von konsistenten und inkonsistenten Aufgabenvergleichen vorgenommen. Allerdings wird bei der Gewichtung die Identifizierung beider variierten Aufgabenmerkmale vorausgesetzt. Allerdings kann die richtige Antwort auch zustande kommen, falls das weniger relevante Merkmal nicht identifiziert wurde. Daher sollte in einer weiteren Studie die Urteilskonsistenz in den gegebenen Antworten, also die Frage nach dem gleichen Verhalten in vergleichbaren Urteils-

situationen genauer untersucht werden. So könnte die Identifizierung der Merkmale als Voraussetzung für die Gewichtung berücksichtigt werden.

6.4 Ausblick und Fazit

Die Studie bestätigt die zentralen Modellannahmen zur Informationsverarbeitung im Urteilsprozess: Es wurde gezeigt, dass die der Aufgabeneinschätzung zugrundeliegenden Urteilsprozesse durch Veränderungen der Personen- und Situationscharakteristika beeinflusst werden.

Neben der getrennten Erfassung der beiden kognitiven Prozesse Identifizierung und Gewichtung hat die vorliegende Studie eine hohe Praxisrelevanz: Die in dieser Studie konzipierte Kurzintervention und deren Ergebnisse liefern erste Hinweise zur Schulung diagnostischer Kompetenz in der Lehramtsausbildung. Durch eine kurze Intervention kann die Urteilsgüte im Allgemeinen und die Identifizierung und Gewichtung der für die Aufgabenschwierigkeit entscheidenden Aufgabenmerkmale im Besonderen gestärkt werden. Die Vermittlung von *fachdidaktischem Wissen* in der vorliegenden Studie diente allerdings der Überprüfung der Modellannahmen und beschränkte sich daher auf nur wenige schwierigkeitsgenerierende Merkmale. In einer realistischen Lehrveranstaltung wäre es wichtig, ein breiteres Spektrum schwierigkeitsgenerierender Merkmale zu vermitteln. Außerdem müsste eine Fortbildung im Rahmen der Lehreraus- bzw. -fortbildung noch größeren Wert auf die hierarchische Anordnung der Merkmale und deren Gewichte legen und die Möglichkeit der Verallgemeinerung von Ergebnissen der Bruchrechenaufgaben in andere Bereiche bedacht werden.

Mit der Vermittlung von fachdidaktischem Wissen wurde ein Faktor für die Verbesserung der Beurteilungsfähigkeit von Aufgabenschwierigkeit ermittelt. Für den schulischen Kontext stellt sich darüber hinaus die Frage, welche Rolle Erfahrung bei der Identifizierung und Gewichtung von Aufgabenmerkmalen spielt. Dazu könnten in einer weiteren Studie erfahrene Lehrkräfte denselben Urteilsituationen ausgesetzt werden.

Neben den praxisrelevanten Aspekten der hier vorgestellten Studie soll auch ein wesentlicher forschungsstrategischer Impuls herausgehoben werden: Um kognitive Prozesse der Urteilsgenese aufzuklären, bedarf es einer theoretischen kognitiven Modellierung (hier: im DiaCoM-Rahmenmodell) sowie die systematische Strukturierung und Variation relevanter Faktoren (hier: Aufgabenmerkmale und darauf abgestimmtes fachdidaktisches Wissen). Eine solche Forschungsstrategie erlaubt eine empirische Fundierung einer Theorie diagnostischer Kompetenz, wie sie vielfach gefordert wird (Schrader 2013; Herppich et al. 2017).

Förderung Das Projekt wurde im Rahmen des Forschungs- und Nachwuchskollegs DiaKom (Diagnostische Kompetenzen von Lehrkräften: Einflüsse, Struktur und Förderung) der Pädagogischen Hochschulen Freiburg und Heidelberg vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert.

Funding Open Access funding provided 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

jedlichem 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

- v. Aufschnaiter, C., Cappell, J., Dübbele, G., Ennemoser, M., Mayer, J., Stiensmeier-Pelster, J., Wolgast, A., et al. (2015). Diagnostische Kompetenz. Theoretische Überlegungen zu einem zentralen Konstrukt der Lehrerbildung. *Zeitschrift für Pädagogik*, 61(5), 738–758.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. *Journal of Teacher Education*, 59(5), 389–407. <https://doi.org/10.1177/0022487108324554>.
- Betsch, T., Funke, J., & Plessner, H. (2011). *Denken – Urteilen, Entscheiden, Problemlösen: Allgemeine Psychologie für Bachelor*. Berlin, Heidelberg: Springer. from <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10455739>
- Binder, K., Krauss, S., Hilbert, S., Brunner, M., Anders, Y., & Kunter, M. (2018). Diagnostic skills of mathematics teachers in the COACTIV study. In I. T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic competence of mathematics teachers: unpacking a complex construct in teacher education and teacher practice*. Mathematics Teacher Education, (Bd. 11, S. 33–53). Cham: Springer. https://doi.org/10.1007/978-3-319-66327-2_2.
- Black, P., & Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. *The Phi Delta Kappan*, 80(2), 139–148.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>.
- Böhmer, I., Hösternmann, T., Gräsel, C., Krolak-Schwerdt, S., & Glock, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung: Welcher Urteilsregel folgen Lehrkräfte? *Journal for Educational Research Online*, 7(2), 59–81.
- Böhmer, M., Englich, B., & Böhmer, I. (2017). Schülerbeurteilung aus der Perspektive dualer Prozessmodelle der sozialen Urteilsbildung. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen*. Münster, New York: Waxmann.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hoboken: Taylor and Francis. <http://gbv.eblib.com/patron/FullRecord.aspx?p=1192162>
- Dünnebier, K., Gräsel, C., & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. *Zeitschrift für Pädagogische Psychologie*, 23(34), 187–195. <https://doi.org/10.1024/1010-0652.23.34.187>.
- Eichmann, A., Narciss, S., Schnaubert, L., & Melis, E. (2012). Typische Fehler bei der Addition und Subtraktion von Brüchen – Ein Review zu empirischen Fehleranalysen. *Journal für Mathematik-Didaktik*, 33(1), 29–57. <https://doi.org/10.1007/s13138-011-0031-5>.
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 5(9), 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>.
- Gürsoy, E., Benholz, C., Renk, N., Prediger, S., & Büchter, A. (2013). Erlös = Erlösung? – Sprachliche und konzeptuelle Hürden in Prüfungsaufgaben zur Mathematik. In *Deutsch als Zweitsprache* (S. 14–24).
- Hammer, S. (2016). *Professionelle Kompetenz von Mathematiklehrkräften im Umgang mit Aufgaben in der Unterrichtsplanung: Theoretische Grundlegung und empirische Untersuchung (1. Auflage)*. Texte zur mathematischen Forschung und Lehre
- Hartig, J., & Frey, A. (2012). Konstruktvalidierung und Skalenbeschreibung in der Kompetenzdiagnostik durch die Vorhersage von Aufgabenschwierigkeiten. *Psychologische Rundschau*, 63(1), 43–49. <https://doi.org/10.1026/0033-3042/a000109>.

- Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2003). Diagnosekompetenz in Ausbildung und Beruf entwickeln. *Karlsruher Pädagogische Beiträge*, 55, 15–34.
- Herppich, S., Praetorius, A.-K., Hetmanek, A., Glogger-Frey, I., Ufer, S., Leutner, D., & Südkamp, A. (2017). Ein Arbeitsmodell für die empirische Erforschung der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen*. Münster, New York: Waxmann.
- vom Hofe, R. (1995). *Grundvorstellungen mathematischer Inhalte*. Heidelberg: Spektrum.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, 59(3), 297–313. <https://doi.org/10.3102/00346543059003297>.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., et al. (2006). Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt. *Materialien aus der Bildungsforschung*, Nr. 81. Berlin: Max-Planck-Institut für Bildungsforschung.
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M., & Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *JMD*, 29(2008), 233–258.
- Krolak-Schwerdt, S., & Rummer, R. (2005). Der Einfluss von Expertise auf den Prozess der schulischen Leistungsbeurteilung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37(4), 205–213. <https://doi.org/10.1026/0049-8637.37.4.205>.
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. *Zeitschrift für Pädagogische Psychologie*, 23(34), 175–186. <https://doi.org/10.1024/1010-0652.23.34.175>.
- Lindmeier, A. (2011). *Modeling and measuring knowledge and competencies of teachers: a threefold domain-specific structure model for mathematics*. Empirische Studien zur Didaktik der Mathematik.
- 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>.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Ullrich, M., et al. (2009). Diagnostische Fähigkeiten von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(34), 223–235. <https://doi.org/10.1024/1010-0652.23.34.223>.
- Musch, J., & Rösler, P. (2011). Schnell-Lesen: Was ist die Grenze der menschlichen Lesegeschwindigkeit? In M. Dresler (Hrsg.), *Kognitive Leistungen*. Heidelberg: Spektrum Akademischer Verlag.
- Müsseler, J., & Rieger, M. (2017). *Allgemeine Psychologie* (3. Aufl.). Berlin, Heidelberg: Springer.
- Neubrand, M., Jordan, A., Krauss, S., Blum, W., & Löwen, K. (2011). Aufgaben im COACTIV-Projekt: Einblicke in das Potenzial für kognitive Aktivierung im Mathematikunterricht. In M. Kunter (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 115–132).
- Neubrand, M., Klieme, E., Lüdtke, O., & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung. *Unterrichtswissenschaft*, 30(1), 100–119.
- Ostermann, A., Leuders, T., & Nückles, M. (2015). Wissen, was Schülerinnen und Schülern schwer fällt. Welche Faktoren beeinflussen die Schwierigkeitseinschätzung von Mathematikaufgaben? *Journal für Mathematik-Didaktik*, 36(1), 45–76. <https://doi.org/10.1007/s13138-015-0073-1>.
- Ostermann, A., Leuders, T., & Nückles, M. (2018). Improving the judgment of task difficulties: prospective teachers' diagnostic competence in the area of functions and graphs. *Journal of Mathematics Teacher Education*, 21(6), 579–605. <https://doi.org/10.1007/s10857-017-9369-z>.
- Ostermann, A., Leuders, T., & Philipp, K. (2019). Fachbezogene diagnostische Kompetenzen von Lehrkräften – Von Verfahren der Erfassung zu kognitiven Modellen zur Erklärung. In T. Leuders, M. Nückles & S. Mikelskis-Seifert (Hrsg.), *Pädagogische Professionalität in Mathematik und Naturwissenschaften* (S. 93–116). Wiesbaden: Springer.
- Oudman, S., van de Pol, J., Bakker, A., Moerbeek, M., & van Gog, T. (2018). Effects of different cue types on the accuracy of primary school teachers' judgments of students' mathematical understanding. *Teaching and Teacher Education*, 76, 214–226. <https://doi.org/10.1016/j.tate.2018.02.007>.
- Padberg, F. (1986). Über typische Schülerschwierigkeiten in der Bruchrechnung – Bestandsaufnahme und Konsequenzen. *Der Mathematikunterricht*, 32(3), 58–77.
- Padberg, F., & Wartha, S. (2017). *Didaktik der Bruchrechnung*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- 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.

- 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*. Mathematics Teacher Education, (Bd. 11, S. 129–148). Cham: Springer. https://doi.org/10.1007/978-3-319-66327-2_7.
- Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, 127(2), 258–276. <https://doi.org/10.1016/j.actpsy.2007.05.004>.
- Rösler, F. (2011). *Psychophysiologie der Kognition*. Heidelberg: Spektrum Akademischer Verlag.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt: Peter Lang.
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerbildung*, 31, 2.
- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27–52.
- Schult, J., & Lindner, M. A. (2018). Diagnosegenauigkeit von Deutschlehrkräften in der Grundschule: Eine Frage des Antwortformats? *Zeitschrift für Pädagogische Psychologie*, 32(1–2), 75–87. <https://doi.org/10.1024/1010-0652/a000216>.
- Schulz, A., Leuders, T., & Rangel, U. (2019). The use of a diagnostic competence model about children's operation sense for criterion-referenced individual feedback in a large-scale formative assessment. *Journal of Psychoeducational Assessment*. <https://doi.org/10.1177/0734282918823590>.
- Sherman, J. W., Krieglmeyer, R., & Calanchini, J. (2014). Process models require process measures. In J. W. Sherman, B. Gawronski & Y. Trope (Hrsg.), *Dual-process theories of the social mind* (S. 121–138). New York: Guilford.
- Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>.
- Statistisches Bundesamt (Hrsg.). (2019). Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus. Ausgabe 2018. Wiesbaden: Statistisches Bundesamt.
- 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.
- Thames, M. H., & Ball, D. L. (2010). What math knowledge does teaching require? *Teaching Children Mathematics*, 17(4), 220–229.
- Van Ophuysen, S. (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnempfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38(4), 154–161. <https://doi.org/10.1026/0049-8637.38.4.154>.
- Wartha, S. (2007in). *Kompetenzen im Bruchrechnen: Die Rolle von Grundvorstellungen*. 41. Jahrestagung der Gesellschaft für Didaktik der Mathematik, Berlin, 25.3. bis 30.3.2007. Beiträge zum Mathematikunterricht 2007.
- Winter, H. (1999). *Mehr Sinnstiftung, mehr Einsicht, mehr Leistungsfähigkeit, dargestellt am Beispiel der Bruchrechnung*. Aachen: RWTH Aachen.