Zeitschrift für Erziehungswissenschaft
https://doi.org/10.1007/s11618-026-01430-7

SCHWERPUNKT

Die Rolle des fachdidaktischen Wissens und subjektiver Einstellungen gegenüber den Bildungsstandards für kognitiv aktivierenden Mathematikunterricht in PISA-Ceco

Stefanie Hohenleitner · Stefan Krauss · Nathalie Stegmüller · Alfred Lindl

Eingegangen: 7. Mai 2025 / Überarbeitet: 7. Juli 2026 / Angenommen: 21. Juli 2026
© The Author(s) 2026

Zusammenfassung Das fachdidaktische Wissen (FDW) zählt zu den entscheidenden professionellen Kompetenzen einer Lehrkraft. In der PISA-Begleitstudie Ceco 2022 wurde für das Fach Mathematik eine aus sechs Items bestehende Kurzskala zum FDW bei Lehrkräften implementiert. Der vorliegende Beitrag beleuchtet einerseits die Zusammenhänge des FDW mit dem Kenntnisstand und den Einstellungen von Mathematiklehrkräften gegenüber den Bildungsstandards (BiSta). Andererseits wird die Bedeutung des FDW und der BiSta-Variablen für die Gestaltung eines kognitiv aktivierenden Unterrichts untersucht. Zur Messung der kognitiven Aktivierung werden verschiedene Operationalisierungen verwirklicht. Zum einen wird dabei auf Einschätzungen von Lehrkräften sowie von Schülerinnen und Schülern zurückgegriffen und zum anderen auf diesbezügliche Kodierungen von eingesammelten Klassenarbeiten. Es zeigt sich, dass das FDW signifikant mit dem von den Schülerinnen und Schülern wahrgenommenen Niveau des mathematischen Argumentierens im Unterricht zusammenhängt, dieser Effekt aber von der Schulform moderiert wird. Einstellungen zu den Bildungsstandards hatten keine vergleichbaren Effekte.

✉ Stefanie Hohenleitner · Prof. Dr. Stefan Krauss · Nathalie Stegmüller
Didaktik der Mathematik, Universität Regensburg, Universitätsstraße 31, 93053 Regensburg, Deutschland
E-Mail: stefanie.hohenleitner@ur.de

Prof. Dr. Stefan Krauss
E-Mail: stefan.krauss@ur.de

Nathalie Stegmüller
E-Mail: nathalie.stegmueller@ur.de

Prof. Dr. Alfred Lindl
Professionalisierungsforschung im Bildungsbereich, Universität Hildesheim, Universitätsplatz 1, 31141 Hildesheim, Deutschland
E-Mail: alfred.lindl@uni-hildesheim.de

Published online: 26 October 2026

Schlüsselwörter Mathematiklehrkräfte · Fachdidaktisches Wissen · Bildungsstandards · Kognitive Aktivierung

The role of pedagogical content knowledge and attitudes towards the "Bildungsstandards" for cognitive activation in mathematics teaching in PISA-Ceco

Abstract Pedagogical content knowledge (PCK) is considered one of the key professional competencies of a teacher. In the PISA-Ceco study accompanying the Program for International Student Assessment 2022 (PISA 2022), a short scale consisting of six items was implemented to assess PCK among mathematics teachers. We examined the relationship between PCK and both teachers' level of knowledge regarding the German *Bildungsstandards* (educational standards; BiSta) and their attitudes toward them. Further, we investigated the role of PCK and BiSta variables for the design of cognitively activating instruction. Various operationalizations were employed to measure cognitive activation, including teachers' and students' evaluations as well as corresponding coding of collected classroom tests. The findings indicate that PCK is significantly related to the level of mathematical argumentation perceived by students in the classroom, although this effect is moderated by the type of school. Attitudes toward the BiSta do not show comparable effects.

Keywords Mathematics teachers · Pedagogical content knowledge · Educational standards ("Bildungsstandards") · Cognitive Activation

1 Einleitung

Nach dem Konstanzer Beschluss 1997 nahm Deutschland zur Jahrtausendwende erstmals an internationalen Bildungsvergleichsstudien wie TIMSS 1999 und PISA 2000 teil (vgl. KMK 2004). Das unbefriedigende Abschneiden in diesen Studien („PISA-Schock"; Baumert et al. 2001) hatte mehrere Konsequenzen: Im deutschen Sprachraum rückten professionelle Kompetenzen von Lehrkräften und Merkmale qualitätsvollen Unterrichts in den Fokus der empirischen Bildungsforschung (z. B. Kunter et al. 2011). Darüber hinaus setzte ein umfassender Reformprozess im deutschen Schulsystem ein, der unter anderem in der Entwicklung bundesweit geltender Bildungsstandards mündete (für das Fach Mathematik: Blum et al. 2006). Mittlerweile folgen Mathematikcurricula und Schulbücher in allen Bundesländern den Vorgaben der Bildungsstandards (z. B. ISB 2025).

In PISA 2022 war das Fach Mathematik zum dritten Mal (nach 2003 und 2012) die untersuchte Hauptdomäne. Die Begleitstudie PISA-Ceco (Classroom Experience, Characteristics and Outcome) erweiterte das Design der PISA-Studie um zusätzliche Erhebungen zu professionellen Kompetenzmerkmalen von Lehrkräften in den Fächern Mathematik, Biologie, Chemie und Physik sowie zur Qualität des mathematisch-naturwissenschaftlichen Unterrichts in Deutschland (Schiepe-Tiska et al. 2025, 2026).

Der vorliegende Beitrag widmet sich aus mathematikdidaktischer Perspektive ausgewählten Aspekten von PISA-Ceco und nimmt insbesondere das fachdidaktische Wissen (FDW), die Bildungsstandards für das Fach Mathematik (BiSta) sowie das Potential für kognitive Aktivierung (KA) im Unterricht der PISA-Klassen in den Blick. Während FDW und KA in empirischen Studien bereits wiederholt als entscheidend für Unterrichtsqualität und Lernzuwächse von Schülerinnen und Schülern identifiziert wurden (z. B. Fukaya et al. 2025; Praetorius und Charalambous 2018), liegen vergleichbare Analysen zur Rolle der Bildungsstandards bislang nicht vor. Im vorliegenden Artikel wird zunächst untersucht, in welchem Zusammenhang der Kenntnisstand über die Bildungsstandards (Ke-B) und die Einstellung zu deren Nutzen (Nu-B) mit dem FDW von Mathematiklehrkräften stehen. Anschließend wird analysiert, inwiefern FDW, Ke-B und Nu-B zur kognitiven Aktivierung im Unterricht beitragen können.

Sowohl FDW als auch KA wurden in PISA-Ceco zunächst übergreifend für mathematisch-naturwissenschaftlichen Unterricht, d. h., gemeinsam für die Fächer Mathematik, Biologie, Chemie und Physik konzeptualisiert (vgl. die Skalendokumentation Schiepe-Tiska et al. 2025). Für das FDW von Mathematiklehrkräften kam eine Kurzskala des entsprechenden COACTIV-Tests zum Einsatz, dessen drei Facetten – Wissen über 1) Erklären, 2) über Schülerfehler und 3) über das Potential von Lernmaterialien – dann auch sinngemäß für die FDW-Tests in den drei naturwissenschaftlichen Fächern übernommen wurden (z. B. Tepner et al. 2026).

Für die Analysen von KA wird für den vorliegenden Beitrag zum einen auf die Fragebogenskala zum wahrgenommenen Anspruch des Argumentierens im Unterricht sowohl aus Sicht der Schülerinnen und Schüler als auch der Lehrkräfte zurückgegriffen. Zum anderen wird der mathematikdidaktischen Tradition folgend das *Potential* zur kognitiven Aktivierung über die in den PISA-Klassen jeweils eingesetzten Aufgaben abgeschätzt. Im Mathematikunterricht gelten mit Bezug zu den Bildungsstandards vor allem die prozessbezogenen Kompetenzen Argumentieren (K1), Problemlösen (K2) und Modellieren (K3) als fachlich gehaltvolle und kognitiv aktivierende Tätigkeiten, aber beispielsweise auch das begriffliche Denken (im Gegensatz zum rein rechnerischen Arbeiten). Als Indikatoren von KA werden demzufolge die von externen Raterinnen und Ratern kodierten Aufgabendimensionen „Argumentieren“, „innermathematisches Problemlösen“, „Modellieren“ und „Anteil begrifflicher Aufgaben (vs. rein rechnerischer Aufgaben)“ betrachtet.

2 Theorie

Die Notwendigkeit einer stärkeren Fokussierung fachspezifischer Merkmale in der Unterrichtsforschung betonen ganz allgemein beispielsweise Seidel und Shavelson (2007), Begrich et al. (2023) oder Reusser (2024). In der mathematikdidaktischen Forschung werden seit etwa 20 Jahren verstärkt das FDW als professioneller Kompetenzaspekt von Lehrkräften sowie KA als fachspezifisches Qualitätsmerkmal von Unterricht in den Blick genommen (z. B. Heinle et al. 2026; Praetorius et al. 2024; Yang und Kaiser 2022). Nicht zuletzt aufgrund der großen Bedeutung, die beide Konstrukte in der COACTIV-Studie für Lernzuwächse von Schülerinnen und Schü-

lern hatten (Kunter et al. 2011), haben das FDW und die (fachbezogene) KA auch in der universitären Ausbildung von Mathematiklehrkräften in Deutschland in den letzten beiden Jahrzehnten an Bedeutung gewonnen (KMK 2024).

Im Folgenden soll ein Überblick über die Konstrukte fachdidaktisches Wissen (FDW), Kenntnisstand zu den Bildungsstandards (Ke-B), Einstellungen gegenüber dem Nutzen der Bildungsstandards (Nu-B) sowie die ausgewählten Aspekte von kognitiver Aktivierung (KA) gegeben werden. Eine Einbettung dieser fachdidaktischen Aspekte in den Gesamtrahmen von PISA-Ceco findet sich in Schiepe-Tiska et al. (2026).

2.1 Fachdidaktisches Wissen (FDW)

Shulman (1986, 1987) leitet anhand einer detaillierten Anforderungsanalyse für den Lehrberuf insgesamt sieben Kategorien des Lehrerwissens ab. Aus seiner einflussreichen theoretischen Wissenstaxonomie stehen heute vor allem das Fachwissen (FW), das pädagogisch-psychologische Wissen (PW) und das fachdidaktische Wissen (FDW) im Forschungsfokus (Ball und Hill 2023; Lindl et al. 2023; Schwarz et al. 2023). FDW und FW bilden die beiden fachspezifischen Wissensbereiche, das PW wird in der Regel unabhängig von den unterrichteten Fächern einer Lehrkraft konzeptualisiert (König et al. 2023). Zu allen drei Wissenskategorien gibt es mittlerweile zahlreiche psychometrische Testkonstruktionen und empirische Befunde (für Überblicksarbeiten siehe z. B. Charalambous et al. 2020; Fukaya et al. 2025; Yang und Kaiser 2022).

In der deutschsprachigen Forschungslandschaft hat sich die Untersuchung des FDW vor allem im Rahmen der Messung professioneller Lehrkraftkompetenzen im Zuge von Large-Scale-Studien wie TEDS (Blömeke et al. 2011, 2013) oder COACTIV (Kunter et al. 2011) etabliert (derzeit verfügbare FDW-Tests für Mathematiklehrkräfte werden detailliert vorgestellt in Krauss und Lindl 2023). Fast allen quantitativen FDW-Testkonstruktionen ist gemein, dass sie – Shulman (1986, 1987) folgend – sowohl die Facette „Erklärungswissen von Fachinhalten“ als auch „Wissen über typische Schülerfehler und/oder -schwierigkeiten“ berücksichtigen (Ball und Hill 2023; Depaepe et al. 2013; Park und Oliver 2008). Im FALKO-Projekt, in dem analog zu COACTIV psychometrische FDW-Tests für sechs verschiedene Unterrichtsfächer (Deutsch, Englisch, Latein, Musik, Religion, Physik) konstruiert wurden, wurde als dritte Facette für alle beteiligten Fächer noch das Wissen über das Potential von fachspezifischen Lernmaterialien für verständnisvolle Lernprozesse operationalisiert (Krauss et al. 2017). In Anlehnung an FALKO wurde diese dritte FDW-Facette auch in PISA-Ceco 2022 zur Konstruktion der vier FDW-Tests aufgegriffen (Mathematik, Biologie, Chemie, Physik).

Neben allen Differenzierungen, variierenden Konzeptualisierungen und Möglichkeiten der Operationalisierung besteht nach Depaepe et al. (2013) – auch fächerübergreifend – Konsens über folgende vier Aspekte:

1. Das FDW ist eine Form des praxisbezogenen Wissens von Lehrkräften, das notwendig für das Unterrichten ist.
2. Das FW ist eine wichtige und notwendige Voraussetzung für das FDW.

3. Das FDW verbindet das FW und das PW.
4. Das FDW ist stets fachspezifisch.

In der COACTIV-Studie hatte das FDW bei über 1000 erfassten Variablen den größten Effekt auf die Unterrichtsqualität (speziell auf KA), aber auch auf den Leistungszuwachs von Schülerinnen und Schülern (Baumert et al. 2010). In einer aktuellen Metaanalyse zu Effekten des FDW von Mathematik- und Naturwissenschaftslehrkräften können diese positiven Effekte – trotz einiger gegenläufiger Befunde – im Wesentlichen bestätigt werden (Fukaya et al. 2025).

Auch im internationalen Diskurs wird das FDW mittlerweile als ein zentraler Kompetenzaspekt von Lehrkräften betrachtet. In einem Literaturreview zur Expertise von Lehrkräften (Anderson und Taner 2023), das sich vor allem auf qualitative Studien bezieht, gehört das FDW mittlerweile zu den drei am häufigsten untersuchten Expertisemerkmalen in der Lehrkräfte- bzw. Unterrichtsforschung.

Für eine ausführliche Charakterisierung des FDW sowie einen diesbezüglichen historischen Forschungsüberblick mit Blick auf die PISA-Ceco Testkonstruktionen in Mathematik, Biologie, Chemie und Physik sei an dieser Stelle auf die Skalendokumentation verwiesen (vgl. Hohenleitner et al. 2025).

2.2 Bildungsstandards für den mittleren Schulabschluss im Fach Mathematik (BiSta)

Die Grundlage für die Formulierung der Bildungsstandards war eine Expertise von Klieme et al. (2001), die ursprünglich als Reaktion auf die unbefriedigenden TIMSS-Ergebnisse 1999 und somit noch vor dem PISA-Schock 2000 verfasst wurde. Für das Fach Mathematik definieren die bundesweiten BiSta für den mittleren Schulabschluss (d. h., bis zu Klasse 10), welche Kompetenzen Schülerinnen und Schüler zu bestimmten Leitideen erwerben sollen (Blum et al. 2006). Mittlerweile wurden auch für die gymnasiale Oberstufe (Hochschulreife) bundeseinheitliche Bildungsstandards für das Fach Mathematik verabschiedet (KMK 2012). Die BiSta verfolgen nach Klieme (2003) zwei Ziele: Zum einen dienen sie Lehrkräften als Orientierung für die Entwicklung und Planung von Unterricht, zum anderen erlauben sie der Bildungsadministration die Konzeption standardisierter Tests, um das Erreichen der erwünschten Kompetenzen zu überprüfen. Die Standards unterliegen einer ständigen Weiterentwicklung auf der Basis von Schulpraxis, fachdidaktischer Forschung und Evaluationstests (KMK 2004, 2022).

Auch wenn sich alle Bundesländer vor 20 Jahren zur Umsetzung der BiSta bis Klasse 10 in Prüfungen, Lehrplänen, Unterricht sowie Lehrkräfteaus- und -fortbildungen verpflichtet haben, wurde den Ländern die Wahl des Zeitpunkts der Implementation in das Schulsystem prinzipiell freigestellt. So wurden beispielsweise in Bayern die BiSta erst mit dem LehrplanPlus seit dem Schuljahr 2017/2018 in allen Schulformen im Sekundarbereich (Mittelschule, Realschule, Gymnasium) explizit und „offiziell“ umgesetzt. Für den mittleren Schulabschluss gibt es bereits eine Überarbeitung und Neuauflage, die vor allem der zunehmenden Digitalisierung Rechnung trägt (KMK 2022). Diese war den Lehrkräften der PISA-Ceco Studie zum Zeitpunkt der Erhebung allerdings noch unbekannt.

Für den vorliegenden Beitrag ist die für die Mathematiklehrkräfte der PISA-Ceco Studie gültige Version der BiSta mit folgenden sechs prozessbezogenen Kompetenzen relevant (KMK 2004): K1: Mathematisch argumentieren; K2: Probleme mathematisch lösen; K3: Mathematisch modellieren; K4: Mathematische Darstellungen verwenden; K5: Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen; K6: Kommunizieren (das mathematische Arbeiten mit Medien wurde dann als siebte Kompetenz im Rahmen der Überarbeitung ergänzt; KMK 2022).

Zentrale Faktoren für eine gelingende Implementation der BiSta sind der Kenntnisstand (Ke-B) sowie diesbezügliche Einstellungen (Nu-B) der Lehrkräfte, die sie letztlich umsetzen (Holder 2019). Pant et al. (2008) betonen, dass die Lehrkräfte den BiSta zwar grundsätzlich offen gegenüberstehen, zugleich aber auch Informationsdefizite bestehen. Weiterhin zeigt sich bei Deutschlehrkräften, dass eine positive Einstellung zu Bildungsstandards nicht zwangsläufig auch deren optimale Umsetzung bedeutet, vor allem, wenn keine zusätzlichen Unterstützungsangebote bereitgestellt werden (Böttcher und Dicke 2008). Beide Umfragen erfolgten noch zu einer Zeit, in der die Bildungsstandards für alle Unterrichtsfächer zwar bereits verabschiedet und auch in Schulen schon umfänglich diskutiert wurden, aber in den meisten Bundesländern noch nicht explizit im Curriculum verankert waren.

In Bezug auf den Mathematikunterricht stellt Frenzel (2018) fest, dass zumindest bei zusätzlichen Unterstützungsmaßnahmen eine größere Orientierung hin zum mathematischen Argumentieren, Modellieren und Problemlösen möglich ist. Dies sind die drei prozessbezogenen Kompetenzen, die wiederholt auch als zentrale Indikatoren von KA genannt werden (vgl. 2.3). Die Kompetenz „mathematisch argumentieren“ (K1) umfasst sowohl das Entwickeln als auch das Verstehen und kritische Bewerten von mathematischen Argumentationen – von einfachen Plausibilitätsüberlegungen bis hin zu formal korrekten Beweisen (Blum et al. 2006; KMK 2022). Die Kompetenz „Probleme mathematisch lösen“ (K2) beinhaltet die Fähigkeit, mathematische Fragestellungen zu bearbeiten, bei denen der Lösungsweg nicht unmittelbar erkennbar ist (d. h., es keine Standardprozedur gibt). Hierfür sollen Schülerinnen und Schüler geeignete heuristische Strategien wählen oder entwickeln sowie (neuartige) Lösungswege erproben und reflektieren (Blum et al. 2006; KMK 2022). Die Kompetenz „mathematisch Modellieren“ (K3) bedeutet, reale Problemstellungen mithilfe der Mathematik zu verstehen, zu vereinfachen und zu lösen. Zentraler Bestandteil dieser Kompetenz ist dabei das Übersetzen zwischen einem Sachkontext (Realsituation) und einem mathematischen Modell (Blum et al. 2006; KMK 2022).

Aktuelle Überblicksarbeiten (z. B. Albers und Jude 2025) beschäftigen sich nicht mehr so sehr mit der Akzeptanz der BiSta (diese gelten vielmehr als gegeben), als eher mit Möglichkeiten der Überprüfung, ob Schülerinnen und Schüler die in den BiSta vorgegebenen Kompetenzen erreichen (beispielsweise durch die bundesweit vorgeschriebenen VERA-Vergleichstests). Auch heute wird allerdings von Seiten der Lehrkräfte noch Bedarf an Unterstützungsangeboten vor allem dahingehend angemeldet, wie Vergleichsarbeiten – statt „nur“ zur Diagnose – zur datengestützten Entwicklung „kompetenzorientierten“ Unterrichts genutzt werden können (z. B. Albers und Jude 2025).

Inwieweit der Kenntnisstand über die BiSta (Ke-B) sowie diesbezügliche Einstellungen (Nu-B) mit dem FDW von Mathematiklehrkräften oder deren Unterrichtsqualität (z. B. KA) zusammenhängen, wurde unseres Wissens noch nicht explizit untersucht.

2.3 Kognitive Aktivierung (KA)

Unterricht, der die Lernenden zu eigenen elaborierenden Gedankengängen und zu einem vertieften Nachdenken über den Unterrichtsgegenstand anregt, wird als *kognitiv aktivierend* bezeichnet (Ufer et al. 2023). Die kognitive Aktivierung (KA) gilt mittlerweile als zentrale Tiefenstruktur von Unterrichtsqualität, die sowohl im Basismodell (Klieme et al. 2006) als auch im neueren Syntheseframework von Unterrichtsqualität (Praetorius et al. 2020) eine hervorgehobene Rolle spielt (weitere Tiefenstrukturen sind z. B. konstruktive Unterstützung und Klassenführung). Dabei hat sich KA nicht nur wiederholt als prädiktiv für Lernfortschritte erwiesen (z. B. Blömeke et al. 2022; Kunter et al. 2013; Praetorius und Charalambous 2018), sondern ist mittlerweile ein sowohl aus theoretischer als auch aus empirischer Sicht sehr vielschichtiger und schillernder Begriff geworden (Praetorius et al. 2024).

Im Mathematikunterricht gelten beispielsweise das Begründen, Analysieren, Verknüpfen, Schlussfolgern oder das begriffliche (statt nur rechnerische) Denken als intendierte Aktivitäten höherer Ordnung (z. B. Ufer et al. 2023). Im Hinblick auf die Kompetenzen der BiSta werden in diesem Zusammenhang vor allem das Argumentieren (K1), das Problemlösen (K2), und das Modellieren (K3) als kognitiv aktivierende mathematische Denkprozesse angeführt (Frenken et al. 2024; Scheja et al. 2024; vgl. 2.2). Bruder et al. (2008) bieten theoretische Modelle zur Klassifizierung von Aufgaben und erläutern hierbei, warum begrifflich-strukturelle Anforderungen im Gegensatz zu rein prozeduralen Abläufen eine tiefere kognitive Aktivierung auslösen.

Neben verschiedenen inhaltlichen Schwerpunktsetzungen werden in der Mathematikdidaktik unterschiedliche Beobachterperspektiven zur Erfassung von KA vorgeschlagen und empirisch umgesetzt (Praetorius et al. 2024; Schlesinger et al. 2018). Zum einen können die Unterrichtsprotagonisten selbst – das heißt Lehrkräfte beziehungsweise Schülerinnen und Schüler – zu Qualitätsaspekten befragt werden, die für fachlich gehaltvollen und kognitiv aktivierenden Mathematikunterricht stehen (Blum 2001). Zum anderen können externe Kodiererinnen und Kodierer herangezogen werden, um die von Mathematiklehrkräften genutzten Aufgaben nach deren Potential für kognitiv aktivierende mathematische Tätigkeiten zu beurteilen (Herbert und Schweig 2021; Neubrand et al. 2011).

Die überarbeitete Konzeptualisierung von KA in PISA-Ceco (vgl. Schiepe-Tiska et al. 2025) legt außerdem Wert darauf, dass es sich nicht nur um verschiedene Beobachtungsperspektiven handelt, sondern dass damit auch das betrachtete Konstrukt variiert (zur differentiellen Validität der verschiedenen Messmethoden siehe z. B. auch Clausen 2002; Fauth et al. 2020; Kleickmann et al. 2019; Praetorius et al. 2024). Über Fragebogenskalen beispielsweise können Lehrkräfte oder Schülerinnen und Schüler das Unterrichtsangebot retrospektivisch beurteilen. Eine solche

„Rekonstruktion“ der Unterrichtsqualität aus der Erinnerung kann durch subjektive Faktoren beeinflusst sein.

Die Analyse von im Unterricht eingesetzten Aufgaben kann dagegen – zumindest theoretisch – objektiv erfolgen. In diesem Zusammenhang wird auch oft betont, dass die Bearbeitung von Aufgaben *den* zentralen Bestandteil des Mathematikunterrichts darstellt (z. B. Johnson et al. 2017). Kodiert werden kann allerdings nur das den Aufgaben inhärente *Potential* für kognitive Aktivierung, da die Raterinnen und Rater in der Regel nichts über den tatsächlichen Einsatz der Aufgabe im Unterrichtsverlauf wissen. Das Potential für kognitive Aktivierung wird in der Praxis auch oft über Klassenarbeiten eingeschätzt, da die dort gestellten Aufgaben als repräsentativ für den Unterricht einer Lehrkraft gelten und darüber hinaus die Erwartungen der Lehrkräfte an ihre Schülerinnen und Schüler widerspiegeln (Baumert et al. 2010; Herbert und Schweig 2021; Neubrand et al. 2011).

Die Lehrkräfte schätzen also ihr eigenes Unterrichtsangebot ein, gestalten darüber hinaus aber auch Aufgaben, die ein Potential für kognitive Aktivierung beinhalten, das im Unterricht realisiert werden muss. Schülerinnen und Schüler wiederum beschreiben eher, inwiefern das Potential solcher Aufgaben während des Unterrichts tatsächlich genutzt wird. Deswegen unterscheiden wir im Folgenden zwischen der *wahrgenommenen* KA im Unterricht, wobei wir zwischen Lernenden und Lehrkräften differenzieren (jeweils über Fragebögen erfasst), und dem *Potential* für KA (über Aufgabenkodierungen abgeschätzt).

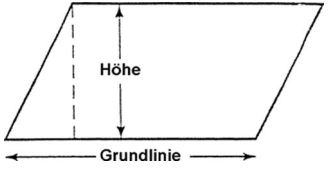
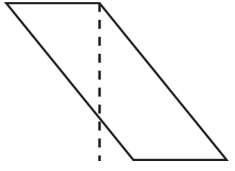
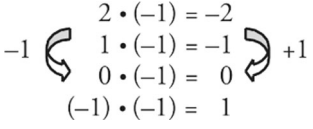
Wie auch schon in der COACTIV-Studie – hier konnte über beide Messmethoden (Unterrichtsfragebögen, Aufgabenkodierungen) ein mediierender Effekt der KA zwischen FDW und dem Leistungszuwachs von Schülerinnen und Schülern nachgewiesen werden (Kunter et al. 2011) – werden für die folgenden Analysen ebenfalls beide Messmethoden umgesetzt.

3 Forschungsziele und Fragestellungen

Im vorliegenden Beitrag sollen auf der Basis der Daten von PISA-Ceco 2022 die wechselseitigen Zusammenhänge zwischen FDW, Ke-B, Nu-B und KA untersucht werden. Während entsprechende Variablen zu den BiSta erstmalig in solchen Analysen berücksichtigt werden, sollen unterschiedliche Operationalisierungen von KA – sowohl über Unterrichtsfragebögen als auch über Aufgabenkodierungen – die zusätzliche Untersuchung der differentiellen Validität verschiedener Perspektiven ermöglichen. Konkret werden folgende Forschungsfragen analysiert:

1. *Zusammenhang der Lehrkräftemerkmale*: Wie hängen der Kenntnisstand zu den Bildungsstandards (Ke-B) sowie Einstellungen zu deren Nutzen (Nu-B) mit dem fachdidaktischen Wissen von Mathematiklehrkräften (FDW) zusammen?
2. *Zusammenhang verschiedener Perspektiven von kognitiver Aktivierung (d.h., unterschiedlicher Operationalisierungen)*: Wie hängen Einschätzungen von Lehrkräften sowie von Schülerinnen und Schülern zum wahrgenommenen Niveau des Argumentierens im Unterricht damit zusammen, wie häufig und wie intensiv die

Tab. 1 Beispielitems aus dem PISA-Ceco-Test zum fachdidaktischen Wissen (FDW) von Mathematiklehrkräften mit exemplarischen Lösungsskizzen (vgl. Krauss et al. 2011)

Facette	Beispielitem	Exemplarische Lösungsskizze
Wissen über kognitive Schülerfehler und -schwierigkeiten	*Parallelogramm* Die Fläche des Parallelogramms lässt sich berechnen aus Länge der Grundlinie mal Länge der Höhe.  Geben Sie bitte ein Beispiel eines Parallelogramms (anhand einer Skizze), an dem Schülerinnen und Schüler bei dem Versuch, diese Formel anzuwenden, eventuell scheitern könnten.	Eine mögliche Antwort könnte hier thematisieren, dass Schülerinnen und Schüler Schwierigkeiten bekommen könnten, wenn ein Höhenfußpunkt außerhalb des Parallelogramms liegt, z. B.: 
Wissen über Erklären und Repräsentieren von Inhalten	*Minus 1 mal Minus 1* Eine Schülerin sagt: „Ich verstehe nicht, warum $(-1) \cdot (-1) = 1$ ist“ Bitte skizzieren Sie kurz möglichst viele verschiedene Wege, mit denen Sie der Schülerin diesen Sachverhalt klarmachen könnten.	Auch wenn die Permanenzreihe die Aussage nicht beweist, kann ihr Einsatz hier konzeptuelles Verständnis fördern und innermathematische Vernetzungen herstellen: 
Wissen über das multiple Lösungspotential von Aufgaben	*Quadrat* *Wie ändert sich der Flächeninhalt eines Quadrats, wenn man die Seitenlänge verdreifacht? Begründe deine Antwort!* Bitte schreiben Sie möglichst viele verschiedene Lösungsmöglichkeiten (Begründungen) zu dieser Aufgabe kurz auf.	*z.B.* *Algebraisch:* Flächeninhalt des Ursprungsquadrats: a^2 Flächeninhalt des „neuen“ Quadrats: $(3a)^2 = 9a^2$, also 9-mal so viel. *Geometrisch:* 9-mal das Ursprungsquadrat (z. B. zeichnen)

folgenden Kompetenzen in Klassenarbeiten eingefordert werden: Argumentieren, Problemlösen, Modellieren, begriffliches Denken?
3. *Zusammenhang der Lehrkräftemerkmale mit den Perspektiven von kognitiver Aktivierung*: Welchen Einfluss haben das FDW sowie der Informationsstand über (Ke-B) und die Einstellung zu (Nu-B) den Bildungsstandards auf die verschiedenen Perspektiven von KA?

4 Methode

Im Folgenden werden die untersuchten fachdidaktischen Konstrukte vorgestellt sowie deren Messung erläutert. Die Studienanlage von PISA 2022 sowie wichtige Ergebnisse in Bezug auf die mathematischen Kompetenzen von Schülerinnen und Schülern finden sich in Lewalter et al. (2023). Rationale und Gesamtdesign von PISA-Ceco werden in Schiepe-Tiska et al. (2026) beschrieben.

4.1 Fachdidaktisches Wissen (FDW): *Schülerfehler, Erklären, Aufgabenpotential*

In PISA-Ceco wurde das FDW für alle vier Unterrichtsfächer (Mathematik, Biologie, Chemie, Physik) mit einer aus jeweils sechs Items bestehenden FDW-Kurzskala erfasst. Dabei wurden für jedes Fach in Anlehnung an das FALKO-Projekt (Krauss et al. 2017) die folgenden drei Facetten von FDW operationalisiert: Wissen über typische Schülerfehler und -schwierigkeiten, Wissen über Erklären und Repräsentieren sowie Wissen über das Potential von Lernmaterialien (ausführlich siehe Hohenleitner et al. 2025).

Für das Fach Mathematik konnten konkret sechs Items aus dem FDW-Test der COACTIV-Studie ausgewählt werden (vgl. Tab. 1 für Beispielitems). Mit den zwei Items zur Facette *Wissen über typische Schülerfehler und -schwierigkeiten* werden die Kenntnisse einer Lehrkraft über inhaltliche Schülerkognitionen getestet, wobei hier das Erkennen typischer Probleme von Schülerinnen und Schülern im Vordergrund steht (Metcalfe 2017). Lehrkräfte sollten weiterhin die Fähigkeit besitzen, Schülerinnen und Schülern fachliche Inhalte verständnisorientiert zugänglich zu machen. Dafür ist ein umfangreiches Wissensrepertoire darüber nötig, wie Unterrichtsstoff durch verschiedene Erklärungsansätze beziehungsweise Repräsentationsformen dargestellt werden kann (z. B. Depaepe et al. 2013; Shulman 1987). Dieses wird durch die beiden Items der Facette *Wissen über Erklären und Repräsentieren von Inhalten* überprüft. Da Aufgaben im Mathematikunterricht als Lerngelegenheiten eine zentrale Rolle spielen, testen die zwei Items der Facette *Wissen über das multiple Lösungspotential von Aufgaben* die Fähigkeit einer Lehrkraft, sowohl die auf unterschiedlichen Repräsentationsmöglichkeiten basierende Lösbarkeit von Aufgaben als auch die strukturellen Unterschiede solcher Lösungen zu erkennen (Krauss et al. 2011, 2017).

Die jeweils zwei Items pro Facette wurden aus den 22 FDW-Items des originalen COACTIV-Tests nach folgenden Kriterien für die PISA-Ceco Kurzskala ausgewählt: Hohe Trennschärfen der Items bezüglich des Gesamttests, Ausgewogenheit zwischen

Inhaltsgebieten (Geometrie, Algebra, Arithmetik) sowie eine hohe prädiktive Validität für den Leistungszuwachs in den COACTIV-Daten. In Tab. 1 ist für jede Facette ein Beispielitem inklusive einer Lösungsmöglichkeit skizziert. Eine Auflistung aller in PISA-Ceco verwendeter FDW-Items – auch für die drei naturwissenschaftlichen Fächer – findet sich in Hohenleitner et al. (2025).

Die sechs Items wurden in einen Onlinefragebogen implementiert. Sie konnten prinzipiell ohne Zeitbegrenzung bearbeitet werden, auch wenn von den Lehrkräften pro Item fünf Minuten eingeplant werden sollten und hierzu auch jeweils eine kurze Erinnerung erfolgte (für weitere administrative Details siehe Krauss et al. 2026). Das aus der COACTIV-Studie bereits existierende Kodiermanual wurde für die Kodierung der sechs FDW-Items bei PISA-Ceco auf der Basis aktueller Forschungsergebnisse optimiert und differenziert: Während bei COACTIV beispielsweise pro Antwort einer Lehrkraft maximal ein Punkt vergeben wurde, wurde der Range nun auf 0 bis 2 mögliche Punkte erweitert, um auch im Ansatz richtige oder stichpunktartige Antworten zu berücksichtigen. Die Antworten der Lehrkräfte wurden von zwei unabhängigen Kodiererinnen (Doktorandinnen der Mathematikdidaktik) geratet. Hierbei ergaben sich Werte für Cohens Kappa von $0{,}69 < k < 0{,}86$ (Hohenleitner et al. 2025).

Für die Bildung der FDW-Gesamtskala wurden nur diejenigen Lehrkräfte berücksichtigt, die die sechs Items mit hoher Plausibilität auch ernsthaft bearbeitet haben. Lehrkräfte mit zu kurzer Verweildauer bei der FDW-Skala – d. h., die sich vermutlich nur durch die FDW-Items „durchgeklickt" haben – wurden ebenso ausgeschlossen wie diejenigen, die bei mindestens drei Items keinerlei Antwortversuch zu verzeichnen hatten (vgl. Hohenleitner et al. 2025). Die Reliabilität der sechs Items in der PISA-Ceco Stichprobe betrug $\alpha = 0{,}67$ und ist demnach vergleichbar mit der entsprechenden Reliabilität dieser sechs Items in der COACTIV-Stichprobe von $\alpha = 0{,}65$ (vgl. Krauss et al. 2026).

Tab. 2 Lehrkräfte-Merkmale ($N = 125$): Fachdidaktisches Wissen (FDW), Informationsstand zu den Bildungsstandards (Ke-B) sowie Einstellung zur Nützlichkeit der Bildungsstandards (Nu-B)

Merkmal	Facetten/Beispielitem	α	*M* (*SD*) [range]
FDW: Fachdidaktisches Wissen (6 Items)	Typische Schülerfehler und -schwierigkeiten (2 Items) Erklären und Repräsentieren von Inhalten (2 Items) Multiples Lösungspotential von Aufgaben (2 Items) *Ein Beispielitem für jede Facette befindet sich in* Tab. 1.	0,67	10,20 (4,44) [0; 20]
Ke-B: Informationsstand Bildungsstandards† (1 Item)	Wie gut schätzen Sie Ihren Informationsstand bezüglich der KMK-Bildungsstandards für den mittleren Schulabschluss ein? *(Bitte ordnen Sie sich selbst eine Note von 1 (sehr gut) bis 6 (ungenügend) zu.)*	–	2,99 (1,02) [1; 6]
Nu-B: Einstellung zum Nutzen der Bildungsstandards (9 Items)	Die KMK-Bildungsstandards für den mittleren Schulabschluss helfen mir bei der Unterrichtsplanung. (Skalierung: 1 = trifft gar nicht zu, ..., 4 = trifft völlig zu)	0,80	2,38 (0,47) [1; 3,6]

α Cronbachs Alpha, *M* arithmetischer Mittelwert, *SD* Standardabweichung, range: Spannweite
† Zur leichteren Interpretation wird die Variable Ke-B für alle Zusammenhangsanalysen invertiert.

4.2 Bildungsstandards: Kenntnisstand (Ke-B) und eingeschätzter Nutzen (Nu-B)

Nach der Bearbeitung der FDW-Items hatten die Lehrkräfte im Ceco-Online-instrument Fragen zu ihrer Einschätzung zu den Bildungsstandards zu beantworten (Tab. 2). Der Informationsstand der Lehrkräfte über die Bildungsstandards (Ke-B) wurde mit einem Einzelitem erhoben (Selbsteinschätzung mit einer Note zwischen 1 bis 6). Anschließend gaben die Lehrkräfte ihre Einschätzung zu Items der Skalen „Umgang mit Bildungsstandards: Bildungsstandards als Diagnoseinstrument“ und „Wahrgenommene Nützlichkeit Kompetenzorientierung“ ab (jeweils vierstufige Likertskala; für alle Items siehe Todtenhöfer et al. 2025). Wie durch eine explorative Faktorenanalyse indiziert, wurden diese insgesamt neun Items für den vorliegenden Beitrag zur Skala „Nützlichkeit der Bildungsstandards“ (Nu-B) zusammengefasst (die Reliabilität dieser Skala betrug $\alpha = 0{,}80$; vgl. Tab. 2).

4.3 Aspekte kognitiver Aktivierung (KA)

Zur Operationalisierung der verschiedenen Perspektiven auf KA wird zum einen auf eine Fragebogenskala zum *wahrgenommenen* Anspruch des Argumentierens im Unterricht sowohl aus Schüler- als auch aus Sicht der Lehrkräfte zurückgegriffen (siehe weiter unten: a und b). Zum anderen wird das *Potential* zur kognitiven Aktivierung über die in den Klassenarbeiten eingesetzten Aufgaben abgeschätzt (vgl. c, d, e und f). Als Indikatoren für KA werden dabei die von externen Raterinnen und Ratern kodierten Dimensionen „Argumentieren“, „innermathematisches Problemlösen“, „Modellieren“ und „Anteil begrifflicher Aufgaben (vs. rein rechnerischer Aufgaben)“ verwendet. Zusammengefasst werden demnach betrachtet:

Wahrgenommene kognitive Aktivierung (internationale PISA-Fragebögen):

a) Schülerskala Argumentieren (Fragebogen-Skala)
b) Lehrkraftskala Argumentieren (Fragebogen-Skala)

Potential zur kognitiven Aktivierung (Kodierung von Klassenarbeiten):

c) Aufgabendimension „Argumentieren“
d) Aufgabendimension „innermathematisches Problemlösen“
e) Aufgabendimension „(außermathematisches) Modellieren“
f) Aufgabendimension „Anteil begrifflicher Aufgaben“

Da das mathematische Argumentieren als fachlich gehaltvolle und kognitiv aktivierende Tätigkeit gilt (Blum et al. 2001; Jahnke et al. 2023), wurde für die Wahrnehmungsperspektive (a und b) die aus neun Items bestehende Skala „Mathematisches Argumentieren“ aus dem internationalen PISA-Fragebogen gewählt (Müller et al. 2025). Diese war für Lehrkräfte bzw. Schülerinnen und Schüler jeweils identisch aufgebaut, das heißt, Iteminhalte und Formulierungen waren abgesehen von der Beobachtungsperspektive gleich (vgl. Beispielitems in Tab. 3, oben). Das mathematische Argumentieren (c) war inhaltlich auch eine der betrachteten Dimensionen im Hinblick auf die Einschätzung des Potentials für KA über die eingesammelten

Tab. 3 Betrachtete Indikatoren kognitiver Aktivierung (KA): oben: Schülerinnen und Schüler (Skala Argumentieren), Lehrkräftefragebogen (Skala Argumentieren); unten: Kodierung von in Klassenarbeiten eingesetzten Aufgaben (geratete Dimensionen: Argumentieren, innermathematisches Problemlösen, Modellieren, Anteil begrifflicher Aufgaben)

Merkmal	Fragebogen/Beispielitem/Skalierung	*M* (*SD*)
Mathematisches Argumentieren ($N = 59$ Klassen, 9 Items, $\alpha = 0{,}84$)	*Schülerskala* Die Lehrkraft forderte uns auf, zu erklären, wie wir ein mathematisches Problem gelöst haben. *Skalierung: 1=(fast) nie, ..., 5=in (fast) jeder Stunde*	3,23 (0,42)
Mathematisches Argumentieren ($N = 81$ Lehrkräfte, 9 Items, $\alpha = 0{,}82$)	*Lehrkraft-Skala* Ich forderte die Schülerinnen und Schüler auf, zu erklären, wie sie ein mathematisches Problem gelöst haben. *Skalierung: 1=(fast) nie, ..., 5=in (fast) jeder Stunde*	3,48 (0,64)
Aufgaben (je $N = 70$ Lehrkräfte)	*Eingesammelte Klassenarbeiten* *Kodierung jeweils:* *Score 1 für mind. 1 zutreffenden Indikator, sonst Score 0*	*Anteil Aufg. Score 1 (SD)*
Argumentieren	Wiedergabe/Anwendung von Routineargumentationen, z. B. Sätze/Verfahren/Herleitungen	0,04 (0,05)
	Geben von einfachen rechnerischen Begründungen	
	Argumentation mit Alltagswissen	
	Nachvollziehen/Erläutern/Entwickeln überschaubarer mehrschrittiger Argumentationen	
	Beschreibung/Begründung von Lösungsweg	
	Bewertung von Ergebnissen bzgl. ihrer Anwendungskontexte	
	Erläuterung von Zusammenhängen/Ordnungen/Strukturen	
	Nutzen/Erläutern/Entwickeln komplexer Argumentationen	
	Bewertung verschiedener Argumentationen	
	Stellen von Fragen, die für Mathematik charakteristisch sind	
	Äußerung begründeter mathematischer Vermutungen	
Innermathematisches Problemlösen	Entwurf/Umsetzung einer umfassenden Lösungsstrategie	0,004 (0,01)
	Treffen allgemeiner Aussagen zur Aufgabe	
	Kritische Reflexion über den Lösungsweg	
Modellieren	Übersetzungsprozesse zwischen „Realität“ und Mathematik	0,15 (0,14)
	Verwendung/Vergleichen/Reflexion/Kritisches Beurteilen mathematischer Modelle	
	Überprüfung von Modellergebnissen	
	Entwicklung komplexer mathematischer Modelle	
Begriffliche Aufgaben	Wissen über mathematische Konzepte/Beziehungen	0,27 (0,13)
	keine Ausführungen/Berechnungen notwendig	

N (maximale) Stichprobengröße/Anzahl Lehrkräfte, *M* arithmetischer Mittelwert, *SD* Standardabweichung, α Cronbachs Alpha.

Klassenarbeiten, was zumindest im Prinzip eine wechselseitige Validierung ermöglicht (zu den anderen drei kodierten Dimensionen d–f lagen in den internationalen Fragebögen keine vergleichbaren Schüler- bzw. Lehrkraft-Skalen vor).

Alle vier anhand der Klassenarbeiten kodierten Dimensionen (c, d, e, f) werden in der Literatur häufig als kognitiv aktivierende Tätigkeiten bezeichnet (vgl. 2.3).

Die Dimensionen c, d und e entsprechen dabei den Kompetenzen 1, 2 und 3 der Bildungsstandards (vgl. 2.2). Begriffliches Arbeiten (f) im Mathematikunterricht wird ebenfalls als fachlich gehaltvoll und kompetenzorientiert betrachtet, da dieses konzeptuelles Argumentieren statt nur rein rechnerisches Arbeiten erfordert (Neubrand et al. 2011). Tab. 3 listet jeweils exemplarische Indikatoren in Bezug auf die vier Dimensionen auf. Bei mindestens einem zutreffenden Indikator wurde pro Aufgabe auf dieser Dimension Code 1 (zutreffend) vergeben, ansonsten Code 0 (nicht zutreffend). Anschließend wurde dann der Anteil des Codes 1 unter allen von einer Lehrkraft in ihren Klassenarbeiten eingesetzten Aufgaben bestimmt (vgl. Tab. 3). Eine ausführliche Beschreibung des methodischen Vorgehens bei der Kodierung der Klassenarbeiten – inklusive des vollständigen Kodiermanuals – findet sich in Heinle et al. (2025).

4.4 Stichprobe

Von den insgesamt $N = 199$ Mathematiklehrkräften, die an der Onlinebefragung von PISA-Ceco teilnahmen, öffneten 150 Lehrkräfte (zumindest teilweise) den darin enthaltenen FDW-Test. Von diesen können wiederum 125 als ernsthafte Bearbeitungen gesehen werden (vgl. 4.1). Von den $N = 125$ Lehrkräften (42,4 % weiblich, 40,8 % männlich, 16,8 % ohne Angabe) unterrichten 63 an Gymnasien und 61 an nicht-gymnasialen Schulformen. Die durchschnittliche Berufserfahrung dieser für den vorliegenden Beitrag gewählten Referenzstichprobe beträgt 10,7 Jahre ($SD = 7{,}75$).

Für die drei Lehrkräfteskalen wurden jeweils manifeste Skalenwerte gebildet (Tab. 2: FDW, Nu-B bzw. Tab. 3: Skala Argumentieren). Die verfügbaren Antworten der Schülerinnen und Schüler (Tab. 3: Skala Argumentieren) wurden pro Item auf Klassenebene aggregiert und anschließend daraus ebenfalls ein manifester Skalenwert gebildet (für das sog. within construct rotated design in PISA vgl. OECD 2023). Zur optimalen Ausnutzung der PISA-Ceco Stichprobe basieren die Mittelwerte und Standardabweichungen (in Tab. 2 und 3) auf der jeweils pro Konstrukt/Dimension maximal vorliegenden Stichprobe.

Da nicht allen Lehrkräften eine der beiden gezogenen PISA-Klassen zugeordnet werden kann und auch nicht alle Lehrkräfte, die den FDW-Test bearbeitet haben, Klassenarbeiten eingereicht haben, reduziert sich die Stichprobe je nach Kombination der in die statistischen Zusammenhangsanalysen eingehenden Konstrukte (Tab. 4, 5, 6, 7 und 8).

5 Ergebnisse

5.1 Deskriptive Ergebnisse

Vor Beantwortung der Forschungsfragen sollen zunächst einige deskriptive und psychometrische Ergebnisse zusammengefasst werden (Tab. 2 und 3).

Tab. 2 kann entnommen werden, dass die Lehrkräfte ihren Kenntnisstand bezüglich der Bildungsstandards insgesamt als befriedigend einschätzen (Einzelitem

Ke-B; Durchschnittsnote 2,99). Auch die Beurteilung deren Nutzens liegt im mittleren Bereich ($M_{\text{Nu-B}} = 2{,}38$ bei vierstufiger Skala).

Tab. 3 präsentiert deskriptive Ergebnisse zu den verschiedenen Operationalisierungen von KA. In Bezug auf die Wahrnehmungsperspektive wurden Umfang und Anspruch des „Mathematischen Argumentierens" im Unterricht sowohl von Lehrkräften ($M = 3{,}48$) als auch von Schülerinnen und Schülern ($M = 3{,}23$) als durchschnittlich beurteilt (mit jeweils guten Reliabilitäten). Im Hinblick auf die Kodierung der Klassenarbeiten fällt auf, dass bezüglich der Dimensionen „Argumentieren" und „innermathematisches Problemlösen" beinahe alle Aufgaben Code 0 erhalten haben. Weiterhin wird deutlich, dass 15 % der Aufgaben Modellierungsprozesse erfordern und dass es sich bei etwa 75 % der betrachteten Aufgaben um reine Rechenaufgaben handelt (der Anteil begrifflicher Aufgaben betrug 27 %; Tab. 3).

5.2 Forschungsfrage 1: Zusammenhang der Lehrkraftmerkmale FDW, Ke-B und Nu-B

Für eine leichtere Interpretierbarkeit der Ergebnisse wurde die Notenskala des Einzelitems Ke-B für alle folgenden Analysen invertiert. Bezüglich Forschungsfrage 1 ergibt sich zwischen eingeschätztem Kenntnisstand (Ke-B) und eingeschätztem Nutzen der Bildungsstandards (Nu-B) eine auf dem 1 %-Niveau signifikante bivariate Pearson-Korrelation von $r = 0{,}36$ ($p < 0{,}01$). Beide Konstrukte sind jedoch unabhängig vom FDW der Lehrkräfte: Der Korrelationskoeffizient für den Zusammenhang zwischen FDW und Ke-B beträgt $r = 0{,}15$ ($p = 0{,}10$) und der entsprechende Koeffizient für den Zusammenhang zwischen FDW und Nu-B ist $r = 0$ ($p = 0{,}98$). Das FDW von Mathematiklehrkräften ist also unabhängig davon, wie diese den Nutzen der BiSta einschätzen. Der fehlende Zusammenhang zwischen FDW und den beiden BiSta-Variablen schließt somit auch jegliche wechselseitige kausale Beeinflussung aus.

5.3 Forschungsfrage 2: Zusammenhang der verschiedenen Perspektiven auf KA

Tab. 4 enthält die wechselseitigen bivariaten Korrelationen zwischen den betrachteten Indikatoren von KA (wobei auf die Dimension Problemlösen aufgrund des geringen Anteils von 0,004 in Tab. 3 verzichtet wurde). Auffällig ist die faktische Null-Korrelation zwischen den Einschätzungen der Lehrkräfte und derjenigen der Schülerinnen und Schüler in Bezug auf die Anforderungen des mathematischen Argumentierens im Unterricht ($r = -0{,}02$, $p = 0{,}88$). Dieser fehlende Zusammenhang ergibt sich im Übrigen auch auf der Ebene der neun – bis auf die Perspektive jeweils identisch formulierten – Einzelitems. Für Hinweise zur Validität dieser Schüler- bzw. Lehrkrafteinschätzungen sollen im Folgenden zusätzlich die Aufgabenkodierungen betrachtet werden.

Für die kodierte Aufgabendimension Argumentieren ergibt sich ein kleiner positiver Zusammenhangseffekt mit der Schülerwahrnehmung ($r = 0{,}12$, $p = 0{,}53$) und ein kleiner negativer Effekt mit der entsprechenden Wahrnehmung der Lehrkräfte ($r = -0{,}13$, $p = 0{,}39$). Beide Korrelationen sind – auch aufgrund der jeweiligen

Tab. 4 Bivariate Zusammenhänge (Pearson Produkt-Moment-Korrelationen) zwischen den Indikatoren kognitiver Aktivierung (jeweils basierend auf der für die beiden Konstrukte maximal vorliegenden Stichprobe)

Methode	Fragebögen	Aufgabenkodierungen		
Perspektive	PISA-Schüler:innen	PISA-Ceco-Lehrkräfte		
Merkmal	Skala Mathematisches Argumentieren	Skala Mathematisches Argumentieren	Argumentieren	Modellieren
FB-Skala Lehrkräfte	–	–	–	–
Mathematisches Argumentieren	–0,02 ($p = 0{,}88$; $N = 48$)	–	–	–
Aufgabenkodierungen	–	– –	–	–
Argumentieren	0,12 ($p = 0{,}53$; $N = 31$)	–0,13 ($p = 0{,}39$; $N = 46$)	–	–
Modellieren	–0,14 ($p = 0{,}45$; $N = 31$)	0,12 ($p = 0{,}43$; $N = 46$)	0,26* ($p = 0{,}03$; $N = 70$)	–
Anteil begrifflicher Aufgaben	0,39* ($p = 0{,}03$; $N = 31$)	0,04 ($p = 0{,}80$; $N = 46$)	0,28* ($p = 0{,}02$; $N = 70$)	–0,28* ($p = 0{,}02$; $N = 70$)

N Stichprobengröße (Anzahl Lehrkräfte bzw. Klassen), signifikanter Zusammenhang: $^{*}p \leq 0{,}05$; $^{**}p \leq 0{,}01$.

Einschränkungen der Stichprobe – allerdings nicht signifikant. Da mathematisches Argumentieren prinzipiell nur bei begrifflichen und nicht bei rechnerischen Aufgaben erforderlich ist, ist auch der diesbezügliche Zusammenhang mit den Wahrnehmungen von Interesse. Der Anteil begrifflicher Aufgaben in den Klassenarbeiten korreliert signifikant mit der Wahrnehmung der Schülerinnen und Schüler zum Argumentieren ($r = 0{,}39$, $p = 0{,}03$), aber nicht mit der entsprechenden Einschätzung der Lehrkräfte ($r = 0{,}04$, $p = 0{,}80$). Daher scheinen in diesem Fall die Einschätzungen der Schülerinnen und Schüler valider zu sein als die Einschätzungen der Lehrkräfte. Für die Validität der Aufgabenkodierungen wiederum sprechen die erwartungskonformen Zusammenhänge der kodierten Dimension Argumentieren mit dem Modellieren ($r = 0{,}26$, $p = 0{,}03$) beziehungsweise mit dem Anteil begrifflicher Aufgaben ($r = 0{,}28$, $p = 0{,}02$), da sowohl bei realitätsbezogenen Modellierungsaufgaben als auch beim Arbeiten mit mathematischen Begriffen oft auf Argumente zurückgegriffen werden muss.

5.4 Forschungsfrage 3: Einfluss von FDW, Ke-B und Nu-B auf KA

Interessant ist nun die gemeinsame Betrachtung von Lehrkraftmerkmalen und KA (Tab. 5, 6, 7 und 8). Hierfür wurden Regressionsmodelle geschätzt, in denen FDW, Ke-B und Nu-B den jeweiligen Indikator von KA vorhersagen (jeweils links Modell a). Neben dem Problemlösen wurde hierbei weiterhin auf die Wahrnehmung des Argumentationsniveaus durch die Lehrkräfte verzichtet (in der Tat beträgt die bivariate Korrelation der Lehrkräfte-Skala mit FDW $r = -0{,}23$). Alle Modelle (Tab. 5,

Tab. 5 Hierarchische lineare Regressionsmodelle zur abhängigen Variable „Mathematisches Argumentieren“ (Schüler:innen-Skala) und den unabhängigen Variablen fachdidaktisches Wissen (FDW) von Lehrkräften, deren Informationsstand von (Ke-B) und Einstellung zum Nutzen (Nu-B) der Bildungsstandards (Modell 1a) bzw. unter zusätzlicher Kontrolle der Schulform (0 = nicht-gymnasial, 1 = gymnasial; Modell 1b), jeweils unter Berücksichtigung der Klassenstruktur (ICC = 14,2 %)

	Modell 1a (ohne Schulform)					Modell 1b (mit Schulform)				
	β	*SE*	*df*	*t*	*p*	*β*	*SE*	*df*	*t*	*p*
Konstante	–0,01	0,06	598,97	–0,23	0,82	–0,17	0,08	428,52	–2,23	0,03
FDW	0,12	0,06	692,94	2,12	0,03	0,04	0,06	739,52	0,75	0,45
Ke-B †	–0,09	0,06	448,40	–1,48	0,14	–0,07	0,06	394,23	–1,16	0,25
Nu-B	0,06	0,06	674,11	0,91	0,36	0,05	0,06	647,57	0,90	0,37
Schulform	–	–	–	–	–	0,36	0,12	683,36	3,16	<0,01
Marg. R^2 | kond. R^2	0,02 | 0,13					0,05 | 0,13				

$N_{\text{Lehrkräfte}} = 57$, $N_{\text{Klassen}} = 63$; in einzelnen Variablen fehlende Werte wurden mittels multipler Imputation ($n = 50$) geschätzt. *ICC* = Intraklassenkorrelation; β = standardisierte Regressionskoeffizienten; *SE* = Standardfehler; *df* = Freiheitsgrade; *t* = *t*-Wert; *p* = Risiko für Fehler 1. Art; R^2 Determinationskoeffizient.
† Zur leichteren Interpretation wird die Variable Ke-B für alle Zusammenhangsanalysen invertiert (vgl. Tab. 2)

Tab. 6 Multiple lineare Regressionsmodelle zur abhängigen Variable „Argumentieren“ (Aufgaben-Rating) und den unabhängigen Variablen fachdidaktisches Wissen (FDW) von Lehrkräften, deren Informationsstand von (Ke-B) und Einstellung zum Nutzen (Nu-B) der Bildungsstandards (Modell 2a) bzw. unter zusätzlicher Kontrolle der Schulform (0 = nicht-gymnasial, 1 = gymnasial; Modell 2b)

	Modell 2a (ohne Schulform)				Modell 2b (mit Schulform)			
	β	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Konstante	–0,05	0,19	–0,29	0,77	–0,44	0,27	–1,59	0,12
FDW	0,16	0,17	0,95	0,35	0,02	0,18	0,10	0,92
Ke-B †	0,03	0,18	0,15	0,89	–0,08	0,18	–0,46	0,65
Nu-B	–0,11	0,20	–0,58	0,57	–0,14	0,19	–0,75	0,46
Schulform	–	–	–	–	0,72	0,39	1,85	0,07
R^2 | korr. R^2	0,04 | 0,00				0,14 | 0,02			

$N = 33$ Klassen; β = standardisierte Regressionskoeffizienten; *SE* = Standardfehler; *t* = *t*-Wert; *p* = Risiko für Fehler 1. Art; R^2 Determinationskoeffizient.
† Zur leichteren Interpretation wird die Variable Ke-B für alle Zusammenhangsanalysen invertiert (vgl. Tab. 2)

6, 7 und 8) wurden anschließend zusätzlich noch unter Kontrolle der Schulform gerechnet (jeweils rechts Modell b).

Während das FDW der Lehrkräfte in allen acht Modellen deskriptiv einen positiven Regressionskoeffizienten aufweist, variieren die diesbezüglichen Vorzeichen des Kenntnisstands (Ke-B) bzw. der eingeschätzten Nützlichkeit (Nu-B) der BiSta unsystematisch. Aufgrund der kleinen Stichprobenzahlen vor allem der Modelle mit Aufgabenkodierungen als abhängige Variable – hier verbleiben nur noch 33 Klassen (Tab. 6, 7 und 8) – gibt es lediglich bei den Modellen zur Vorhersage der Schülerwahrnehmung (auf der Basis von 63 Klassen) signifikante Koeffizienten (Tab. 5). Hier ist das FDW ein signifikanter Prädiktor für das wahrgenommene Niveau des Argumentierens (Modell 1a), dieser Effekt kann jedoch beinahe komplett durch die Schulform erklärt werden (Modell 1b). Dieser Befund wird – zumindest deskriptiv – durch die beiden entsprechenden in Tab. 6 dargestellten Modelle 2a und 2b gestützt,

Tab. 7 Multiple lineare Regressionsmodelle zur abhängigen Variable „Modellieren“ (Aufgaben-Rating) und den unabhängigen Variablen fachdidaktisches Wissen (FDW) von Lehrkräften, deren Informationsstand von (Ke-B) und Einstellung zum Nutzen (Nu-B) der Bildungsstandards (Modell 3a) bzw. unter zusätzlicher Kontrolle der Schulform (0 = nicht-gymnasial, 1 = gymnasial; Modell 3b)

	Modell 3a (ohne Schulform)				Modell 3b (mit Schulform)			
	β	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Konstante	–0,09	0,18	–0,52	0,61	0,03	0,28	0,11	0,92
FDW	0,30	0,16	1,84	0,08	0,34	0,18	1,89	0,07
Ke-B †	0,09	0,18	0,53	0,60	–0,08	0,18	–0,42	0,68
Nu-B	0,05	0,19	0,25	0,81	0,06	0,19	0,29	0,77
Schulform	–	–	–	–	–0,23	0,39	–0,59	0,56
R^2 | korr. R^2	0,13 | 0,04				0,14 | 0,01			

N = 33 Klassen; β = standardisierte Regressionskoeffizienten; *SE* = Standardfehler; *t* = *t*-Wert; *p* = Risiko für Fehler 1. Art; R^2 Determinationskoeffizient.
† Zur leichteren Interpretation wird die Variable Ke-B für alle Zusammenhangsanalysen invertiert (vgl. Tab. 2)

Tab. 8 Multiple lineare Regressionsmodelle zur abhängigen Variable „Anteil begrifflicher Aufgaben“ (Aufgaben-Rating) und den unabhängigen Variablen fachdidaktisches Wissen (FDW) von Lehrkräften, deren Informationsstand von (Ke-B) und Einstellung zum Nutzen (Nu-B) der Bildungsstandards (Modell 4a) bzw. unter zusätzlicher Kontrolle der Schulform (0 = nicht-gymnasial, 1 = gymnasial; Modell 4b)

	Modell 4a (ohne Schulform)				Modell 4b (mit Schulform)			
	β	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Konstante	–0,04	0,19	–0,19	0,85	–0,08	0,29	–0,26	0,79
FDW	0,19	0,17	1,15	0,26	0,18	0,19	0,94	0,36
Ke-B †	–0,13	0,18	–0,72	0,48	–0,13	0,19	–0,67	0,51
Nu-B	0,06	0,19	0,30	0,77	0,06	0,20	0,28	0,78
Schulform	–	–	–	–	0,08	0,41	0,19	0,86
R^2 | korr. R^2	0,06 | 0,00				0,06 | 0,00			

N = 33 Klassen; β = standardisierte Regressionskoeffizienten; *SE* = Standardfehler; *t* = *t*-Wert; *p* = Risiko für Fehler 1. Art; R^2 Determinationskoeffizient.
† Zur leichteren Interpretation wird die Variable Ke-B für alle Zusammenhangsanalysen invertiert (vgl. Tab. 2)

in denen das Argumentationspotential alternativ über die Aufgaben eingeschätzt wird.

Relativ konstant unter der Kontrolle der Schulform bleibt dagegen der Effekt des FDW in Bezug auf die Aufgabendimension Modellieren (Tab. 7: Modell 3a und 3b), auch wenn beide Effekte (gerade) nicht signifikant sind. Der Anteil begrifflicher Aufgaben konnte durch keine der drei Lehrkraftvariablen (FDW, Ke-B, Nu-B) vorhergesagt werden (Tab. 8).

6 Diskussion

6.1 Zusammenfassung und Interpretation

Im vorliegenden Beitrag wurden Merkmale der Mathematiklehrkräfte von PISA-Ceco sowie ausgewählte fachspezifische Aspekte aus deren Unterricht untersucht. Für die zentrale Professionswissenskategorie FDW wurde eine Kurzskala aus sechs

offenen Items implementiert, die in einen Onlinefragebogen eingebettet war. Im Rahmen dieses Fragebogens hatten die Lehrkräfte weiterhin subjektive Einschätzungen zu den Bildungsstandards im Fach Mathematik abzugeben.

Bei der Beantwortung von Forschungsfrage 1 zeigte sich, dass das FDW der Lehrkräfte in keinem Zusammenhang zum Kenntnisstand über die Bildungsstandards (Ke-B) beziehungsweise zu den Einschätzungen über deren Nutzen (Nu-B) stand, das FDW von Mathematiklehrkräften variiert vielmehr unabhängig von den beiden BiSta-Variablen. Da die Bildungsstandards offensichtlich starke fachdidaktische Bezüge aufweisen (Blum et al. 2006), ist der fehlende Zusammenhang von Ke-B und Nu-B zum FDW ein überraschender Befund, da hierdurch auch kausale Effekte – und zwar in beide Richtungen – weitestgehend ausgeschlossen werden können. Denkbar wäre allerdings, dass die Messung subjektiver Einschätzungen nicht ausreicht, um solche möglichen wechselseitigen Beeinflussungen sichtbar zu machen (vgl. 6.2).

In Bezug auf den Unterricht wurden verschiedene Operationalisierungen von KA betrachtet, wobei explizit zwischen der *wahrgenommenen* KA (Fragebogenskala zum Argumentieren) und dem *Potential* für KA (jeweils in Klassenarbeiten kodierte Dimensionen: Argumentieren, Problemlösen, Modellieren, Anteil begrifflicher Aufgaben) unterschieden wurde. In Bezug auf Forschungsfrage 2 waren die Einschätzungen zum mathematischen Argumentieren im Unterricht zwischen Lehrkräften und Schülerinnen bzw. Schülern unkorreliert. Daten aus den eingesammelten Klassenarbeiten legen nahe, dass das Schülerurteil in diesem Fall als valider einzuschätzen ist. Im Lichte der bislang vielfältigen und teils widersprüchlichen Befunde zur differentiellen Validität der Einschätzungen von Schülerinnen und Schülern sowie Lehrkräften (z. B. Clausen 2002; Fauth et al. 2020) sind die plausibleren Korrelationen der Wahrnehmung der Schülerinnen und Schüler zu den Aufgabenkodierungen bemerkenswert, da diesen gelegentlich nachgesagt wird, gerade Aspekte von KA nicht valide einschätzen zu können (Kunter und Baumert 2006). Möglicherweise trifft dies aber nur für explizite Tiefenstrukturen zu, die mit dem didaktisch planvollen Handeln sowie den (verborgenen) unterrichtlichen Zielen der Lehrkräfte assoziiert sind. Die Bedeutung der in der vorliegenden Studie implementierten Items zum Argumentieren sind dagegen unmittelbar transparent und direkt verständlich (für alle neun Items siehe Müller et al. 2025).

Die wechselseitigen Korrelationen der betrachteten Aufgabendimensionen sprechen insgesamt für die Validität dieser Kodierungen, auch wenn sowohl die Notwendigkeit zum mathematischen Argumentieren als auch zum Problemlösen in solchen Aufgaben kaum vertreten zu sein scheint. Da Argumentieren und Problemlösen nicht nur seit zwanzig Jahren in den Bildungsstandards explizit hervorgehoben werden (als K1 und K2), sondern seit vielen Jahren auch deutschlandweit in Lehrplänen und Schulbüchern abgebildet sind, ist das geringe Potential zur kognitiven Aktivierung in Bezug auf diese beiden prozessbezogenen Kompetenzen in Klassenarbeiten durchaus erstaunlich (das Niveau entspricht in etwa dem in der COACTIV-Studie gemessenen, Neubrand et al. 2011). Möglicherweise lässt sich durch bildungspolitische Maßnahmen wie z. B. Standardsetzungen aber auch weniger Einfluss auf die Unterrichtsqualität nehmen als gewünscht (siehe hierzu auch den entsprechenden Vergleich zwischen COACTIV und PISA-Ceco in Krauss et al. 2026).

Zur Beantwortung von Forschungsfrage 3 wurden in Regressionsmodellen die drei Konstrukte der Lehrkraftebene (FDW, Ke-B, Nu-B) als unabhängige Variablen gemeinsam zur Vorhersage der verschiedenen Indikatoren von KA implementiert, wobei auf die Lehrkraftskala zum Argumentieren (aufgrund mangelnder Validität) sowie auf die Aufgabendimension Problemlösen (aufgrund eines zu geringen Anteils entsprechender Aufgaben) als abhängige Variablen verzichtet wurde. Auch wenn sich aufgrund geringer Stichprobengrößen nur wenige signifikante Regressionskoeffizienten ergeben, legen die Modellergebnisse nahe, dass eher das FDW der Lehrkräfte – nicht aber die erhobenen Variablen zu den Bildungsstandards – mit den Indikatoren von KA zusammenhängen.

Die Rolle des FDW für die vier abhängigen Variablen der KA in den Regressionsmodellen wird teilweise von der Schulform beeinflusst. Vor allem der Effekt von FDW auf das Argumentieren kann dadurch erklärt werden, dass diese Tätigkeit prinzipiell häufiger im Gymnasium praktiziert wird (vgl. die positiven Koeffizienten der Schulform in Tab. 5 und 6) und gymnasiale Lehrkräfte über höheres FDW verfügen. Für Aufgabendimensionen ohne Schulformunterschiede (z. B. den Anteil begrifflicher Aufgaben; Tab. 8) oder für Tätigkeiten, die tendenziell sogar häufiger in nicht-gymnasialen Schulformen vorkommen (Modellieren; Tab. 4) bleibt der Effekt des FDW auch bei Kontrolle der Schulform dagegen relativ konstant.

6.2 Limitationen

Die Messung von FDW erfolgte im vorliegenden Beitrag lediglich mit sechs Items. Krauss et al. (2026) illustrieren jedoch, dass sich zumindest in den COACTIV-Daten ähnliche Ergebnisse zeigen, wenn man diese aus sechs Items bestehende FDW-Skala mit der entsprechenden Originalskala aus 22 Items vergleicht. Auch die Bildungsstandards wurden lediglich über zwei subjektive Einschätzungen adressiert, nämlich zum eigenen Kenntnisstand und über deren Nutzen. Dies sind sicherlich eher rudimentäre Aspekte, die nicht als eine Evaluation der Bildungsstandards interpretiert werden dürfen. Immerhin wird einerseits klar, dass der Nutzen umso höher eingeschätzt wird, je größer der diesbezügliche Kenntnisstand ist. Andererseits wird aber auch deutlich, dass keine der beiden Variablen einen Zusammenhang zum FDW oder zur kognitiven Aktivierung aufweist.

Kognitive Aktivierung ist ein äußerst vielschichtiges Konzept, für das zahlreiche alternative Messmethoden, inhaltliche Ausrichtungen sowie Beobachterperspektiven denkbar sind (Praetorius et al. 2024). Insofern unterliegen ausgewählte Indikatoren von KA prinzipiell einer gewissen Willkür. Im vorliegenden Beitrag wurde dieser Vielschichtigkeit durch die Implementierung sechs verschiedener Operationalisierungen begegnet. Eine gemeinsame Analyse fachspezifischer und überfachlicher Aspekte von KA auf der Basis der PISA-Ceco Daten findet sich in Heinle et al. (2026).

Die größte Einschränkung ist sicherlich die verfügbare Stichprobengröße, die umso geringer ausfällt, je größer die Zahl der gleichzeitig betrachteten Konstrukte wird. Insofern können vor allem die Ergebnisse der Modelle in den Tab. 6, 7 und 8 – jeweils auf der Basis von 33 Klassen – nur als Hinweise verstanden werden. Krauss et al. (2026) zeigen, dass auf der Basis der PISA-Ceco Daten jedoch auch abwei-

chende Indikatoren von KA – beispielsweise basierend auf Items zum Alltagsbezug oder zur Behandlung alternativer Lösungswege – in ähnlicher Größenordnung von FDW vorhergesagt werden können.

Autorenrollen Alle Autorinnen und Autoren waren gleichermaßen an der Erstellung dieses Beitrags beteiligt.

Förderung Die Erhebung der Daten und die Dokumentation der Erhebungsinstrumente wurden vom Bundesministerium für Bildung und Forschung (BMBF) und der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) gefördert (ZIB2022).

Open Access funding enabled and organized by Projekt DEAL.

Datenverfügbarkeit Die im Rahmen der Studie erhobenen und analysierten Daten wurden beim Forschungsdatenzentrum (FDZ) des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) hinterlegt. Zum Zeitpunkt der Veröffentlichung befinden sich die Daten noch im Aufbereitungs- und Prüfprozess und sind daher noch nicht öffentlich zugänglich. Nach Abschluss dieses Prozesses werden die Daten gemäß den Zugangsregelungen des FDZ IQB zur Nutzung bereitgestellt.

Einhaltung ethischer Richtlinien

Interessenkonflikt S. Hohenleitner, S. Krauss, N. Stegmüller und A. Lindl geben an, dass kein Interessenskonflikt besteht.

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

Albers, A., & Jude, N. (2025). *Blickpunkt Bildungsmonitoring – Bilanzen und Perspektiven*. Beltz Juventa. https://doi.org/10.3262/978-3-7799-7843-5.

Anderson, J., & Taner, G. (2023). Building the expert teacher prototype: A metasummary of teacher expertise studies in primary and secondary education. *Educational Research Review*, *38*, 100485. https://doi.org/10.1016/j.edurev.2022.100485.

Ball, D. L., & Hill, H. C. (2023). Das Professionswissen von Mathematiklehrkräften in der Grundschule. In S. Krauss & A. Lindl (Hrsg.), *Professionswissen von Mathematiklehrkräften: Implikationen aus der Forschung für die Praxis* (S. 231–257). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64381-5_7.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K. J., & Weiß, M. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich.* Leske + Budrich. https://doi.org/10.1007/978-3-322-83412-6.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157.

Begrich, L., Praetorius, A. K., Decristan, J., Fauth, B., Göllner, R., Herrmann, C., & Kunter, M. (2023). Was tun? Perspektiven für eine Unterrichtsqualitätsforschung der Zukunft. *Unterrichtswissenschaft*, *51*(1), 63–97. https://doi.org/10.1007/s42010-023-00163-4.

Blömeke, S., Bremerich-Vos, A., Haudeck, H., Kaiser, G., Nold, G., & Schwippert, K. (2011). *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT*. Waxmann Verlag.

Blömeke, S., Bremerich-Vos, A., & Kaiser, G. (2013). *Professionelle Kompetenzen im Studienverlauf: Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerausbildung aus TEDS-LT*. Waxmann Verlag.

Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, *79*(1), 101600. https://doi.org/10.1016/j.learninstruc.2022.101600.

Blum, W. (2001). Was folgt aus TIMSS für Mathematikunterricht und Mathematiklehrerausbildung? In BMBF (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 75–83). Bonn.

Blum, W., Drüke-Noe, C., Hartung, R., & Köller, O. (2006). *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen*. Cornelsen.

Böttcher, W., & Dicke, J. N. (2008). Implementation von Standards. Empirische Ergebnisse einer Umfrage bei Deutschlehrern. In W. Böttcher, W. Bos, H. Döbert & H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive: Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB)*. Waxmann.

Bruder, R., Leuders, T., & Büchter, A. (2008). *Mathematikunterricht entwickeln. Bausteine für kompetenzorientiertes Unterrichten*. Cornelsen Scriptor.

Charalambous, C. Y., Hill, H. C., Chin, M. J., & McGinn, D. (2020). Mathematical content knowledge and knowledge for teaching: exploring their distinguishability and contribution to student learning. *Journal of Mathematics Teacher Education*, *23*(6), 579–613. https://doi.org/10.1007/s10857-019-09443-2.

Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Waxmann Verlag. https://doi.org/10.25656/01:4014.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, *34*, 12–25. https://doi.org/10.1016/j.tate.2013.03.001.

Fauth, B., Göllner, R., Lenske, G., Praetorius, A. K., & Wagner, W. (2020). Who sees what? Conceptual considerations on the measurement of teaching quality from different perspectives (S. 138–155). https://doi.org/10.25656/01:25870.

Frenken, L., Klinger, M., Greefrath, G., & Barzel, B. (2024). Kognitives Potenzial deutscher Abituraufgaben. *Journal für Mathematik-Didaktik*, *45*(8), 1–37. https://doi.org/10.1007/s13138-024-00233-9.

Frenzel, J. (2018). *Mathematiklehrkräfte im Implementationsprozess der nationalen Bildungsstandards* (Dissertation). Humboldt-Universität Berlin. https://doi.org/10.18452/19593.

Fukaya, T., Nakamura, D., Kitayama, Y., & Nakagoshi, T. (2025). A systematic review and meta-analysis of research on mathematics and science pedagogical content knowledge: Exploring its associations with teacher and student variables. *Teaching and Teacher Education*, *155*, 104881. https://doi.org/10.1016/j.tate.2024.104881.

Heinle, A., Todtenhöfer, P., Dümig, P., Reinhold, F., Reiss, K., Krauss, S., Neumann, K., & Schiepe-Tiska, A. (2025). Kodierschema zur Aufgaben-Analyse in PISA-Ceco: Potenzial zur kognitiven Aktivierung, motivationales Potenzial und Orientierung an den Bildungsstandards. In A. Schiepe-Tiska, S. Krauss, K. Reiss & D. Lewalter (Hrsg.), *PISA-Ceco – Dokumentation der Erhebungsinstrumente der Begleitstudie zu PISA 2022. Einblicke in den mathematisch-naturwissenschaftlichen Unterricht*. Waxmann. https://doi.org/10.31244/9783818850692.

Heinle, A., Reinhold, F., Todtenhöfer, P., Holzberger, D., & Schiepe-Tiska, A. (2026). Generic and subject-specific dimensions of Cognitive Activation: Assessing the Potential of Cognitive Activation in Mathematics and Science Tasks [in Vorbereitung]. *Zeitschrift für Erziehungswissenschaft.*

Herbert, B., & Schweig, J. (2021). Erfassung des Potenzials zur kognitiven Aktivierung über Unterrichtsmaterialien im Mathematikunterricht. *Zeitschrift für Erziehungswissenschaft*, *24*(4), 955–983. https://doi.org/10.25656/01:25456.

Hohenleitner, S., Stegmüller, N., Dittmer, A., Ehras, C., Neumann, K., Sorge, S., Tepner, O., Telser, V., Münch, B., & Krauss, S. (2025). Erhebung des fachdidaktischen Wissens von Lehrkräften in PISA-Ceco. In A. Schiepe-Tiska, S. Krauss, K. Reiss & D. Lewalter (Hrsg.), *PISA-Ceco – Dokumentation der Erhebungsinstrumente der Begleitstudie zu PISA 2022. Einblicke in den mathematisch-naturwissenschaftlichen Unterricht*. Waxmann. https://doi.org/10.31244/9783818850692.

Holder, K. (2019). *Inklusion und Bildungsstandards – gegensätzliche Anforderungen für Unterricht und Leistungsbeurteilung?* (Dissertation). Freie Universität Berlin. https://doi.org/10.17169/refubium-1909.

ISB (2025). *Lehrplan Plus – Gymnasium. Fachprofil Mathematik*. https://www.lehrplanplus.bayern.de/fachprofil/gymnasium/mathematik

Jahnke, H. N., Sommerhoff, D., & Ufer, S. (2023). Argumentieren, Begründen und Beweisen. In R. Bruder, A. Büchter, H. Gasteiger, B. Schmidt-Thieme & H. G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (2. Aufl., S. 369–398). Springer Spektrum. https://doi.org/10.1007/978-3-662-66604-3_12.

Johnson, H. L., Coles, A., & Clarke, D. (2017). Mathematical tasks and the student: Navigating "tensions of intentions" between designers, teachers, and students. *ZDM*, *49*(6), 813–822. https://doi.org/10.1007/s11858-017-0894-0.

Kleickmann, T., Praetorius, A. K., & Riecke-Baulecke, T. (2019). Beurteilung von Unterrichtsqualität. In F. Zimmermann, J. Möller & T. Riecke-Baulecke (Hrsg.), *Schulische Diagnostik und Leistungsbeurteilung* (S. 207–223). Kallmeyer.

Klieme, E. (2003). Benotungsmaßstäbe an Schulen: Pädagogische Praxis und institutionelle Bedingungen. Eine empirische Analyse auf der Basis der PISA-Studie. In H. Döbert, B. von Kopp, R. Martini & M. Weiß (Hrsg.), *Bildung vor neuen Herausforderungen* (S. 195–210). Luchterhand.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur" und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (Hrsg.), *TIMSS – Impulse für Schule und Unterricht, Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–58). Bonn: BMBF.

Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchung zur Bildungsqualität von Schule* (S. 127–146). Waxmann.

KMK (2004). *Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung*. Wolters Kluwer.

KMK (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf

KMK (2022). *Bildungsstandards für das Fach Mathematik Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA)*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-ESA-MSA-Mathe.pdf

KMK (2024). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf

König, J., Felske, C., & Kaiser, G. (2023). Professionelle Kompetenz von Mathematiklehrkräften aus einer pädagogischen Perspektive. In S. Krauss & A. Lindl (Hrsg.), *Professionswissen von Mathematiklehrkräften: Implikationen aus der Forschung für die Praxis* (S. 293–331). Springer. https://doi.org/10.1007/978-3-662-64381-5_9.

Krauss, S., & Lindl, A. (2023). *Professionswissen von Mathematiklehrkräften. Implikationen aus der Forschung für die Praxis*. Springer Spektrum. https://doi.org/10.1007/978-3-662-64381-5.

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., Besser, M., & Elsner, J. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, K. Baumert, W. Blum, U. Klusmann & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 163–192). Waxmann.

Krauss, S., Lindl, A., Schilcher, A., Fricke, M., & Göhring, A. (2017). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch Englisch Latein Physik Musik Evangelische Religion und Pädagogik: mit neuen Daten aus der COACTIV-Studie*. Waxmann.

Krauss, S., Hohenleitner, S., Stegmüller, N., Schiepe-Tiska, A., Reiss, K., & Lindl, A. (2026). PISA-Ceco 2022 und COACTIV 2003/04: Fachdidaktisches Wissen von Mathematiklehrkräften und die Qualität ihres Unterrichts im Vergleich. *Zeitschrift für Erziehungswissenschaft*.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften, Ergebnisse des Forschungsprogramms COACTIV*. Waxmann. https://doi.org/10.31244/9783830974338.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*(3), 805–820. https://doi.org/10.1037/a0032583.

Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O., & Reiss, K. (Hrsg.). (2023). *PISA 2022: Analyse der Bildungsergebnisse in Deutschland*. Waxmann Verlag. https://doi.org/10.25656/01:28666.

Lindl, A., Krauss, S., & Buchholtz, N. (2023). Professionswissen von Mathematiklehrkräften – eine einleitende Übersicht. In S. Krauss & A. Lindl (Hrsg.), *Professionswissen von Mathematiklehrkräften: Implikationen aus der Forschung für die Praxis* (S. 1–35). Springer. https://doi.org/10.1007/978-3-662-64381-5_1.

Metcalfe, J. (2017). Learning from Errors. *Annual Review of Psychology*, *68*, 465–489. https://doi.org/10.1146/annurev-psych-010416-044022.

Müller, M., Teufele, L., Bias, A., Diedrich, J., González Rodríguez, E., González Uriarte, A., & Lewalter, D. (2025). *PISA 2022 Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Waxmann. https://doi.org/10.25656/01:34036.

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 *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 115–132).

OECD (2023). *PISA 2022 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en.

Pant, H. A., Vock, M., Pöhlmann, C., & Köller, O. (2008). Offenheit für Innovationen. Befunde aus einer Studie zur Rezeption der Bildungsstandards bei Lehrkräften und Zusammenhänge mit Schülerleistungen. *Zeitschrift für Pädagogik*, *64*, 827–845. https://doi.org/10.25656/01:4379.

Park, S., & Oliver, J. S. (2008). Revisiting the Conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals. *Research in Science Education*, *38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6.

Praetorius, A. K., & Charalambous, C. (2018). Classroom observation frameworksfor studying instructional quality: Looking back and looking forward. *ZDM*, *50*(3), 535–553. https://doi.org/10.1007/s11858-018-0946-0.

Praetorius, A. K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B., & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, *48*(3), 409–446. https://doi.org/10.1007/s42010-020-00082-8.

Praetorius, A. K., Wemmer-Rogh, W., Schreyer, P., & Brinkmann, M. (2024). *Kognitive Aktivierung unter der Lupe: Bestandsaufnahme und Möglichkeiten der Weiterentwicklung eines prominenten Konstrukts*. Waxmann. https://doi.org/10.31244/9783830999010.

Reusser, K. (2024). Empirische Unterrichtsforschung – quo vadis? Plädoyer für eine didaktische Fundierung und mehr Praxisorientierung. *Unterrichtswissenschaft*, *52*(2), 167–182. https://doi.org/10.1007/s42010-024-00209-1.

Scheja, B., Schreck, A., & Rott, B. (2024). Prozessbezogene Kompetenzen in zentralen Abschlussarbeiten im Fach Mathematik? – Eine Analyse aller ZP10-Aufgaben Nordrhein-Westfalens von 2007 bis 2019. *Zeitschrift für Erziehungswissenschaft*, *27*(1), 197–227. https://doi.org/10.1007/s11618-023-01191-7.

Schiepe-Tiska, A., Krauss, S., Reiss, K., & Lewalter, D. (2025). *PISA-Ceco – Dokumentation der Erhebungsinstrumente der Begleitstudie zu PISA 2022. Einblicke in den mathematisch-naturwissenschaftlichen Unterricht*. Waxmann.

Schiepe-Tiska, A., Reiss, K., Krauss, S., Heinle, A., Reinhold, F., Holzberger, D., Todtenhöfer, P., & Lewalter, D. (2026). Mehrdimensionale Kompetenzorientierung im mathematisch-naturwissenschaftlichen Unterricht und das Studiendesign von PISA-Ceco. *Zeitschrift für Erziehungswissenschaft*.

Schlesinger, L., Jentsch, A., Kaiser, G., König, J., & Blömeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM*, *50*, 475–490. https://doi.org/10.1007/s11858-018-0917-5.

Schwarz, B., Buchholtz, N., & Kaiser, G. (2023). Professionelle Kompetenz von Mathematiklehrkräften aus einer mathematikdidaktischen Perspektive. In S. Krauss & A. Lindl (Hrsg.), *Professionswissen von Mathematiklehrkräften: Implikationen aus der Forschung für die Praxis* (S. 39–73). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64381-5_2.

Seidel, T., & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research*, *77*(4), 454–499. https://doi.org/10.3102/0034654307310317.

Shulman, L. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189x015002004.

Shulman, L. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411.

Tepner, O., Münch, B., Telser, V., Dittmer, A., Ehras, D., Krauss, S., Neumann, K., Sorge, S., Todtenhöfer, P., & Schiepe-Tiska, A. (2026). Wie hängen das Fachdidaktische Wissen und motivationale Kompetenzfacetten von Naturwissenschaftslehrkräften mit ihren Selbsteinschätzungen zur Unterrichtsgestaltung zusammen? *Zeitschrift für Erziehungswissenschaft*.

Todtenhöfer, P., Heinle, A., Dümig, P., Weis, M., Reinhold, F., Holzberger, D., Reiss, K., Lewalter, D., & Schiepe-Tiska, A. (2025). Fragebögen und Beobachtungsinstrumente in PISA-Ceco. In A. Schiepe-Tiska, S. Krauss, K. Reiss & D. Lewalter (Hrsg.), *PISA-Ceco – Dokumentation der Erhebungsinstrumente der Begleitstudie zu PISA 2022. Einblicke in den mathematisch-naturwissenschaftlichen Unterricht*. Waxmann. https://doi.org/10.31244/9783818850692.

Ufer, S., Heinze, A., & Lipowsky, F. (2023). Unterrichtsqualität und Instruktionsstrategien. In R. Bruder, A. Büchter, H. Gasteiger, B. Schmidt-Thieme & H. G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (2. Aufl. S. 465–492). Springer Spektrum. https://doi.org/10.1007/978-3-662-66604-3_15.

Yang, X., & Kaiser, G. (2022). The impact of mathematics teachers' professional competence on instructional quality and students' mathematics learning outcomes. *Current Opinion in Behavioral Sciences*, *48*, 101225. https://doi.org/10.1016/j.cobeha.2022.101225.

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.