Zeitschrift für Erziehungswissenschaft
https://doi.org/10.1007/s11618-026-01427-2

SCHWERPUNKT

PISA-Ceco 2022 und COACTIV 2003/04: Fachdidaktisches Wissen von Mathematiklehrkräften und die Qualität ihres Unterrichts im Vergleich

Stefan Krauss · Stefanie Hohenleitner · Nathalie Stegmüller · Anja Schiepe-Tiska · Kristina Reiss · Alfred Lindl

Eingegangen: 18. Mai 2025 / Überarbeitet: 24. August 2026 / Angenommen: 10. September 2026
© The Author(s) 2026

Zusammenfassung Sowohl bei PISA-Ceco 2022 als auch bei COACTIV 2003/04 wurden das fachdidaktische Wissen von Mathematiklehrkräften sowie Aspekte der Unterrichtsqualität in PISA-Klassen gemessen. Das komplexe Transformationsgeschehen in den letzten 20 Jahren – z. B. bildungspolitisch die Einführung von Standards für Mathematikunterricht und für die Lehrkräftebildung, aber auch Aspekte wie Migration, Pandemie, Digitalisierung – motiviert ein Monitoring potentieller Änderungen in der Bildungsqualität. Im fachdidaktischen Wissen der Lehrkräfte können dabei keine signifikanten Unterschiede zwischen den beiden Large-Scale-Studien gefunden werden. Während die kognitive Aktivierung und die Klassenführung 2004 aus Sicht der Schülerinnen und Schüler bei den COACTIV-Lehrkräften

Prof. Dr. Stefan Krauss · ✉ Stefanie Hohenleitner · 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. Anja Schiepe-Tiska
Erziehungswissenschaft, Universität Koblenz, Universitätsstraße 1, 56070 Koblenz, Deutschland
E-Mail: schiepe-tiska@uni-koblenz.de

Prof. Dr. Kristina Reiss
Didaktik der Mathematik, Technische Universität München, Arcisstr. 21, 80333 München, Deutschland
E-Mail: kristina.reiss@tum.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

besser bewertet wurden, wird 2022 die konstruktive Unterstützung bei den PISA-Ceco-Lehrkräften signifikant höher beurteilt.

Im Lichte der Replikationskrise in den Sozialwissenschaften ist es darüber hinaus eine offene Frage, inwiefern sich in der COACTIV-Studie gefundene Zusammenhänge zwischen fachdidaktischem Wissen und Unterrichtsqualitätsaspekten, aber auch damals überraschende Befunde wie die Schulartunterschiede bezüglich des fachdidaktischen Wissens zugunsten gymnasialer Lehrkräfte oder die fehlende Abhängigkeit des fachdidaktischen Wissens von der Berufserfahrung mit den Daten von PISA-Ceco replizieren lassen. Die gelungene Replikation wesentlicher Zusammenhänge aus der COACTIV-Studie auf der Basis einschlägiger Konstrukte und bewährter Instrumente ist dabei ein ermutigendes Signal, auch für die empirische Bildungsforschung im Allgemeinen.

Schlüsselwörter PISA-Ceco · COACTIV · Fachdidaktisches Wissen · Unterrichtsqualität · Änderungsmonitoring · Replikation

PISA-Ceco 2022 and COACTIV 2003/04: A comparison of mathematics teachers' pedagogical content knowledge and the instructional quality

Abstract Both PISA-Ceco 2022 and COACTIV 2003/04 measured the pedagogical content knowledge of mathematics teachers as well as aspects of instructional quality in PISA classes. The complex transformations that took place in the nearly 20 years between studies—such as the introduction of standards for mathematics instruction and teacher education, as well as factors such as migration, the COVID-19 pandemic, and digitalization—provided a rationale for monitoring potential changes in educational quality. We found no significant differences between the two large-scale studies in terms of teachers' pedagogical content knowledge. Regarding cognitive activation and classroom management, students rated teachers more highly in the COACTIV study in 2004, whereas constructive learning support was rated significantly higher for teachers in the PISA-Ceco study in 2022.

Given the replication crisis in the social sciences, we deemed it reasonable to ask to what extent the correlations found in the earlier COACTIV study between pedagogical content knowledge and aspects of instructional quality, as well as findings that were surprising at the time—such as differences across school types regarding pedagogical content knowledge in favor of academic-track teachers, or the lack of correlation between pedagogical content knowledge and professional experience—could be replicated using the later PISA-Ceco data. We successfully replicated key findings from the COACTIV study, using pertinent constructs and established instruments. This is an encouraging sign—not only for this study but also for empirical educational research in general.

Keywords PISA-Ceco · COACTIV · Pedagogical content knowledge · Instructional quality · Monitoring of changes · Replication

1 Bildungsmonitoring im 18 Jahres-Vergleich: PISA-Ceco 2022 und COACTIV 2003/2004

Die nicht befriedigenden Ergebnisse der beiden letzten PISA-Studien 2022 und 2025 in Deutschland (Lewalter et al. 2023; Greiff et al. 2026) haben eine öffentliche wie wissenschaftliche Debatte ausgelöst. Diese Diskussionen erinnern an die erste PISA-Studie im Jahr 2000, als Defizite im deutschen Bildungssystem deutlich sichtbar wurden. Nach einer Phase kontinuierlicher Kompetenzsteigerungen bis etwa 2012 zeigen aktuellere Ergebnisse wieder schwache Leistungen der Jugendlichen insbesondere in Mathematik. In Deutschland liegt die mathematische Kompetenz von 15-Jährigen bei PISA 2022 im OECD-Durchschnitt, nachdem sie in einigen Zyklen der Studie signifikant darüber lag.

Der Befund, dass die Testergebnisse der Jugendlichen in Deutschland bei der ersten PISA-Studie 2000 unter dem Durchschnitt der Staaten der OECD lagen, ist heute als „PISA-Schock" bekannt (vgl. Baumert et al. 2001). Eine Reaktion der Bildungsforschung war die Initiierung einer umfassenden begleitenden Lehrkraft- und Unterrichtsstudie im Zuge der zweiten PISA-Erhebung 2003, nämlich die COACTIV-Studie (Kunter et al. 2011), für die in den Jahren 2003 und 2004 umfangreich Daten erhoben wurden. Im Rahmen von COACTIV wurde das mittlerweile einschlägige Modell *professioneller Lehrkräftekompetenz* entwickelt, bestehend aus den vier übergeordneten Aspekten *Professionswissen, professionelle Überzeugungen, motivationale Orientierungen* und *selbstregulative Fähigkeiten*. Weiterhin wurde ein Modell der Unterrichtsqualität mit den drei Basisdimensionen *kognitive Aktivierung, konstruktive Unterstützung* und *Klassenführung* spezifiziert (Kunter und Voss 2011). Sowohl zu den Kompetenzen von Lehrkräften als auch zu den Aspekten von Unterrichtsqualität wurden für die COACTIV-Studie zahlreiche Messinstrumente neu entwickelt oder entsprechend adaptiert (Baumert et al. 2009).

Mit PISA-Ceco („Classroom experience, characteristics and outcome") wurde begleitend zur PISA-Erhebung 2022 nun wieder eine nationale Zusatzstudie durchgeführt, die neben dem mathematischen auch den naturwissenschaftlichen Unterricht in den Blick nimmt (vgl. Schiepe-Tiska et al. 2023). Im Zentrum von PISA-Ceco steht die Frage, wie Kompetenzorientierung im mathematisch-naturwissenschaftlichen Unterricht umgesetzt wird und in welchem Zusammenhang der Unterricht mit kognitiven sowie motivational-affektiven Bildungszielen steht. Neben Unterrichtsbeobachtungen dienten Selbstauskünfte von Lehrkräften, Einschätzungen der Schülerinnen und Schüler zu ihrem Unterricht, Materialien einer Unterrichtsstunde und Klassenarbeiten aus dem laufenden Schuljahr als Datengrundlage. PISA-Ceco ermöglicht eine differenzierte Analyse der intendierten und tatsächlich erlebten Unterrichtsmerkmale und deren Bedeutung für mehrdimensionale Lernziele (vgl. Schiepe-Tiska et al. 2026).

Zur Kontextualisierung wurden ausgewählte Aspekte der professionellen Kompetenz der Lehrkräfte erfasst, etwa ihre Haltung zur Kompetenzorientierung (Hohenleitner et al. 2026) oder ihr fachdidaktisches Wissen. Die PISA-Ceco-Studie greift dabei zentrale theoretische Konzepte – auch aus COACTIV – auf und setzt diese im Hinblick auf aktuelle Fragestellungen der pädagogisch-psychologischen und fach-

didaktischen Unterrichts- sowie Lehrkräfteforschung empirisch um (Schiepe-Tiska et al. 2025).

Für das Fach Mathematik ist es durch die Anlage der beiden PISA-Begleitstudien 2003/2004 bzw. 2022 möglich, das fachdidaktische Wissen von Lehrkräften sowie Aspekte der drei Basisdimensionen der Unterrichtsqualität – kognitive Aktivierung, konstruktive Unterstützung und Klassenführung – im Vergleich zwischen PISA-Ceco und COACTIV zu untersuchen und zu prüfen, inwieweit sich zentrale Befunde der COACTIV-Studie nach knapp zwei Jahrzehnten bestätigen oder welche Weiterentwicklungen zu erkennen sind. Dabei werden zwei Analyseperspektiven eingenommen: eine veränderungssensitive Betrachtung („Änderungsmonitoring“) sowie eine replizierende Analyse zentraler Zusammenhänge („Replikationsansatz“).

Beim Änderungsmonitoring wird anhand von Mittelwertvergleichen zwischen der COACTIV-Erhebung und PISA-Ceco untersucht, ob sich das fachdidaktische Wissen von Mathematiklehrkräften bzw. die Qualität des Mathematikunterrichts – gemessen an den drei Basisdimensionen – im Zeitverlauf verändert hat. So könnten einerseits verschiedene bildungspolitische Maßnahmen einen Einfluss gehabt haben, die erst nach der COACTIV-Studie erfolgten, wie beispielsweise die bundesweite Einführung von Bildungsstandards für das Fach Mathematik (Blum et al. 2006; Hohenleitner et al. 2026) oder die Standards für die fachwissenschaftliche, fachdidaktische und bildungswissenschaftliche Lehrkräftebildung (KMK 2022 [2004], 2024 [2008]). Andererseits könnten gesellschaftliche Herausforderungen und Transformationen wie Migration, Pandemie oder Digitalisierung einen Einfluss auf Lehrkräfte und deren Unterrichtsqualität gehabt haben.

Unter dem Schlagwort „Replikationskrise“ manifestierte sich in letzter Zeit zudem in den Sozialwissenschaften und der Psychologie – aber auch in der Bildungsforschung sowie den Fachdidaktiken – die Erkenntnis, dass sich bestimmte Effekte bei erneuter Untersuchung aus verschiedenen Gründen nicht bestätigen ließen (z. B. Bienefeld et al. 2020; Cai et al. 2018). Da dies sowohl die Bedeutsamkeit als auch die Generalisierbarkeit von Befunden der empirischen Bildungsforschung gefährdet, analysiert der vorliegende Beitrag, ob sich zentrale korrelative Befunde von COACTIV wie die differentiellen Zusammenhänge zwischen fachdidaktischem Wissen und den drei Aspekten der Unterrichtsqualität durch Pfadmodelle auf Basis der PISA-Ceco Daten bestätigen lassen. Von besonderem Interesse ist dabei auch, ob sich damals überraschende Resultate wie die in COACTIV identifizierten Schulartunterschiede im fachdidaktischen Wissen – zugunsten gymnasialer Lehrkräfte – sowie der fehlende Zusammenhang zwischen fachdidaktischem Wissen und Berufserfahrung mit den PISA-Ceco Daten replizieren lassen.

Bei der Betrachtung der beiden Perspektiven ist zu beachten, dass sich zwar Mittelwerte von Skalen im Laufe der Zeit ändern können und positive Zuwächse hierbei sogar erwünscht sind (Änderungsmonitoring), dies für Zusammenhänge zwischen – manifesten oder latenten – Konstrukten aber gerade nicht der Fall sein sollte (Replikationsansatz). Zum einen sind insbesondere lineare Modelle – wie Korrelationen – statistisch unabhängig von den Mittelwerten der beteiligten Variablen. Zum anderen bilden solche Modelle strukturelle empirische Beziehungen in Unterrichtsumgebungen ab und es gibt derzeit keine bekannten Theorien zur zeitlichen Veränderung solcher Beziehungen.

2 Fachdidaktisches Wissen und Unterrichtsqualität

Der PISA-Schock 2000 war eine Initialzündung für umfangreiche theoretische und empirische Forschung in Bezug auf die Kompetenz von Lehrkräften und den schulischen Unterricht. Insbesondere betrifft das im deutschsprachigen Raum die empirische Untersuchung von Kategorien des Professionswissens – vorrangig Fachwissen, fachdidaktisches Wissen und pädagogisch-psychologisches Wissen – und die Messung von Unterrichtsqualität in Form der drei Basisdimensionen (Kunter et al. 2011).

2.1 Fachdidaktisches Wissen (FDW)

Psychometrische Testkonstruktionen zum fachdidaktischen Wissen (FDW), zum Fachwissen und zum pädagogischen Wissen berufen sich in der Regel auf die theoretische Wissenstaxonomie von Shulman (1987). Aufgrund der positiven Effekte vor allem des FDW in der COACTIV-Studie (siehe 2.3) erfolgte in PISA-Ceco eine Fokussierung auf diese Wissenskategorie. Für das Fach Mathematik greifen beinahe alle FDW-Konzeptualisierungen die beiden Aspekte „Wissen über Erklären und Repräsentieren“ und „Wissen über typische Schülerfehler bzw. -schwierigkeiten“ auf (für einen Überblick siehe z. B. Lindl et al. 2023; Sarkar et al. 2024). In den Naturwissenschaften wird darüber hinaus oft zwischen verschiedenen Wissensarten wie „theoretischem Lehrbuchwissen“, „persönlichem fachdidaktischen Wissen“ und „im Unterricht umgesetzten Wissen“ unterschieden (Carlson et al. 2019). Im FALKO-Projekt, in dem FDW-Tests sowohl für sprachliche Unterrichtsfächer (Latein, Englisch, Deutsch) als auch für Musik und Religion entwickelt wurden, ist – neben dem Wissen über Erklären und Repräsentieren sowie über Schülerfehler bzw. -schwierigkeiten – ein dritter Aspekt von FDW das „Erkennen des Potentials von Unterrichtsmaterialien für verständnisvolle Lernprozesse“ (Krauss et al. 2017).

Das FDW zeigt sich nicht nur in quantitativ orientierten Studien als Kernkompetenz von Lehrkräften (z. B. Blömeke et al. 2022; Fukaya et al. 2025). In einem aktuellen Review zur Expertise von Lehrkräften (Anderson und Taner 2023), das vor allem auf qualitativer Forschung beruht, gehört das FDW mittlerweile ebenfalls zu den drei meistuntersuchten Expertisemerkmalen (siehe hierzu auch Krauss et al. 2025). Ein geschichtlicher Abriss zur FDW-Forschung im Allgemeinen sowie der übergeordnete theoretische Rahmen der FDW-Testkonstruktionen in PISA-Ceco für Mathematik, Biologie, Chemie und Physik werden in der Skalendokumentation dargelegt (Hohenleitner et al. 2025). Die konkrete Konzeptualisierung und Operationalisierung des FDW der PISA-Ceco Mathematiklehrkräfte wird in einem weiteren Beitrag im vorliegenden Themenschwerpunkt ausführlich beschrieben (Hohenleitner et al. 2026).

2.2 Unterrichtsqualität (KA, KU, KF)

Die Modellierung der Unterrichtsqualität anhand der drei Tiefenstrukturen kognitive Aktivierung (KA), konstruktive Unterstützung (KU) und Klassenführung (KF) geht auf eine Expertise von Klieme et al. (2001) zurück, die bereits als Reaktion auf die ebenfalls unbefriedigenden deutschen TIMSS-Ergebnisse 1999 verfasst wurde.

Diese Merkmale avancierten seither zu den drei Basisdimensionen von Unterrichtsqualität (z. B. Praetorius et al. 2018), wurden vielfach in Bezug auf unterschiedliche Unterrichtsfächer untersucht (z. B. Alp Christ et al. 2024) und in den letzten Jahren zunehmend auch kritisch hinterfragt (z. B. Reusser 2024; Rothland 2024).

KA ist mittlerweile ein vielschichtiger Begriff (Praetorius et al. 2024); im Mathematikunterricht können Lernende beispielsweise durch fachlich gehaltvolle Tätigkeiten wie Argumentieren, Modellieren oder Problemlösen kognitiv aktiviert werden (Blum et al. 2006). KU ist auf die Erfüllung der Bedürfnisse individueller Schülerinnen und Schüler ausgerichtet, mit dem Ziel der Förderung der intrinsischen Motivation (Deci und Ryan 2008). Hierzu zählen fachliche Hilfestellungen genauso wie Unterstützung bei persönlichen Anliegen. Zu KF gehören effiziente präventive und intervenierende Maßnahmen zur Aufrechterhaltung der Disziplin im Klassenzimmer sowie eine effektive Zeitnutzung (Kunter und Voss 2011).

Bei PISA-Ceco wurde die Unterrichtsqualität mehrperspektivisch erhoben (z. B. Schüler- und Lehrkräftefragebögen, Kodierung eingesammelter Aufgaben, Unterrichtsbesuche). Eine ausführliche Auseinandersetzung mit den verschiedenen Zugängen vor allem zur Messung von KA im Rahmen von PISA-Ceco findet sich in Hohenleitner et al. (2026) bzw. Heinle et al. (2026).

2.3 Effekte des FDW

Mittlerweile gibt es zahlreiche empirisch implementierte Modelle zur Schätzung von Effekten des Professionswissens auf Aspekte der Unterrichtsqualität, aber auch auf den Lernzuwachs von Schülerinnen und Schülern (siehe etwa Krauss et al. 2020). In einer aktuellen Metaanalyse zu diesem Thema verdeutlichen Fukaya et al. (2025) – trotz in Einzelstudien durchaus gemischter Befundlage – die entscheidende Bedeutung des FDW für eine qualitätsvolle Unterrichtsgestaltung. Dieser Befund zeigte sich auch in der COACTIV-Studie, und zwar vor allem in Bezug auf KA sowie auf KU (Kunter et al. 2011). KF ist demgegenüber weniger mit FDW (z. B. Jentsch et al. 2021), sondern mit pädagogisch-psychologischem Wissen assoziiert (z. B. König et al. 2023).

Es liegen auch schon längsschnittliche Studien zu Effekten des FDW auf den Lernzuwachs unter Kontrolle des Vorwissens der Schülerinnen und Schüler vor (z. B. Hill et al. 2005; Kelcey et al. 2019; König et al. 2021; Sadler et al. 2013; Yang et al. 2025). Auch diesbezüglich belegen Überblicksarbeiten insgesamt einen positiven Zusammenhang (Charalambous et al. 2020; Fukaya et al. 2025). Keinen direkten Effekt des FDW auf den Lernzuwachs – sondern nur auf Unterrichtsqualität – finden beispielsweise Blömeke et al. (2022) oder Förtsch et al. (2018). In der COACTIV-Studie dagegen war das FDW unter ca. 1000 erhobenen Variablen zu Lehrkraftmerkmalen und Unterrichtsqualität sogar der mit Abstand stärkste Prädiktor für den Lernzuwachs der Schülerinnen und Schüler (Krauss et al. 2025).

2.4 Überraschende Befunde zum FDW in der COACTIV-Studie

Eine besondere Herausforderung für eine Replikation stellen überraschende Resultate der COACTIV-Studie dar. Da beispielsweise nicht-gymnasiale Lehramts-

studierende an der Universität mehr Didaktik- und Pädagogikveranstaltungen als gymnasiale Studierende belegen müssen, war in der COACTIV-Studie der Vorteil gymnasialer Lehrkräfte beim FDW – deren universitäre Ausbildung sich stark auf fachliche Aspekte konzentriert – ein viel diskutierter Befund (z. B. Bruckmaier et al. 2023). Ebenfalls erstaunlich war der fehlende Zusammenhang von FDW mit der Berufserfahrung. Hier hätte man erwarten können, dass eine breitere praktische Erfahrung zu einem größeren Repertoire an Erklärstrategien sowie einem besseren Wissen über typische Schülerfehler und -schwierigkeiten führt (Krauss et al. 2025). Für beide Ergebnisse ist eine offene Frage, ob es sich lediglich um stichprobenspezifische Befunde handelt oder ob sich auch diese Resultate durch eine Replikation validieren lassen.

3 Ziele & Fragestellungen

Bei PISA-Ceco und COACTIV wurden Aspekte von FDW, KA, KU und KF jeweils auf der Basis identischer oder sehr ähnlicher Items erhoben. Dies bietet die Gelegenheit, diese Merkmale für das Fach Mathematik im Abstand von knapp 20 Jahren zu vergleichen. Im folgenden empirischen Teil werden für alle vier Konstrukte die Ergebnisse eines entsprechenden Mittelwertvergleichs berichtet. Darüber hinaus werden auf Basis der Daten von PISA-Ceco die bei der COACTIV-Studie gefundenen wechselseitigen Zusammenhänge zwischen den vier Konstrukten einem Replikationsversuch unterzogen. Konkret werden folgende Fragestellungen untersucht:

a) Änderungsmonitoring

1. Gibt es Unterschiede im fachdidaktischen Wissen (FDW) von Mathematiklehrkräften zwischen 2004 und 2022?
2. Gibt es Unterschiede in der Qualität des Mathematikunterrichts (KA, KU, KF) zwischen 2004 und 2022?

Das komplexe Transformationsgeschehen in den letzten 20 Jahren (etwa die Einführung von Bildungsstandards oder Herausforderungen wie Pandemie oder Migration) lässt zu diesen Fragestellungen Hypothesen in beide Richtungen zu.

b) Replikationsansatz

3. Lassen sich die in COACTIV gefundenen Schulartunterschiede im FDW zugunsten gymnasialer Lehrkräfte bzw. der fehlende Zusammenhang des FDW mit der Berufserfahrung mit den Daten von PISA-Ceco replizieren?
4. Zeigen sich die in COACTIV gefundenen differentiellen Zusammenhänge von FDW und Unterrichtsqualität (KA, KU, KF) auch in den Daten von PISA-Ceco? Lassen sich entsprechende Koeffizienten von Pfadmodellen replizieren?

Eine erfolgreiche Replikation von Zusammenhängen hätte allgemeine Bedeutung für die Verlässlichkeit und Generalisierbarkeit der (quantitativen) empirischen Bildungsforschung sowie der dort verwendeten Verfahren.

4 Methode

Im Folgenden werden die untersuchten Stichproben (4.1) und die für die Analysen verwendeten Konstrukte sowie deren Operationalisierungen vorgestellt (4.2). Eine ausführliche Übersicht über alle Instrumente findet sich in der PISA-Ceco Skalendokumentation (Schiepe-Tiska et al. 2025).

4.1 Stichprobe und Administration

Sowohl bei COACTIV als auch in PISA-Ceco wurden die Mathematiklehrkräfte von PISA-Schulen für ihre freiwillige Teilnahme an einer Begleitstudie angefragt. Der üblichen Vorgehensweise in der empirischen Bildungsforschung folgend wird bei den Analysen jeweils zwischen gymnasialen und nicht-gymnasialen Schularten unterschieden (GY vs. NGY). Die weitere Differenzierung der NGY-Schularten wird aufgrund der großen diesbezüglichen Heterogenität zwischen den Bundesländern seit PISA 2015 im Berichtsband zur Studie nicht mehr vorgenommen. Im Folgenden wird diese Aufteilung aber jeweils mit angegeben, um potentielle Mittelwertsunterschiede bei den nicht-gymnasialen Lehrkräften zwischen 2004 und 2022 näher beleuchten zu können.

4.1.1 PISA-Ceco 2022: Lehrkräfte

An PISA-Ceco nahmen $N = 150$ Mathematiklehrkräfte teil, von denen 125 den FDW-Test (siehe 4.2) in digitaler Form valide bearbeiteten (d. h., deren Bearbeitungszeit in einem plausiblen Intervall liegt, vgl. Hohenleitner et al. 2025). Von diesen 125 Lehrkräften (42,4 % weiblich, 40,8 % männlich, 16,8 % ohne Angabe) waren 63 GY- und 62 NGY-Lehrkräfte. Die durchschnittliche Berufserfahrung betrug 10,7 Jahre ($SD = 7{,}75$). Insgesamt konnten $N = 63$ Klassen (GY: 26; NGY: 37, davon 14 Realschule, 15 integrierte Gesamtschule, vier Schulen mit mehreren Bildungsgängen, vier Haupt-/Mittelschule) Lehrkräften zugeordnet werden.

4.1.2 PISA 2022: Schülerinnen und Schüler

Die Schülerinnen und Schüler, die an PISA 2022 teilnahmen, wurden im Rahmen eines geschichteten Stichprobenverfahrens ausgewählt. In Deutschland erfolgte zunächst eine zufällige Auswahl der Schulen. Anschließend wurden in jeder dieser Schulen zwei neunte Klassen per Zufallsverfahren gezogen. Daraufhin wurden pro Schule 30 Fünfzehnjährige sowie innerhalb der beiden zuvor ausgewählten neunten Klassen zusätzlich 15 Schülerinnen und Schüler zufällig ausgewählt (Mang et al. 2023). Der Begriff „Klasse“ bezieht sich in diesem Zusammenhang folglich auf die jeweils gezogenen Schülerinnen und Schüler einer Klasse. Von den $N = 63$ Klassen,

denen eine Lehrkraft zugeordnet werden konnte, bearbeiteten pro Klasse im Durchschnitt 16 Schülerinnen und Schüler den Fragebogen zur Erfassung von KA, KU und KF.

4.1.3 COACTIV 03/04: Lehrkräfte

In COACTIV bearbeiteten insgesamt $N = 205$ Mathematiklehrkräfte (darunter 45,5 % weibliche und 46,2 % männliche) den FDW-Test, der zum zweiten COACTIV-Messzeitpunkt 2004 durchgeführt wurde. Die durchschnittliche Berufserfahrung in dieser Stichprobe betrug 22 Jahre ($SD = 9{,}35$), wobei 85 GY-Lehrkräfte und 120 NGY-Lehrkräfte waren. Zu $N = 186$ Lehrkräften (GY: 76; NGY: 110, davon 69 Realschule, 21 integrierte Gesamtschule, 20 Schulen mit mehreren Bildungsgängen) liegen Klassendaten vor.

4.1.4 PISA 2003: Schülerinnen und Schüler

Bei PISA 2003 gab es ein ähnliches Stichprobenverfahren wie bei PISA 2022, allerdings wurden für die COACTIV-Studie pro zufällig gezogener Schule zusätzlich zwei neunte Klassen jeweils vollständig erhoben und PISA national längsschnittlich um einen zweiten Messzeitpunkt 2004 erweitert (Löwen et al. 2011; Prenzel et al. 2006). Von den $N = 186$ COACTIV-Lehrkräften, zu denen längsschnittliche Klassendaten vorliegen, bearbeiteten pro Klasse im Durchschnitt 23 Schülerinnen und Schüler den Fragebogen zur Erfassung von KA, KU und KF.

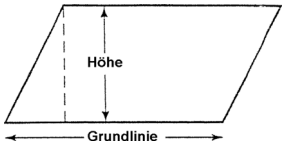
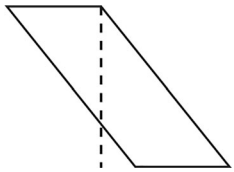
4.2 Instrumente

4.2.1 Fachdidaktisches Wissen (FDW)

Das FDW wird bei PISA-Ceco und COACTIV anhand dreier Facetten konzeptualisiert: Wissen über kognitive Schülerfehler und -schwierigkeiten (Skog), Wissen über Erklären und Repräsentieren von Inhalten (E&R) und Wissen über das multiple Lösungspotential von Aufgaben (Pot). In Tab. 1 ist für jede Facette ein Beispielitem mit exemplarischer Lösungsskizze angeführt (für eine ausführliche Darstellung der Konzeptualisierung und Operationalisierung der drei Facetten vgl. Hohenleitner et al. 2025, 2026).

Der FDW-Test bei COACTIV bestand ursprünglich aus insgesamt 22 Items. Für PISA-Ceco wurden pro Facette zwei Items für eine FDW-Kurzskala ausgewählt. Kriterien für die Auswahl waren dabei eine hohe Trennschärfe der Items bezüglich des Gesamttests, die Ausgewogenheit zwischen Inhaltsgebieten (Geometrie, Algebra, Arithmetik) und eine hohe prädiktive Validität für den Leistungszuwachs bei den COACTIV-Daten. Diese sechs Items wurden dann jeweils auch für die Vergleichsanalysen herangezogen (d. h., auch alle im Folgenden berichteten COACTIV-Ergebnisse beziehen sich auf diese jeweils zwei Items pro Facette). Auch wenn die sechs ausgewählten Items für PISA-Ceco wörtlich übernommen wurden, gibt es dennoch einige Unterschiede in Hinblick auf die Administration.

Tab. 1 Beispielitems aus der in beiden Studien (COACTIV und PISA-Ceco) eingesetzten Kurzskala zum fachdidaktischen Wissen (FDW) von Mathematiklehrkräften (6 Items) mit exemplarisch skizzierten Lösungsansätzen (vgl. Krauss et al. 2011)

Facette	Beispielitem	Exemplarische Lösungsskizze
Wissen über kognitive Schülerfehler und -schwierigkeiten (Skog)	**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 (E&R)	**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: $2 \cdot (-1) = -2$ $1 \cdot (-1) = -1$ $0 \cdot (-1) = 0$ $(-1) \cdot (-1) = 1$ -1 … $+1$
Wissen über das multiple Lösungspotential von Aufgaben (Pot)	**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)

Tab. 2 kann entnommen werden, dass die FDW-Items in der COACTIV-Studie teilweise im Papier- und Bleistift-Format („P&B“) und teilweise computerbasiert („CB“) erhoben wurden. In PISA-Ceco wurden alle sechs FDW-Items in einen Online-Lehrkräftefragebogen eingebettet, der zuhause bearbeitet werden konnte. Zu beachten ist, dass von den sechs für PISA-Ceco gewählten Items in COACTIV nur eines ebenfalls auf dem Computer bearbeitet werden musste (damals allerdings nicht online, sondern auf einem vom Testleiter bzw. der Testleiterin mitgebrachten Laptop).

Den Lehrkräften bei PISA-Ceco standen für jedes Item ein Eingabefeld für die offene Antwort und eine Zeichenfläche für etwaige Skizzen zur Verfügung. Im Gegensatz zu den P&B-Items bei COACTIV gab es nicht die Möglichkeit, bei der

Tab. 2 Vergleich von COACTIV und PISA-Ceco hinsichtlich der Erfassung des fachdidaktischen Wissens FDW (Stichprobe, Durchführungsjahr, # Anzahl Items pro Modus und FDW-Facette, Administrationsdetails)

	COACTIV		PISA-Ceco
Stichprobe	*N* = 217 Mathematiklehrkräfte		*N* = 150 Mathematiklehrkräfte
	Davon 205 gültige Bearbeitungen der FDW-Items		Davon 125 gültige Bearbeitungen des FDW-Tests
	205 Mathematiklehrkräfte, die sowohl den P&B-Test als auch den CB-Fragebogen bearbeiteten		Davon 57 Lehrkräfte von 63 PISA-Klassen (6 Lehrkräfte unterrichteten je zwei Klassen)
Jahr der Durchführung	2004		2022
Gesamtzahl Items [# pro Modus]	Original 22; für vorliegenden Beitrag: 6 [14: P & B; 8: CB]		6 [alle 6 online]
FDW-Facetten	*Original*	*Vorliegender Beitrag*	–
Skog	7 [4: P & B; 3: CB]	Davon 2 gewählt	2 [beide online]
E&R	12 [7: P & B; 5: CB]	Davon 2 gewählt	2 [beide online]
Pot	4 [4: P & B]	Davon 2 gewählt	2 [beide online]
Administration	Einzelsitzungen unter Aufsicht eines Testleiters (Testleiter stellt für CB-Items Laptop zur Verfügung)		Onlinefragebogen (ohne Aufsicht)
			Keine Möglichkeit, zu einem bereits bearbeiteten Item zurückzukehren
Dauer	Kein Zeitlimit		Kein Zeitlimit
	Durchschnittliche Bearbeitungsdauer pro FDW-Item: etwa 3–4 Min.		Instruktion, bitte etwa 5 Min. pro FDW-Item einzuplanen (aber keine Sanktion bei Nichtbeachtung)

FDW Fachdidaktisches Wissen, *Skog* Wissen über typische Schülerfehler und -schwierigkeiten, *E & R* Wissen über Erklären und Repräsentieren von Inhalten, *Pot* Wissen über das multiple Lösungspotential von Aufgaben
Modi: *P & B* Papier und Bleistift, *CB* Computerbasiert (Laptop von Testleiter*in), *online* (zuhause bearbeitbar)

Bearbeitung zu einer vorherigen Aufgabe zurückzukehren (für weitere administrative Unterschiede vgl. Tab. 2).

Die Antworten der Lehrkräfte bei PISA-Ceco wurden von zwei geschulten Kodiererinnen geratet, wobei das damalige COACTIV-Kodiermanual für die sechs Items der FDW-Kurzskala nach aktuellem Forschungsstand überarbeitet wurde. Für die Vergleichsanalysen wurden die Antworten auf diese sechs Items auch bei den COACTIV-Lehrkräften auf der Basis des überarbeiteten Kodiermanuals von denselben beiden Raterinnen neu kodiert. Das Reliabilitätsmaß Cronbachs Alpha für die sechs Items konnte dabei mit den COACTIV-Daten von $\alpha = 0{,}59$ auf $\alpha = 0{,}65$ erhöht werden (für Details siehe Hohenleitner et al. 2025, 2026).

4.2.2 Unterrichtsqualität (KA, KU, KF)

Für die Analysen zur Unterrichtsqualität im Hinblick auf KA, KU und KF wurde jeweils auf die durch einen Fragebogen erhobene Perspektive der Schülerinnen und Schüler zurückgegriffen (Tab. 3 zeigt sämtliche aus beiden Studien genutzten Items).

Tab. 3 Vergleich der Messung von Konstrukten der Unterrichtsqualität (KA, KU, KF) bei COACTIV und PISA-Ceco

Merkmal	COACTIV (*N* = 185 Klassen)	PISA-Ceco (*N* = 61 Klassen)
Kognitive Aktivierung (KA)	Unser Mathematiklehrer/unsere Mathematiklehrerin lässt uns häufiger unsere Gedankengänge genau erklären.	Die Lehrkraft forderte uns auf, unsere Gedankengänge bei der Lösung eines mathematischen Problems zu erklären.
	Unser Mathematiklehrer/unsere Mathematiklehrerin verlangt häufiger, dass wir unsere Arbeitsschritte ausführlich begründen.	
	Unser Mathematiklehrer/unsere Mathematiklehrerin fragt häufiger, ob jemand einen anderen Lösungsweg gefunden hat.	Die Lehrkraft ermutigte uns, darüber nachzudenken, wie man ein mathematisches Problem anders lösen könnte, als es im Unterricht gezeigt wurde.
	Unser Mathematiklehrer/unsere Mathematiklehrerin zeigt uns an Beispielen aus dem täglichen Leben, wozu man Mathematik brauchen kann.	Die Lehrkraft zeigte uns, wie Mathematik in unserem Alltag nützlich sein kann.
Skalierung	*1 = trifft nicht zu, ..., 4 = trifft zu*	*1 = (fast) nie, ..., 4 = in (fast) jeder Stunde*
Konstruktive Unterstützung (KU)	Unser Lehrer/unsere Lehrerin interessiert sich für den Lernfortschritt jedes einzelnen Schülers/jeder Schülerin.	Die Lehrkraft interessiert sich für den Lernfortschritt jeder/jedes einzelnen Schülerin/Schülers.
	Der Lehrer/die Lehrerin unterstützt uns zusätzlich, wenn wir Hilfe brauchen.	Die Lehrkraft unterstützt zusätzlich, wenn die Schülerinnen und Schüler Hilfe brauchen.
	Der Lehrer/die Lehrerin unterstützt uns beim Lernen.	Die Lehrkraft unterstützt die Schülerinnen und Schüler beim Lernen.
	Unser Lehrer/unsere Lehrerin erklärt etwas so lange, bis wir es verstehen.	Die Lehrkraft erklärt etwas so lange, bis es die Schülerinnen und Schüler verstanden haben.
Skalierung	*1 = nie oder fast nie, ..., 4 = in jeder Stunde*	*1 = (fast) nie, ..., 5 = in jeder Stunde*
Klassenführung (KF)	Wir hören nicht auf das, was der Lehrer/die Lehrerin sagt.	Die Schülerinnen und Schüler hören der Lehrkraft nicht zu.
	Im Mathematikunterricht ist es laut und alles geht durcheinander.	Im Klassenzimmer ist es oft laut und es geht drunter und drüber.
	Unser Lehrer/unsere Lehrerin muss lange warten, bis Ruhe eintritt.	Die Lehrkraft muss lange warten, bis die Schülerinnen und Schüler ruhig werden.
	Wir können nicht ungestört arbeiten.	Die Schülerinnen und Schüler können nicht ungestört arbeiten.
	Wir fangen erst lange nach dem Beginn der Stunde an zu arbeiten.	Die Schülerinnen und Schüler fangen erst lange nach Beginn der Stunde an zu arbeiten.
Skalierung	*1 = in jeder Stunde, ..., 4 = nie oder fast nie*	*1 = in jeder Stunde, ..., 4 = (fast) nie*

Während für die beiden Konstrukte KU und KF nahezu identische Items verwendet werden konnten, finden sich für KA in beiden Studien lediglich einige ähnliche Items. Zur besseren Vergleichbarkeit des Merkmals KA wurde die fünfstufige Skala bei PISA 2022 linear zu einer entsprechend bei COACTIV implementierten vierstufigen Skala transformiert. Darüber hinaus wurden für das erste Item bei PISA 2022 (vgl. Tab. 3) zwei inhaltlich nahe COACTIV-Items durch Mittelwertbildung zusammengefasst.

Die Gesamtanlage beider Studien (Gesamtrationale, Administration, Stichproben, Spezifika) findet sich für PISA-Ceco in Schiepe-Tiska et al. (2025, 2026) bzw. für COACTIV in Löwen et al. (2011) und Prenzel et al. (2006).

4.3 Datenaufbereitung und -auswertung

Die sechs Items zum FDW von Lehrkräften wurden in beiden Studien zu einer (manifesten) Summenskala zusammengefasst. Die Items zur Unterrichtsqualität waren bei PISA-Ceco Teil der internationalen PISA-2022-Erhebung (Müller et al. 2025). Beim umgesetzten „within construct rotated design“ (vgl. OECD 2023) variierte die Itemauswahl zur Erfassung eines Konstrukts systematisch, sodass jede Schülerin bzw. jeder Schüler nur eine Teilmenge des vollständigen Itempools bearbeitete. Deshalb wurden zunächst die einzelnen Itemantworten für die Schülerinnen und Schüler auf Klassenebene aggregiert und anschließend manifeste Skalenwerte für KA, KU und KF gebildet, wie sie auch für COACTIV vorliegen. Der Anteil fehlender Werte liegt dabei jeweils unter zehn Prozent. Die interne Konsistenz aller gebildeten Skalen ist gut (für die jeweiligen Werte des Reliabilitätsindikators Cronbachs α siehe Tab. 4), ein Ausschluss einzelner Variablen war nicht erforderlich.

Die Analysen zu den Forschungsfragen 1, 2 und 3 basieren auf der jeweils maximal vorliegenden Stichprobe pro Konstrukt. In die Analysen zu Forschungsfrage 4 wurden nur diejenigen Lehrkräfte einbezogen, für die sowohl Daten zu FDW als auch Klassendaten zu KA, KU und KF vorlagen.

5 Ergebnisse

Zur Validierung der FDW-Kurzskala (6 Items) wurden alle im Folgenden berichteten COACTIV-Ergebnisse zusätzlich auf Basis der ursprünglichen Langskala (22 Items) berechnet. Dabei ergaben sich weder für den Schulartvergleich zwischen GY und NGY (Tab. 4) noch für den Zusammenhang mit der Berufserfahrung (vgl. 5.2) noch für die Pfadkoeffizienten (Abb. 1 und 2) substantielle Unterschiede. Weiterhin konnte für die aus sechs Items bestehende FDW-Skala anhand einer konfirmatorischen Faktorenanalyse (Multigruppenmodell mit beiden Studien) partielle skalare Messinvarianz nachgewiesen werden (der Intercept eines Items musste frei geschätzt werden). Diesbezügliche Mittelwertvergleiche zwischen beiden Studien können also sinnvoll durchgeführt werden.

Tab. 4 Stichprobengröße (*N*), interne Konsistenz (Cronbachs α), Mittelwerte (*M*) und Standardabweichungen (*SD*) in beiden Studien sowie Unterschiede zwischen nicht-gymnasialen (*NGY*) und gymnasialen (*GY*) Schularten innerhalb von COACTIV (*links*) und PISA-Ceco (*Mitte*) bzw. zwischen beiden Studien (*rechts*)

	COACTIV 2004						PISA-Ceco 2022						GY	NGY
	N	α	Gesamt *M* (*SD*)	GY *M* (*SD*)	NGY *M* (*SD*)	$d_{GY–NGY}$ (*p*-Wert)	*N*	α	Gesamt *M* (*SD*)	GY *M* (*SD*)	NGY *M* (*SD*)	$d_{GY–NGY}$ (*p*-Wert)	$d_{Ceco–COACTIV}$ (*p*-Wert)	$d_{Ceco–COACTIV}$ (*p*-Wert)
FDW: Fachdidaktisches Wissen (6 Items)	205	0,65	10,76 (4,11)	12,74 (3,84)	9,36 (3,72)	0,90** (<0,01)	125	0,70	10,20 (4,44)	11,86 (4,25)	8,52 (3,99)	0,82** (<0,01)	–0,22 (0,67)	–0,22 (0,16)
KA: Kognitive Aktivierung (4/3 Items)	185	0,78	2,85 (0,36)	2,88 (0,37)	2,84 (0,36)	0,12 (0,43)	61	0,65	2,64 (0,42)	2,78 (0,36)	2,53 (0,43)	0,64* (0,02)	–0,26 (0,25)	–0,82** (<0,01)
KU: Konstruktive Unterstützung (4 Items)	186	0,96	2,55 (0,48)	2,42 (0,51)	2,64 (0,44)	–0,48** (<0,01)	63	0,92	2,80 (0,37)	2,76 (0,37)	2,83 (0,38)	–0,19 (0,46)	0,72** (<0,01)	0,44* (0,02)
KF: Klassenführung (5 Items)	186	0,96	3,02 (0,48)	3,06 (0,53)	2,99 (0,44)	0,14 (0,36)	62	0,87	2,82 (0,40)	2,97 (0,43)	2,71 (0,33)	0,68* (0,02)	–0,18 (0,38)	–0,67** (<0,01)

Die Effektstärke *d* berechnet sich aus der Mittelwertdifferenz der jeweiligen Gruppen ($M_{GY}–M_{NGY}$ bzw. $M_{Ceco}–M_{COACTIV}$) dividiert durch deren gepoolte Standardabweichung. Nach Cohen (1992) entspricht eine Effektstärke von $d = 0{,}20$ einem kleinen, von $d = 0{,}50$ einem mittleren und von $d = 0{,}80$ einem großen Effekt

*$p \leq 0{,}05$; **$p \leq 0{,}01$

5.1 Änderungsmonitoring

Tab. 4 stellt Mittelwerte und Standardabweichungen zu den vier betrachteten Konstrukten FDW, KA, KU und KF im Überblick dar. Während links Ergebnisse zu COACTIV präsentiert werden, befinden sich in der Mitte der Tabelle die entsprechenden Resultate aus PISA-Ceco. Außerdem sind die entsprechenden Mittelwerte jeweils nach Schulart getrennt angegeben (GY bzw. NGY).

In den rechten beiden Spalten von Tab. 4 werden die Forschungsfragen 1 und 2 beantwortet (obere Zeile: FDW; untere drei Zeilen: KA, KU, KF), zur Ermöglichung eines validen Vergleichs jeweils eingeschränkt auf GY- bzw. NGY-Lehrkräfte. Die sechs negativen Vorzeichen in den letzten beiden Spalten illustrieren, dass in drei von vier Konstrukten (FDW, KF, KA) für beide Schularten die im Rahmen von PISA-Ceco befragten Lehrkräfte deskriptiv geringere Werte aufweisen als die COACTIV-Lehrkräfte (bei KU verhält es sich genau umgekehrt). Die kleinen negativen Effekte beim FDW zu Ungunsten der PISA-Ceco Lehrkräfte sind für beide Schularten jedoch nicht signifikant.

In Bezug auf die drei Unterrichtsqualitätsaspekte sind bei den NGY-Lehrkräften (rechte Spalte) die differentiellen Unterschiede jeweils signifikant, wobei die Effekte bei KA und KF zugunsten der COACTIV-Lehrkräfte auf dem 1 %-Niveau und die Effekte bezüglich KU zugunsten der PISA-Ceco Lehrkräfte auf dem 5 %-Niveau signifikant ausfallen. Bei den GY-Lehrkräften (zweite Spalte von rechts) fallen die Unterschiede bezüglich KU auf dem 1 %-Niveau signifikant zugunsten der PISA-Ceco Lehrkräfte aus.

5.2 Replikationsansatz

Tab. 4 (obere Zeile) liefert auch bereits eine erste Antwort auf Forschungsfrage 3 bezüglich der Replizierbarkeit von Zusammenhängen des FDW mit externen Variablen. Ein Aspekt dabei war die Frage nach der Stabilität von Schulartunterschieden. In der Tat lassen sich die in COACTIV gefundenen Schulartunterschiede zugunsten von GY-Lehrkräften ($d = 0{,}90^{**}$) mit den PISA-Ceco-Daten in derselben Größenordnung replizieren ($d = 0{,}82^{**}$).

Ein weiterer Aspekt war die Frage nach der Replizierbarkeit des bei COACTIV nicht nachweisbaren Zusammenhangs von FDW und Berufserfahrung (gemessen durch die Anzahl der Jahre als berufstätige Lehrkraft). Hierbei ergibt sich weder auf der Basis der COACTIV-Daten ($r = -0{,}05$; $p = 0{,}52$) noch bei den PISA-Ceco Lehrkräften ($r = 0{,}16$; $p = 0{,}27$) eine signifikante Produkt-Moment-Korrelation, was ebenfalls als Bestätigung eines fehlenden Zusammenhangs interpretiert werden kann.

Die Daten aus beiden Studien erlauben jedoch noch mehr: Bei der Bildung eines *gemeinsamen* Datensatzes lassen sich die Lehrkräfte wieder getrennt nach Schulart betrachten (diesmal über die beiden Studien hinweg). Betrachtet man nun alle sowohl damals als auch heute untersuchten 111 GY-Lehrkräfte, so zeigt sich auch bei dieser Stichprobe kein Zusammenhang zwischen FDW und Berufserfahrung ($r = 0{,}04$; $p = 0{,}76$). Bei der analogen Zusammenlegung der 157 NGY-Lehrkräfte über beide Studien hinweg ergibt sich ein nahezu identischer Wert ($r = 0{,}05$; $p = 0{,}56$).

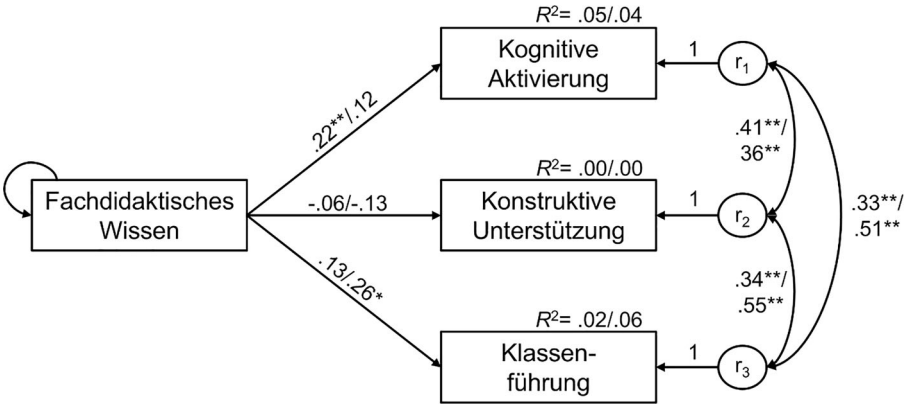


Abb. 1 Pfadmodell zu Zusammenhängen zwischen fachdidaktischem Wissen (FDW) und Merkmalen von Unterrichtsqualität (KA, KU, KF) bei COACTIV/PISA-Ceco.
Bem.: Der jeweils erste Koeffizient (links neben dem Schrägstrich) bezieht sich auf COACTIV ($N = 205$), der zweite Koeffizient (rechts neben dem Schrägstrich) auf PISA-Ceco ($N = 63$). Rechtecke symbolisieren manifeste Skalenwerte, Kreise (r_i) Fehlerterme, R^2 die durch die exogene Variable aufgeklärte Varianz. Fehlende Werte bei einzelnen Variablen (< 10 %) wurden mittels der Full-Information-Maximum-Likelihood-Prozedur geschätzt. Da das Modell gerade identifiziert ist (keine Freiheitsgrade), sind keine globalen Modellfitindizes (z. B. CFI, RMSEA, SRMR) verfügbar; $^*p \leq 0{,}05$; $^{**}p \leq 0{,}01$

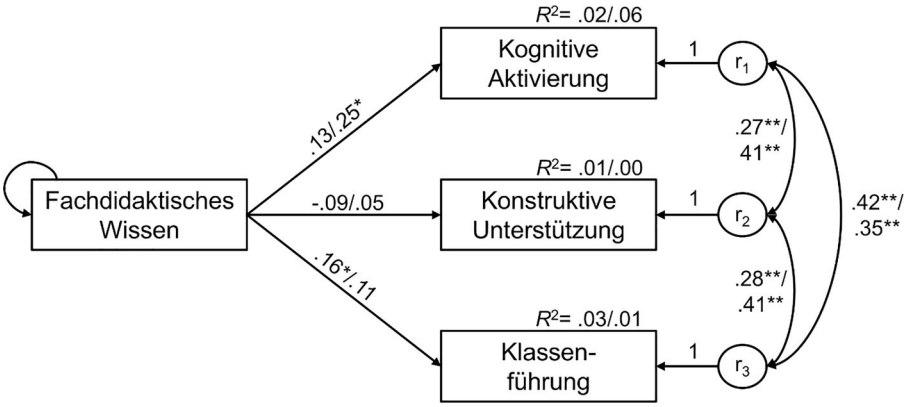


Abb. 2 Pfadmodell zu Zusammenhängen zwischen fachdidaktischem Wissen (FDW) und Merkmalen von Unterrichtsqualität (KA, KU, KF) nach nicht-gymnasialer/gymnasialer Schulart (NGY/GY).
Bem.: Der jeweils erste Koeffizient (links neben dem Schrägstrich) bezieht sich auf NGY ($N = 157$), der zweite Koeffizient (rechts neben dem Schrägstrich) auf GY ($N = 111$). Rechtecke symbolisieren manifeste Skalenwerte, Kreise (r_i) Fehlerterme, R^2 die durch die exogene Variable aufgeklärte Varianz. Fehlende Werte bei einzelnen Variablen (< 10 %) wurden mittels der Full-Information-Maximum-Likelihood-Prozedur geschätzt. Da das Modell gerade identifiziert ist (keine Freiheitsgrade), sind keine globalen Modellfitindizes (z. B. CFI, RMSEA, SRMR) verfügbar; $^*p \leq 0{,}05$; $^{**}p \leq 0{,}01$

Für *alle* 268 Lehrkräfte – d. h. bei gemeinsamer Betrachtung aller Schularten und beider Studien – ergibt sich bei einer solchen Analyse $r = 0{,}03$ ($p = 0{,}72$).

Zur Beantwortung von Forschungsfrage 4 werden alle vier untersuchten Konstrukte gemeinsam betrachtet. Da lineare Zusammenhänge nicht von den Mittelwerten der untersuchten Merkmale abhängen (sollten), müssten sich auch die Koeffizienten von Pfadmodellen zur Vorhersage der Unterrichtsqualität durch FDW im Wesentlichen replizieren lassen. Abb. 1 stellt die entsprechenden Pfadkoeffizienten für den Prädiktor FDW sowie die Korrelationen zwischen den drei Unterrichtsqualitätsaspekten KA, KU und KF pro Pfeil jeweils nebeneinander dar (vor dem Schrägstrich: COACTIV; nach dem Schrägstrich: PISA-Ceco). In der Tat lassen sich diese Koeffizienten – zumindest der Größenordnung nach – in etwa replizieren. Die Korrelationen zwischen den Residuen von KA, KU und KF entsprechen dabei näherungsweise den manifesten Zusammenhängen zwischen diesen Merkmalen.

Man beachte, dass ein signifikanter Koeffizient in einer Studie und ein nicht-signifikanter in der anderen Studie nicht bedeuten, dass dieser Unterschied zwischen den Studien auch signifikant ist. Chi-Quadrat-Differenztests für alle sechs paarweisen Koeffizientenvergleiche (Abb. 1) liefern vielmehr kein signifikantes Ergebnis. Auch wenn das FDW in COACTIV also den größten Zusammenhang mit der kognitiven Aktivierung hat, während in PISA-Ceco vor allem die Klassenführung mit dem FDW zusammenhängt, ist im Hinblick auf eine Replikation das entscheidende Resultat, dass sich weder die Pfadkoeffizienten noch die Zusammenhänge zwischen KA, KU und KF zwischen beiden Studien signifikant unterscheiden.

Auch für das Pfadmodell ist es wieder sinnvoll, zur Analyse einer größeren Stichprobe bezüglich beider Schularten jeweils über beide Studien hinweg zu aggregieren (Abb. 2). In diesem Fall liefern die sechs Chi-Quadrat-Differenztests (jeweils vor dem Schrägstrich: NGY; nach dem Schrägstrich: GY) ebenfalls keinen signifikanten Unterschied. Das Pfadmodell ist also nicht nur zeitlich und zwischen Studien relativ stabil, sondern liefert darüber hinaus auch ein vergleichbares Bild für GY- und NGY-Schularten.

6 Diskussion

6.1 Zusammenfassung und Interpretation

Im vorliegenden Beitrag wurden die Daten der beiden Erweiterungsstudien COACTIV (zu PISA 2003) und PISA-Ceco (zu PISA 2022) genutzt, um sowohl die Professionswissenskategorie FDW als auch die drei Basisdimensionen der Unterrichtsqualität (KA, KU, KF) einem Vergleich zu unterziehen (Änderungsmonitoring), aber auch, um die Beziehungen zwischen diesen Konstrukten auf Stabilität zu untersuchen (Replikationsansatz).

6.1.1 Änderungsmonitoring

Während sich bezüglich FDW bei PISA-Ceco sowohl für GY- als auch NGY-Lehrkräfte keine signifikanten Unterschiede im Vergleich zu den Ergebnissen von COAC-

TIV zeigen (Forschungsfrage 1), hat KA und KF vor allem in den nicht-gymnasialen Schularten signifikant abgenommen und gleichzeitig KU in beiden Schularten zugenommen (Forschungsfrage 2).

Zunächst muss festgehalten werden, dass die etwas abweichende Zusammensetzung der NGY-Stichprobe in beiden Studien keinen substantiellen Einfluss auf die gefundenen Unterschiede hatte (COACTIV: größerer Anteil Realschullehrkräfte; PISA-Ceco: größerer Anteil von Lehrkräften an Gesamtschulen und Schulen mit mehreren Bildungsgängen; vgl. 4.1). Auch bei einer entsprechenden Gewichtung der verschiedenen NGY-Schularten bleiben die Unterschiede in dieser Gruppe von Lehrkräften (Tab. 4: NGY: $d_{\text{Ceco-COACTIV}}$) zwischen COACTIV und PISA-Ceco in FDW, KA, KF und KU weitestgehend erhalten.

Für den Vergleich von FDW muss berücksichtigt werden, dass die Items der FDW-Kurzskala zwar wörtlich von der COACTIV-Studie übernommen wurden, aber die Bearbeitung in PISA-Ceco am Computer vor allem bei Items schwieriger gewesen sein könnte, bei denen Skizzen angefertigt werden mussten. So wurden beispielsweise beim Item „Parallelogramm“ im PISA-Ceco-Test online weniger Skizzen angefertigt als bei der damaligen P&B-Version des COACTIV-Tests (Hohenleitner et al. 2025). Das einzige Item, das in COACTIV und PISA-Ceco identisch am Computer bearbeitet werden musste, war das Item „Trapez“ (Hohenleitner et al. 2025), bei dem die Performanz zwischen beiden Studien vergleichbar war. Auch die administrativen Abweichungen (z. B. Zeithinweise bzw. keine Rückkehrmöglichkeit zu vorherigem Item bei PISA-Ceco; Tab. 2) könnten Unterschiede bei der Bearbeitung erklären.

Allerdings muss auch berücksichtigt werden, dass die PISA-Ceco Lehrkräfte während der Befragung nicht beaufsichtigt waren und somit nicht sichergestellt werden kann, dass sie die Items ohne fremde (menschliche oder technische) Hilfe beantwortet haben. Da sich mögliche Vor- und Nachteile zwischen beiden Studien in etwa ausgleichen und sich bei identischer Durchführung (Item Trapez) kein Unterschied zeigt, kann man – auch in Anbetracht der trotz aller administrativen Unterschiede festgestellten partiellen skalaren Messinvarianz (vgl. Abschnitt 5) – insgesamt mit Vorsicht konstatieren, dass das FDW von Mathematiklehrkräften in einem Zeitraum von 18 Jahren auf etwa gleichem Niveau geblieben ist.

Angesichts der breiten Rezeption der COACTIV-Ergebnisse sowie deren Nutzung in der Lehrkräftebildung stellt sich die Frage, welche Effekte realistisch von curricularen Reformen und Standardsetzungen zu erwarten sind. De facto ist in der universitären Lehrkräfteausbildung für das FDW im Vergleich zum Fachwissen nach wie vor nur ein geringer Anteil vorgesehen (z. B. in Bezug auf ECTS-Punkte). Jedenfalls scheinen sich zentrale Wissenskomponenten wie FDW durch bildungspolitische Reformen langfristig nur begrenzt verändern zu lassen.

Im Jahr 2011 äußerte Jürgen Baumert (der federführende Direktor der ersten PISA-Studie 2000 sowie einer der Projektleiter der COACTIV-Studie) große Bedenken, dass die nach dem PISA-Schock einsetzenden Reformbemühungen, die in der Tat zu einem kontinuierlichen Ansteigen der deutschen PISA-Testleistungen bis 2012 geführt haben, nun – zehn Jahre nach dem PISA-Schock – aus seiner Sicht wieder stark nachlassen würden (vgl. Köller 2025). Er befürchtete ohne bildungspolitisches Gegenlenken ein wiederholtes Absinken zukünftiger PISA-Leistungen,

welches tatsächlich beginnend von PISA 2015 bis PISA 2025 auch eingetreten ist (Lewalter et al. 2023; Greiff et al. 2026). Diese Überlegungen weisen darauf hin, dass im 18-Jahresabstand gemessene Unterschiede zeitlich nur punktuell interpretiert werden dürfen. Lineare Interpolationen für den Zeitraum dazwischen sind nur bedingt valide und auch langfristige Trends können aus einem solchen Vergleich nur mit Einschränkungen abgeleitet werden.

In Bezug auf die Unterrichtsqualität war die Administration bei beiden Studien – im Gegensatz zum FDW – nahezu identisch (geschlossene Items in einem Fragebogen, den die Schülerinnen und Schüler im Rahmen von PISA bearbeiteten). Aus Tab. 3 wird zwar deutlich, dass es bei KU und KF leichte sprachliche Abweichungen zwischen den Items gibt (z. B. „wir" statt „die Schülerinnen und Schüler"; „ruhig werden" statt „Ruhe eintritt" etc.). Bei KA sind diese Abweichungen am stärksten (hier war auch die Vergleichbarkeit der Antwortskalen – d. h., 5- vs. 4-stufig – nur durch vorherige lineare Transformation möglich). Diese Abweichungen sind jedoch nicht ausreichend, um die gefundenen – teilweise starken – Unterschiede alternativ zu erklären (auch die Zusammensetzung der NGY-Stichprobe liefert keine Erklärung hierfür; vgl. oben).

Im Folgenden soll deshalb auf eine auffallende Kompatibilität der gefundenen Unterschiede in der Unterrichtsqualität zu Befunden von Studien zum Belastungserleben von Lehrkräften während der Corona-Pandemie (Hansen et al. 2022) hingewiesen werden, die mutmaßlich auch zu Einschränkungen in der Unterrichtsqualität geführt haben könnten (Anders 2022a). Besonders interessant sind hierzu die Ergebnisse des deutschen Schulbarometers (Anders 2022b), das im April 2022 – also genau im selben Monat wie PISA-Ceco – durchgeführt wurde. Hier berichten Lehrkräfte – z. B. im Vergleich zum Vorjahr 2021 – von deutlich größeren Lernrückständen der Schülerinnen und Schüler, die offenbar erst im Laufe des wiedereinsetzenden Präsenzunterrichts zutage getreten sind. Neben der Aufarbeitung dieser Rückstände mussten Lehrkräfte in dieser Zeit Corona-Richtlinien überwachen, die Digitalisierung in Rekordtempo nachholen und bereits eine steigende Zahl von geflüchteten ukrainischen Kindern und Jugendlichen in den Unterricht integrieren (Anders 2022b).

Klusmann konstatiert (vgl. Anders 2022a), dass 95 % der Lehrkräfte in dieser Zeit einen Anstieg negativer Verhaltensweisen bei Schülerinnen und Schülern wie etwa Konzentrationsmängel oder Motivationsprobleme beobachtet haben, was vollständig kompatibel mit dem gemessenen Abfall von KA und KF (v. a. auch in NGY-Schularten) bei PISA-Ceco ist. Weiterhin berichtet ein Großteil der Lehrkräfte, dass die Schulen den Jugendlichen in dieser Zeit weder eine adäquate Lernunterstützung noch eine Förderung des psychischen Wohlbefindens bieten konnten, was beides eigentlich erforderlich gewesen wäre (Anders 2022b). Diese Wahrnehmungen in der ausklingenden Corona-Situation könnten durchaus zu verstärkten Bemühungen von Lehrkräften im Hinblick auf die Förderung individueller KU von Schülerinnen und Schülern geführt haben, wie es in PISA-Ceco gemessen wurde (Tab. 4).

6.1.2 Replikationsansatz

Die Erkenntnisse zur Replizierbarkeit von Zusammenhängen stimmen insgesamt zuversichtlich. Bezüglich des FDW konnten zwei bei COACTIV zunächst überraschende Befunde – nämlich der Vorteil der GY-Lehrkräfte und der fehlende Zusammenhang mit der Berufserfahrung – auch in den PISA-Ceco-Daten gefunden werden (Forschungsfrage 3). Beide FDW-Effekte sind bei Mathematiklehrkräften zeitlich stabil und müssen nicht mehr durch Spezifika der COACTIV-Stichprobe erklärt werden.

Der FDW-Vorteil von GY-Lehrkräften konnte mittlerweile auch in anderen Unterrichtsfächern nachgewiesen werden (z. B. Deutsch, Englisch, Physik, Religion, Musik; Krauss et al. 2017). In den Daten der COACTIV-Studie waren bei statistischer Kontrolle des Fachwissens – das in PISA-Ceco nicht erhoben wurde – die NGY-Lehrkräfte beim FDW im Vorteil (Krauss et al. 2017). Demnach gibt es zwei mögliche Wege zu einem hohen fachdidaktischen Wissen, nämlich einen direkten über eine umfangreichere Pädagogik- und Didaktikausbildung und einen indirekten über Fachwissen als Voraussetzung von FDW (vgl. Fukaya et al. 2025; Lindl und Krauss 2017).

Auch der fehlende Zusammenhang des Professionswissens mit der Berufserfahrung hat sich mittlerweile in weiteren Studien und in anderen Unterrichtsfächern gezeigt (Großschedl et al. 2014; Kirschner et al. 2017; zusammenfassend in Lindl und Krauss 2017). Laut der Deliberate-Practice-Theorie (Ericsson et al. 1993) ist die bloße Berufsausübung für eine substantielle Steigerung von Expertise nicht ausreichend, hierzu ist vielmehr intrinsisch motiviertes, gezieltes, wiederholtes Arbeiten an eigenen Schwächen erforderlich, idealerweise durch permanentes Experten-Feedback unterstützt. Eine aktuelle Sekundäranalyse der COACTIV-Daten zeigt in der Tat, dass sich auch bei zahlreichen weiteren Konstrukten der professionellen Lehrkraftkompetenz und der Unterrichtsqualität kein Zusammenhang mit der Berufserfahrung nachweisen lässt (Krauss et al. 2025).

Darüber hinaus ließ sich ein Pfadmodell zur Vorhersage der Unterrichtsqualität durch FDW replizieren (Forschungsfrage 4), d. h. die sechs Koeffizienten unterscheiden sich nicht signifikant zwischen beiden Studien (Abb. 1). Allerdings gibt es zwischen PISA-Ceco und COACTIV eine Verschiebung bei den Signifikanzen. Während das FDW in COACTIV prädiktiv für die kognitive Aktivierung war, zeigte sich in PISA-Ceco ein signifikanter Effekt für die Klassenführung. Dies ist ein bemerkenswerter Befund, da die bisherige Forschungslage eher den in COACTIV gefundenen Effekten entspricht (z. B. Förtsch et al. 2018; Jentsch et al. 2021). Möglicherweise hat das FDW aber auch einen ordnenden Effekt, etwa weil schlechte Erklärungen zur Unruhe in der Klasse beitragen könnten.

6.2 Limitationen und Grenzen der Studie

Limitierend für die Analysen ist die geringe Stichprobengröße bei PISA-Ceco, wodurch die Ergebnisse gerade bei weiteren Untergliederungen (z. B. NGY vs. GY) nur eingeschränkt belastbar und mit Vorsicht zu interpretieren sind. So lag beispielsweise in den Pfadanalysen das Verhältnis der zu schätzenden Parameter zur

Personenanzahl in COACTIV bei etwa 20:1, in PISA-Ceco aber nur bei etwa 6:1 und somit an der vertretbaren Untergrenze (mindestens 5:1; Kline 2023). Es war darüber hinaus keine latente Modellierung möglich, sondern nur gerade identifizierbare Pfadmodelle, so dass auch keine Modellfitindizes bestimmbar waren. Durch die identische Art der Stichprobengewinnung der Lehrkräfte in beiden Studien jeweils im Rahmen von PISA besteht allerdings die Hoffnung, dass ein Vergleich – bei aller gegebenen Vorsicht – gerechtfertigt ist.

Im Hinblick auf geringe (Teil-)Stichprobengrößen ist beim Einsatz identischer Items für Replikationsansätze auch die Zusammenlegung von Datensätzen verschiedener Studien eine interessante Option (vgl. 5.2). Insofern könnten die beiden „Nullkorrelationen" von FDW und Berufserfahrung in einer reinen GY-Stichprobe und in einer reinen NGY-Stichprobe darauf hinweisen, dass durch eine solche Zusammenlegung stichprobenspezifisches Overfitting ggf. abgemildert werden könnte.

Nicht nur die Erfassung des FDW in COACTIV, sondern auch die Messung der Unterrichtsqualität erfolgte in beiden Studien umfassender. Für einen Vergleich konnten jeweils nur die Aspekte herangezogen werden, die in beiden Studien identisch oder zumindest sehr ähnlich gemessen wurden. So waren die individuelle Unterstützung für Schülerinnen und Schüler (Tab. 3, Mitte) als auch die Disziplin im Klassenzimmer (Tab. 3, unten) jeweils nur Teilaspekte von KU und KF (die allerdings hierfür jeweils zentrale Indikatoren darstellen; Kunter und Voss 2011). KA wurde in beiden Begleitstudien ebenfalls vielschichtiger untersucht, die Items (Tab. 3, oben) stellen dabei die „jeweils ähnlichsten" Items dar, die aus Skalen wie z. B. *Argumentieren* oder *Alltagsbezug* entnommen wurden (für alternative Zugänge zur kognitiven Aktivierung in PISA-Ceco siehe Hohenleitner et al. 2026; bzw. Heinle et al. 2026).

Bezüglich der Modellierung von Unterrichtsqualität gibt es mittlerweile auch Weiterentwicklungen des Basismodells, die mehr als drei Dimensionen der Unterrichtsqualität beinhalten (Alp Christ et al. 2024; Brunner und Star 2024; Praetorius et al. 2020). Praetorius et al. (2018) weisen allerdings explizit darauf hin, dass die drei Basisdimensionen für eine sparsame Erhebung auch nach aktuellem Forschungsstand gut geeignet sind.

Für einen Mittelwertunterschied von Konstrukten, die in einem Abstand von fast 20 Jahren erhoben wurden, kann es aufgrund des multiplen Transformationsgeschehens mehrere alternative Erklärungen geben. Ohne experimentelle Kontrolle sind kausale Schlussfolgerungen nicht valide möglich, über isolierte Effekte wie z. B. die der Pandemie oder der Einführung von Bildungsstandards lässt sich deshalb auf Basis der vorliegenden Analysen nur spekulieren (im vorliegenden Beitrag werden solche Spekulationen über die Rolle der Corona-Pandemie vorgenommen, Überlegungen zur Bedeutung der Bildungsstandards finden sich in Hohenleitner et al. 2026).

6.3 Fazit

Während das FDW von Mathematiklehrkräften im 18-Jahresabstand relativ stabil geblieben ist, zeigen sich Unterschiede in der Unterrichtsqualität zwischen den beiden PISA-Begleitstudien 2003/2004 und 2022. Die gefundenen Unterschiede in

KA, KU und KF sind kompatibel mit den Einschätzungen von Lehrkräften in der ausklingenden Corona-Pandemie, die von unkonzentrierten und schwer zu aktivierenden Schülerinnen und Schülern berichten, die auf deutlich mehr konstruktive Unterstützung angewiesen sind, um Wissensrückstände aufzuholen.

In Bezug auf Forderungen aus der Replikationskrise indizieren die vorliegenden Analysen eine gewisse Stabilität von Zusammenhangsmaßen. Da die betrachteten Konstrukte (FDW, KA, KU und KF) sowie die hierfür eingesetzten Messinstrumente zum Standardrepertoire der empirischen Bildungsforschung zählen, stimmen diese Ergebnisse zuversichtlich und sind auch von Bedeutung für die Verlässlichkeit und Generalisierbarkeit der (quantitativen) empirischen Bildungsforschung im Allgemeinen.

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. Krauss, S. Hohenleitner, N. Stegmüller, A. Schiepe-Tiska, K. Reiss und A. Lindl geben an, dass kein Interessenkonflikt 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

Alp Christ, A., Capon-Sieber, V., Köhler, C., Klieme, E., & Praetorius, A. K. (2024). Revisiting the Three Basic Dimensions model: A critical empirical investigation of the indirect effects of student-perceived teaching quality on student outcomes. *Frontline Learning Research*, *12*(1), 66–123. https://doi.org/10.14786/flr.v12i1.1349.

Anders, F. (2022a). *Deutsches Schulbarometer: Gesunde Lehrkräfte sind die Voraussetzung für gelingende Lernprozesse*. https://deutsches-schulportal.de/unterricht/gesunde-lehrkraefte-sind-die-voraussetzung-fuer-gelingende-lernprozesse/. Zugegriffen: 9. Juni 2022.

Anders, F. (2022b). *Deutsches Schulbarometer: Erschöpft: Lehrer-Umfrage zeigt ein System an der Belastungsgrenze*. https://deutsches-schulportal.de/unterricht/deutsches-schulbarometer-auswirkungen-corona-pandemie-april-2022/. Zugegriffen: 9. Juni 2022.

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.

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_3.

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M., & Yi-Miau, T. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente*. Max-Planck-Institut für Bildungsforschung.

Bienefeld, M., Böhm-Kasper, O., & Demmer, C. (2020). Highly recommended and yet neglected. The rarity of replication studies in educational science. *Journal for educational research online*, *12*(3), 3–22.

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*, 101600. https://doi.org/10.1016/j.learninstruc.2022.101600.

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

Bruckmaier, G., Krauss, S., Blum, W., & Neubrand, M. (2023). Die Messung des fachdidaktischen Wissens in der COACTIV-Studie. In S. Krauss & A. Lindl (Hrsg.), *Professionswissen von Mathematiklehrkräften: Implikationen aus der Forschung für die Praxis*. *Mathematik Primarstufe und Sekundarstufe I + II* (S. 111–142). Springer Spektrum.

Brunner, E., & Star, J. R. (2024). The quality of mathematics teaching from a mathematics educational perspective: what do we actually know and which questions are still open? *ZDM Mathematics Education*, *56*, 775–787. https://doi.org/10.1007/s11858-024-01600-z.

Cai, J., Morris, A. K., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018). The role of replication studies in educational research. *Journal for Research in Mathematics Education*, *49*(1), 2–8. https://doi.org/10.5951/jresematheduc.49.1.0002.

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., & Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (S. 77–94). Springer Nature. https://doi.org/10.1007/978-981-13-5898-2_2.

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.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne*, *49*(3), 182–185. https://doi.org/10.1037/a0012801.

Ericsson, K. A., Krampe, R. Th , & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406. https://doi.org/10.1037/0033-295X.100.3.363.

Förtsch, S., Förtsch, C., Von Kotzebue, L., & Neuhaus, B. J. (2018). Effects of teachers' professional knowledge and their use of three-dimensional physical models in biology lessons on students' achievement. *Education Sciences*, *8*(3), 118. https://doi.org/10.3390/educsci8030118.

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.

Greiff, S., Diedrich, J., Kastorff, T., Goldhammer, F., & Köller, O. (Hrsg.) (2026). *PISA 2025. Bildung im Spannungsfeld von Innovation und Transformation. Zusammenfassung.* Waxmann. https://www.waxmann.com/buecher/PISA-2025

Großschedl, J., Mahler, D., Kleickmann, T., & Harms, U. (2014). Content-related knowledge of biology teachers from secondary schools: Structure and learning opportunities. *International Journal of Science Education*, *36*(14), 2335–2366. https://doi.org/10.1080/09500693.2014.923949.

Hansen, J., Klusmann, U., & Hanewinkel, R. (2022). Emotionale Erschöpfung und Berufszufriedenheit von Lehrpersonal während der COVID-19-Pandemie. *Bundesgesundheitsblatt*, *65*, 776–783. https://doi.org/10.1007/s00103-022-03554-7.

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

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, *42*(2), 371–406. https://doi.org/10.3102/00028312042002371.

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.

Hohenleitner, S., Krauss, K., Stegmüller, N., & Lindl, A. (2026). Die Rolle des fachdidaktischen Wissens und subjektiver Einstellungen gegenüber den Bildungsstandards für kognitiv aktivierenden Mathematikunterricht in PISA-Ceco. *Zeitschrift für Erziehungswissenschaft.*

Jentsch, A., Schlesinger, L., Heinrichs, H., Kaiser, G., König, J., & Blömeke, S. (2021). Erfassung der fachspezifischen Qualität von Mathematikunterricht: Faktorenstruktur und Zusammenhänge zur professionellen Kompetenz von Mathematiklehrpersonen. *Journal für Mathematik-Didaktik*, *42*, 97–121. https://doi.org/10.1007/s13138-020-00168-x.

Kelcey, B., Hill, H. C., & Chin, M. J. (2019). Teacher mathematical knowledge, instructional quality, and student outcomes: a multilevel quantile mediation analysis. *School Effectiveness and School Improvement*, *30*(4), 398–431. https://doi.org/10.1080/09243453.2019.1570944.

Kirschner, S., Szczudlek, M., Tepner, O., Borowski, A., Fischer, H. E., Lenske, G., Leutner, D., Neuhaus, B. J., Sumfleth, E., Thillmann, H., & Wirth, J. (2017). Professionswissen in den Naturwissenschaften (ProwiN). In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 113–130). Springer. https://doi.org/10.1007/978-3-658-07274-2_7.

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). BMBF.

Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling* (5. Aufl.). Guilford Press.

KMK. Sekretariat der Kultusministerkonferenz (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf

KMK. Sekretariat der Kultusministerkonferenz (2008). *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

Köller, O. (2025). PISA gestern und heute: Entwicklungslinien über 25 Jahre [Vortrag, 26. Nov. 2025]. FES: 25 Jahre empirische Wende der Bildungspolitik – Versuch einer Bilanz, Berlin.

König, J., Blömeke, S., Jentsch, A., Schlesinger, L., Felske, C., Musekamp, F., & Kaiser, G. (2021). The links between pedagogical competence, instructional quality, and mathematics achievement in the lower secondary classroom. *Educational Studies in Mathematics*, *107*, 189–212. https://doi.org/10.1007/s10649-020-10021-0.

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., 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., Hofmann, B., Kirchhoff, P., & Mulder, R. H. (2017). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik*. Waxmann.

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N., & Blum, W. (2020). Competence as a continuum in the COACTIV study: the "cascade model". *ZDM Mathematics Education*, *52*, 311–327. https://doi.org/10.1007/s11858-020-01151-z.

Krauss, S., Hohenleitner, S., Steib, N., Lindl, A., & Hilbert, S. (2025). Does more teaching experience mean more expertise? A nonlinear perspective from the COACTIV study. *ZDM Mathematics Education*, *57*(6), 1235–1250. https://doi.org/10.1007/s11858-025-01748-2.

Kunter, K., & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, K. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften, Ergebnisse des Forschungsprogramms COACTIV* (S. 85–113). Waxmann. https://doi.org/10.31244/9783830974338.

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

Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O., & Reiss, K. (2023). *PISA 2022: Analyse der Bildungsergebnisse in Deutschland*. Waxmann.

Lindl, A., & Krauss, S. (2017). Transdisziplinäre Perspektiven auf domänenspezifische Lehrerkompetenzen. Eine Metaanalyse zentraler Resultate des Forschungsprojekts FALKO. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R.H. Mulder (Hrsg.), *FALKO – Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik* (S. 381–438). Waxmann.

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. Mathematik Primarstufe und Sekundarstufe I + II* (S. 1–35). Springer Spektrum. https://doi.org/10.1007/978-3-662-64381-5_1.

Löwen, K., Baumert, J., Kunter, M., Krauss, S., & Brunner, M. (2011). Methodische Grundlagen des Forschungsprogramms. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 69–84). Waxmann.

Mang, J., Wagner, S., Gomolka, J., & Meinck, S. (2023). Stichprobenbeschreibung PISA 2022. In D. Lewalter, J. Diedrich, F. Goldhammer, O. Köller & K. Reiss (Hrsg.), *PISA 2022: Analyse der Bildungsergebnisse in Deutschland* (S. 371–384). Waxmann.

Müller, M., Teufele, L., Bias, A., Diedrich, J., González Rodríguez, E., González Uriarte, A., Heine, J.H., Heinle, M., Kastorff, T., Mang, J., Patzl, S., Seßler, S., Todtenhöfer, P., Tupac-Baier, A., Weis, M., Ziernwald, L., & Lewalter, D. (2025). *PISA 2022 Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Waxmann. https://doi.org/10.31244/9783818850616.

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

Praetorius, A.K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality. The German framework of Three Basic Dimensions. *ZDM Mathematics Education*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4.

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.

Prenzel, M., Carstensen, C.H., Schöps, K., & Maurischat, C. (2006). Die Anlage des Längsschnitts bei PISA 2003. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 29–62). Waxmann.

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.

Rothland, M. (2024). Auf Sand gebaut? Über den Versuch, die Basisdimensionen der Unterrichtsqualität in erziehungswissenschaftlicher „Theorie" zu verankern. *Zeitschrift für Bildungsforschung*, *14*(1), 41–53. https://doi.org/10.1007/s35834-024-00411-2.

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The Influence of Teachers' Knowledge on Student Learning in Middle School Physical Science Classrooms. *American Educational Research Journal*, *50*(5), 1020–1049. https://doi.org/10.3102/0002831213477680.

Sarkar, M., Gutierrez-Bucheli, L., Yip, S. Y., Lazarus, M., Wright, C., White, P. J., Ilic, D., Hiscox, T. J., & Berry, A. (2024). Pedagogical content knowledge (PCK) in higher education: A systematic scoping review. *Teaching and Teacher Education*, *144*(2), 104608. https://doi.org/10.1016/j.tate.2024.104608.

Schiepe-Tiska, A., Heinle, A., Todtenhöfer, P., Heine, J. H., Reinhold, F., Krauss, S., Holzberger, D., Lewalter, D., & Reiss, K. (2023). Wie sieht der aktuelle Mathematikunterricht in Deutschland aus? Befunde aus PISA 2022 und PISA-Ceco. In D. Lewalter, J. Diedrich, F. Goldhammer, O. Köller & K. Reiss (Hrsg.), *PISA 2022: Analyse der Bildungsergebnisse in Deutschland* (S. 199–236). Waxmann.

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. https://doi.org/10.31244/9783818850692.

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.*

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.

Yang, X., Li, X., Chen, X., König, J., & Kaiser, G. (2025). The impact of Chinese mathematics teachers' competence on students' affective learning outcomes and the mediating role of instructional quality. *ZDM Mathematics Education*, *57*, 1–14. https://doi.org/10.1007/s11858-025-01738-4.

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