



Nutzung und Wirkung eines kognitiv aktivierenden Unterrichts bei nicht-gymnasialen Schülerinnen und Schülern

Samuel Merk · Andrea Batzel-Kremer · Thorsten Bohl ·
Marc Kleinknecht · Timo Leuders

Eingegangen: 9. November 2018 / Überarbeitet: 11. Februar 2021 / Angenommen: 18. Februar 2021 /
Online publiziert: 26. April 2021
© Der/die Autor(en) 2021

Zusammenfassung Bisherige Untersuchungen zu kognitiv aktivierendem Unterricht belegen überwiegend positive Effekte auf die Leistungsentwicklung der Schülerinnen und Schüler. Jedoch wird die Wirkung eines solchen Lernangebots bei leistungsschwächeren Lernenden selten untersucht. Der Fokus des vorliegenden Beitrags liegt daher auf der Nutzung und Wirkung kognitiv aktivierenden Unterrichts bei Schülerinnen und Schülern an Haupt- und Realschulen. Es wurde untersucht, (1) ob es einen Zusammenhang zwischen der beobachteten kognitiven Aktivierung und dem Ausmaß selbstberichteter kognitiver Aktivität der Schülerinnen und Schüler gibt und (2) ob sich die in den beobachteten Unterrichtsstunden erzielten Lernzuwächse der Schülerinnen und Schüler mit dem Ausmaß kognitiver Aktivierung im Unterricht und der selbstberichteten kognitiven Aktivität erklären lassen. Dazu wurden 60 videographierte Mathematikstunden hinsichtlich ihres Ausmaßes an kognitiver Aktivierung hochinferent bewertet, die entsprechenden Schülerinnen und Schüler ($N=608$) aus der achten Jahrgangsstufe an Haupt- und Realschulen zu ihrer kognitiven Aktivität im Unterricht befragt sowie ihre Leistung im Prä-Post-Design erfasst. Die Vorhersage der kognitiven Aktivität bzw. des Leistungszuwachses wurde mithilfe eines Mehrebenen-Designs untersucht. Erwartungswidrig zeigte sich Evi-

Die AutorInnen Samuel Merk und Andrea Batzel-Kremer haben zu gleichen Teilen zum Manuskript beigetragen.

S. Merk · A. Batzel-Kremer (✉) · T. Bohl
Institut für Erziehungswissenschaft, Eberhard Karls Universität Tübingen, Münzgasse
22–30, 72070 Tübingen, Deutschland
E-Mail: andrea.batzel-kremer@uni-tuebingen.de

M. Kleinknecht
Institut für Bildungswissenschaft, Leuphana Universität Lüneburg,
Universitätsallee 1, 21335 Lüneburg, Deutschland

T. Leuders
Institut für Mathematische Bildung, Pädagogische Hochschule Freiburg,
Kunzenweg 21, 79117 Freiburg, Deutschland

denz für die Nichtexistenz eines Zusammenhangs zwischen kognitiver Aktivierung und kognitiver Aktivität, jedoch bestätigen die Ergebnisse die Annahme, dass der Leistungszuwachs der Schülerinnen und Schüler unter anderem durch das Ausmaß kognitiver Aktivierung in den Lernbegleitungsphasen moderiert wird. Implikationen für die weitere Forschung werden diskutiert.

Schlüsselwörter Kognitive Aktivierung · Kognitive Aktivität · Unterrichtsqualität · Videoanalysen · Lehr-Lernforschung

Use and effect of cognitively activating instruction in classes with students from non-academic tracks

Abstract Previous studies on cognitively activating instruction have mainly shown positive effects on students' performance. However, the impact of such learning opportunities is rarely explored among educationally disadvantaged learners. Hence, this article focuses on the use and effect of cognitively activating instruction with students in non-academic track schools. We examined (1) whether there is a correlation between the observed cognitive activation and the self-reported cognitive activity of the students, and (2) whether the degree of cognitive activation and self-reported cognitive activity could explain the learning gains of the students in the videotaped lessons. For this purpose, cognitive activation was assessed in 60 videotaped mathematics lessons using high-inference rating items; the corresponding eight grade students ($N=608$) were asked about their cognitive activity in the videotaped lessons and their performance was assessed in a pre-post design. Multi-level analyses predicted relevant factors for cognitive activity and performance of the students. Against our expectation, the results revealed that cognitive activation and cognitive activity are not related. Nonetheless, the findings confirm the assumption that students' performance is predicted, among other things, by the extent of cognitive activation moderated by pre-test scores. Implications for further research are discussed.

Keywords Cognitive activation · Cognitive activity · Instructional quality · Video analysis · Research into teaching and learning

1 Einleitung

In jüngerer Zeit wird im deutschsprachigen Raum Unterrichtsqualität vermehrt in Form dreier übergeordneter Faktoren konzeptualisiert (Klieme et al. 2001): (1) kognitive Aktivierung, worunter das Potenzial verstanden wird, die Lernenden zur vertieften Auseinandersetzung mit dem Lerngegenstand anzuregen; (2) Konstruktive Unterstützung und die Schaffung eines schülerorientierten Lernklimas; sowie (3) eine strukturierte, effektive und störungspräventive Unterrichtsführung (Klieme et al. 2006; Pianta und Hamre 2009). Diese Faktoren (auch Tiefendimensionen der Unterrichtsqualität genannt) haben sich in der nationalen wie internationalen Lehr-

Lernforschung der vergangenen zwanzig Jahre immer wieder als überfachlich erwiesen.

Fraglich ist allerdings, ob alle Schülerinnen und Schüler gleichermaßen von qualitativem Unterricht profitieren (Kunter und Ewald 2016). Kritisch hinterfragt wird beispielsweise die Passung des Unterrichtsangebots zu den Voraussetzungen der Lernenden, so etwa in der Diskussion um entdeckendes Lernen vs. direkter Instruktion (Alfieri et al. 2011). Gerade wenn Lernende nur über schlecht strukturiertes Vorwissen verfügen, sind offene Instruktionmethoden nicht effektiv (Kirschner et al. 2006). Ebenso wird jedoch kritisiert, dass Strukturierung und herausfordernde Lernangebote als konträre Konstrukte aufgefasst werden (Klieme et al. 2001).

Die skizzierten Diskussionslinien verdeutlichen, dass Unterrichtsqualität als ein Angebot aufgefasst werden kann, dessen Wirkung je nach Nutzung der Lernenden differenziell ausfallen kann (Klieme und Rakoczy 2008), wobei diese Nutzung von Eigenschaften des Lernenden abhängig ist. Mit Blick auf die Forschung zu kognitiv aktivierendem Unterricht fällt dabei zweierlei auf: Zum einen wird kognitive Aktivierung als Qualitätsmerkmal des Unterrichtsangebots vor allem bei Lernenden mit eher günstigen Lernvoraussetzungen in seiner leistungsförderlichen Wirkung untersucht (Kunter et al. 2013; Lipowsky et al. 2009), zum anderen wird die *Nutzung* des Angebots in Form verstärkter kognitiver Lernaktivitäten auf Seiten der Schülerinnen und Schüler nur selten analysiert (Helm 2016; Hugener 2008; Lipowsky et al. 2009; Rakoczy et al. 2007, 2010). Dem Desiderat, ein kognitiv aktivierendes Unterrichtsangebot bei Lernenden mit (durchschnittlich) vergleichsweise ungünstigen Lernvoraussetzungen in seiner Nutzung und leistungsförderlichen Wirkung zu untersuchen, widmet sich daher die vorliegende Studie.

2 Theoretischer Hintergrund und Forschungsstand

2.1 Kognitive Aktivierung

Dem Qualitätsmerkmal *Kognitive Aktivierung* liegt die auf soziokonstruktivistischen Theorien basierende Annahme zugrunde, dass Lernen ein ko-konstruktiver Prozess ist, bei dem Lernende in aktiver Auseinandersetzung mit dem Lerngegenstand ihre vorhandenen kognitiven Strukturen zum Erwerb eines konzeptuellen Verständnisses verändern müssen. Die Lehrkraft kann daher das Lernen der Schülerinnen und Schüler nicht herbeiführen, sondern lediglich Lerngelegenheiten für kognitive Konstruktionsleistungen zur Verfügung stellen (Baumert und Köller 2000).

Solchermaßen kognitiv aktivierende Lerngelegenheiten werden im Mathematikunterricht durch die ausgewählten Aufgaben und deren Implementation im Unterricht bestimmt: Komplexe Aufgaben, die Aktivierung des Vorwissens sowie die Aufforderung der Lehrkraft an die Lernenden, ihre Lösungen zu diskutieren, kennzeichnen ein kognitiv aktivierendes Handeln der Lehrkraft (Klieme et al. 2006; Kunter et al. 2006). Dabei wird angenommen, dass eine kognitive Herausforderung mit den Fähigkeiten der Lernenden variiert, jedoch für alle Lernenden auf dem für sie adaptierten Schwierigkeitsniveau möglich ist (Baumert et al. 2004).

Für die Wirksamkeit kognitiv aktivierenden Unterrichts auf die Leistungsentwicklung der Schülerinnen und Schüler gibt es einige Hinweise. Baumert und Köller (2000) zeigten etwa, dass die über Schülerbefragungen ermittelte Verständnisorientierung systematisch zur Erklärung der Leistungsvarianz beiträgt. Klieme et al. (2001) fanden einen korrelativen Zusammenhang zwischen der beobachteten kognitiven Aktivierung und dem Leistungszuwachs auf Klassenebene. Analysen zur Wirksamkeit kognitiv aktivierenden Unterrichts von Kunter et al. (2013) ergaben, dass das kognitive Potenzial der Aufgaben einen signifikanten Prädiktor für die Mathematikleistung darstellt. Die beobachtete kognitive Aktivierung hat bei Lipowsky et al. (2009) ebenfalls einen positiven Effekt auf die Leistungsentwicklung der Schülerinnen und Schüler. Kunter (2005) und Waldis et al. (2010) hingegen konnten keinen Effekt zwischen kognitiver Aktivierung und der Leistung belegen. Dies ist insofern von besonderem Interesse, als beide Studien eine der wenigen im deutschsprachigen Raum sind, die alle Schularten in ihre Analysen eingeschlossen und somit auch kognitiv aktivierenden Unterricht bei leistungsschwächeren Lernenden untersucht haben.

Insgesamt handelt es sich bei kognitiver Aktivierung um ein breit untersuchtes Konstrukt, für das vor allem auf der Grundlage von Beobachter- und Aufgabenanalyse-daten belastbare Evidenz für die positive Wirkung auf die Leistungsentwicklung der Lernenden vorliegt (Praetorius et al. 2018). Allerdings fehlen Wirkungsstudien bei nicht-gymnasialen Schülerinnen und Schülern.

2.2 Elaborationsstrategien

Ob kognitiv aktivierender Unterricht bei den Schülerinnen und Schülern wirksam ist, so die Annahme, hängt entscheidend davon ab, ob das Lernangebot in Form verstärkter kognitiver Aktivität genutzt wird (Klieme und Rakoczy 2008). Die vertiefte Auseinandersetzung der Lernenden mit dem Unterrichtsgegenstand kann unter anderem über die genutzten Lernstrategien erfasst werden (Friedrich und Mandl 1992). Nach Weinstein und Mayer (1986) werden dabei u. a. Elaborations- und Organisationsstrategien als Tiefenverarbeitungsstrategien unterschieden: Elaborationsstrategien dienen dazu, sinnkonstituierend einen Lerngegenstand zu verstehen. Organisationsstrategien zielen darauf ab, Verbindungen zwischen neuen Wissens-elementen herauszuarbeiten.

Bereits frühe experimentelle Studien (Bobrow und Bower 1969) belegen, dass auf das Verständnis abzielende Tiefenverarbeitungsstrategien zu einer besseren Behaltensleistung beitragen als oberflächliche Wiederholungsstrategien wie etwa das mehrmalige Lesen eines Satzes. Artelt (1999) untersuchte handlungsnah erhobene Tiefenverarbeitungsstrategien und deren quantitativen und qualitativen Effekt auf den Lernerfolg: Zum einen konnte sie belegen, dass die Anwendung von Tiefenverarbeitungsstrategien zu einer umfassenderen Wiedergabe des gelesenen Textes führt, zum anderen korrelierte deren Einsatz mit der Konzentration der Lernenden auf die wesentlichen Sinngehalte des wiederzugebenden Textes. Auch Leopold und Leutner (2002) belegten enge Zusammenhänge zwischen der Leistung und dem tiefenorientierten Strategiegebrauch der Lernenden.

Andere Studien kommen zu divergenten Ergebnissen: Spörer und Brunstein (2005) befragten Schülerinnen und Schüler der 8. Klasse zu ihren Lernstrategien. Während sich die mittels Interviewanalyse ermittelten Tiefenverarbeitungsstrategien als positive Prädiktoren für die ermittelten Leistungsindikatoren erwiesen, war der Zusammenhang der Leistung mit den mittels Fragebogen erfassten Tiefenverarbeitungsstrategien nicht signifikant.

Qualitativ vertiefende Analysen geben Aufschluss darüber, dass sich leistungsschwächere Schülerinnen und Schüler beim Einsatz von Tiefenverarbeitungsstrategien im Vergleich zu leistungsstärkeren Lernenden unterscheiden: Zimmermann und Martínez-Pons (1990) sowie Lehtinen (1992) konnten zeigen, dass leistungsstarke Schülerinnen und Schüler häufiger Lernstrategien nutzen. Entsprechend belegten Spörer (2004) sowie Artelt et al. (2003), dass Schülerinnen und Schüler am Gymnasium über mehr Elaborationsstrategien berichten als Lernende nicht-gymnasialer Schularten.

2.3 Zusammenhangsbefunde zwischen Unterrichtsangebot und kognitiver Aktivität der Lernenden

Die divergenten Ergebnisse zur Effizienz von Lernstrategien werden theoretisch auch auf den Anregungsgehalt der jeweiligen zugrundeliegenden Lernsituation (z. B. problemorientiertes Lernen) zurückgeführt. McNamara et al. (1996) untersuchten anhand unterschiedlicher Texte, welcher Kohärenzgrad für eine aktive Verarbeitung förderlich ist. Insbesondere bei problemorientierten Testaufgaben, die eine tiefgehende Verarbeitung des Textes voraussetzten, erzielten leistungsschwächere Lernende bessere Ergebnisse bei kohärenten Texten, während leistungsstärkere Lernende nach der Lektüre eines weniger kohärenten Textes besser abschnitten.

Im Rahmen der Videostudie des Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN) konnten Seidel et al. (2002, 2006) zeigen, dass der Anteil unterrichtlicher Arbeitsformen (Sozialformen) in keinem systematischen Zusammenhang mit den von den Lernenden berichteten Elaborationen steht. Die effektive Zeitnutzung hingegen kovarierte mit den Elaborationen der Lernenden: Schülerinnen und Schüler einer Gruppe mit niedriger Zeitnutzung berichteten signifikant weniger vertiefende Elaborationen. Weitere Analysen belegen, dass elaborierende Äußerungen von Schülerinnen und Schülern im Klassengespräch mit verstärkten kognitiven Lernaktivitäten aller Lernenden einhergehen (Seidel 2003). Die Befunde deuten zusammenfassend darauf hin, dass vor allem die Tiefendimensionen des Unterrichts mit den Elaborationen der Lernenden zusammenhängen.

Diese Vermutung wird in weiteren Feldstudien bestätigt: Für die Strukturierung des Unterrichts konnten Rakoczy et al. (2007) einen positiven Effekt auf die nachvollziehende Elaboration (z. B. die Erklärungen der Lehrkraft nachvollziehend verstehen) der Schülerinnen und Schüler aufzeigen; für vertiefende Elaborationen (z. B. sich ein eigenes, zum neuen Stoff passendes Beispiel zu überlegen) ließ sich dieser Effekt jedoch nicht replizieren (Rakoczy et al. 2010). Helm (2016) belegte in einer breit angelegten Fragebogenstudie positive Effekte von Merkmalen kognitiver Aktivierung auf die elaborierte Auseinandersetzung der Schülerinnen und Schüler mit Aufgaben. Bei Hugener (2008) hingegen berichteten die Schülerinnen und Schüler

keine erhöhten kognitiven Lernaktivitäten, wenn der videographierte Unterricht, an dem sie teilgenommen hatten, als kognitiv aktivierend eingeschätzt wurde.

Die referierten Befunde zeigen somit, dass anregende Lernumgebungen nicht zwingend zu einer kognitiv vertieften Verarbeitung auf Seiten der Lernenden führen. Insbesondere bei leistungsschwächeren Lernenden scheint dieser Zusammenhang fraglich zu sein (McNamara et al. 1996).

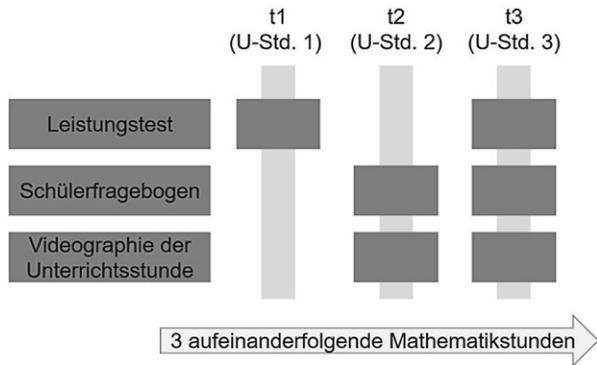
2.4 Fragestellung

Während es mittlerweile eine beträchtliche Anzahl an Studien zur Wirksamkeit kognitiver Aktivierung gibt und auch der lernförderliche Einsatz von Elaborationsstrategien gut belegt ist, sind Studien, die das Gesamtgefüge aus anregendem Unterrichtsangebot, vertiefter kognitiver Nutzung und leistungsförderlicher Wirkung untersuchen, rar (Helm 2016; Hugener 2008; Lipowsky et al. 2009). Angesichts der Ergebnisse schulartspezifischer Wirkungsanalysen zu kognitiv aktivierendem Unterricht (vgl. Abschn. 2.1), differenzieller Effekte des Lernstrategieeinsatzes (vgl. Abschn. 2.2) sowie Befunden dazu, dass Lernende in Abhängigkeit ihres Vorwissens unterschiedlich bezüglich kognitiver Lernaktivitäten auf herausfordernde Lernangebote reagieren (vgl. Abschn. 2.3), scheint es ohne eine weitere empirische Vergewisserung nicht angeraten, bisherige Forschungsbefunde auf den Unterricht bei nicht-gymnasialen Lernenden zu übertragen. Daher wurde im Rahmen der vorliegenden Studie Unterricht dort untersucht und dessen Nutzung und Wirkung analysiert.

Zunächst wird daher der Zusammenhang zwischen der beobachteten kognitiven Aktivierung und dem Ausmaß selbstberichteter Elaborationsstrategien von Schülerinnen und Schülern nicht-gymnasialer Schularten fokussiert (Fragestellung 1). Angesichts der referierten Befunde und theoretischen Überlegungen ist davon auszugehen, dass kognitiv aktivierender Unterricht, da er als kognitiv anregende Lernumgebung prinzipiell auf jedem Schwierigkeitsniveau realisierbar ist, auch bei Lernenden mit vergleichsweise ungünstigen Lernvoraussetzungen zu einem größeren Ausmaß an Elaborationsstrategien führt.

Darüber hinaus ist von Interesse, ob sich die in den beobachteten Unterrichtsstunden erzielten Lernzuwächse der Lernenden mit dem Ausmaß kognitiver Aktivierung im Unterricht und den selbstberichteten Elaborationsstrategien erklären lassen (Fragestellung 2). Hier kann angenommen werden, dass der an anderen Schularten mehrfach als leistungsförderlich belegte kognitiv aktivierende Unterricht auch in der vorliegenden Stichprobe einen positiven Effekt auf die Leistungsentwicklung hat. Ebenso ist davon auszugehen, dass die Elaborationsstrategien prädiktiv für die Leistung sind, sofern sie unmittelbar nach dem Unterricht erfasst und konkret auf die Lernsituation bezogen werden.

Abb. 1 Forschungsdesign:
Datenerhebung je Klasse
(U-Std. Unterrichtsstunde,
 t Messzeitpunkt)



3 Methode

3.1 Stichprobe und Untersuchungsdesign

Datengrundlage für die folgenden Analysen bilden (1) ein Leistungstest, (2) ein Schülerfragebogen sowie (3) Videoanalysen zu Mathematikstunden mit dem Unterrichtsthema „Vermehrter/verminderter Grundwert“. Sämtliche Datenerhebungen wurden unter standardisierten Bedingungen (standardisierte Test- und Fragebogeninstruktion, Kameraskript) durchgeführt (Abb. 1).

Der Leistungstest wurde zeitnah vor der ersten (t_1) sowie nach der zweiten videographierten Unterrichtsstunde (t_3) durchgeführt, wobei Vor- und Nachtest identisch waren. Nach den Unterrichtsstunden (t_2 , t_3) erhielten die Schülerinnen und Schüler einen Fragebogen mit Items zur stundenbezogenen Motivation, zum situationalen Interesse und zu Elaborationsstrategien. Die Lehrkräfte bekamen die Aufforderung, zwei für ihr Unterrichtshandeln typische Stunden zum vorgegebenen Unterrichtsthema zu zeigen.

Die Stichprobe auf Klassenebene umfasste insgesamt 30 Lehrkräfte aus Baden-Württemberg und deren $2 \times 30 = 60$ Unterrichtsstunden. 50% der Lehrkräfte sind weiblich, 23 unterrichteten an einer Haupt- bzw. Werkrealschule, 7 an einer Realschule. Die Stichprobe auf Schülerebene umfasste insgesamt 608 Schülerinnen und Schüler der achten Jahrgangsstufe, wovon 42,1% Mädchen sind. Sowohl auf Klassen- als auch auf Schülerebene stellt die Stichprobe eine Gelegenheitsstichprobe dar.

3.2 Instrumente

3.2.1 Leistungstest

Das konzeptuelle Verständnis des Unterrichtsthemas wurde mithilfe eines schriftlichen Tests erfasst, der von MathematikdidaktikerInnen entwickelt wurde. Der Test beinhaltet Aufgaben mit offenem Antwortformat, die dichotom kodiert wurden und sich auf das konzeptuelle Verständnis des Unterrichtsthemas beziehen (Beispielitem: „Die Preise für das Telefonieren sind in den letzten Jahren auf ein Viertel

gefallen. Drücke diese Veränderung in Prozent aus.“). Anhand einer Pilotierungsstichprobe wurden mittels exploratorischer Faktorenanalyse 12 Items ausgewählt. Das gesamte Testdatenmaterial wurde von zwei Raterinnen unabhängig voneinander kodiert. Die Beurteilerübereinstimmung für die einzelnen Aufgaben erwies sich als gut ($0,79 < \text{Krippendorff's } \alpha < 0,97$). Um die Leistung zu skalieren, wurde ein eindimensionales Rasch-Modell unter Ausschluss eines Items spezifiziert. Anschließend wurden unter Verwendung des Softwarepaketes eRm (Mair und Hatzinger 2007) der Software R (R Core Team 2017) Personenparameter geschätzt. Die Outfit-MSQ-Werte variieren zwischen 0,71 und 1,20 ($IQB = [0,80; 1,04]$), die Infit-MSQ-Werte zwischen 0,82 und 1,03 ($IQB = [0,85; 0,97]$). Zur Überprüfung der internen Konsistenz der endgültigen Testskala ($M = -1,54$; $SD = 1,55$; $IQB = [-3,09; -0,69]$) wurde die Separationsreliabilität berechnet, welche akzeptable Werte zeigt ($r_{\text{WLE_pre}} = 0,64$; $r_{\text{WLE_post}} = 0,69$; $r_{\text{WLE_gesamt}} = 0,68$).

3.2.2 Schülerfragebogen

Die kognitive Lernaktivität der Lernenden wurde über die Skala *Vertiefende Elaboration/organisierende Prozesse* operationalisiert, die in der Studie „Unterrichtsqualität und mathematisches Verständnis in verschiedenen Unterrichtskulturen“ validiert (Rakoczy et al. 2005) und für die vorliegende Studie geringfügig adaptiert wurde. Die Elaborationsstrategien wurden in dieser Studie anhand retrospektiver Selbstauskünfte möglichst zeitnah und auf die Lernsituation bezogen erfasst (Ericsson und Simon 1980).

Für alle Modelle wurden a priori potenzielle Einflussfaktoren als Kontrollvariablen ausgewählt, welche die Zielkriterien beeinflussen und Zusammenhänge verde-

Tab. 1 Skalen im Schülerfragebogen: Beispielitems, deskriptive Statistiken und interne Konsistenz

Skala	Beispielitem	Items	M	SD	McDonald's ω
Vertiefende Elaboration/organisierende Prozesse	In der heutigen Stunde habe ich mir bei der Erklärung des Lehrers ein eigenes Beispiel vorgestellt.	6	2,66 ^a	0,60 ^a	0,75 ^a
			2,66 ^b	0,65 ^b	0,80 ^b
Intrinsische Motivation	In der heutigen Stunde verging die Zeit wie im Flug.	4	2,85 ^a	0,66 ^a	0,77 ^a
			2,87 ^b	0,68 ^b	0,80 ^b
Amotivation	In der heutigen Stunde war ich mit meinen Gedanken woanders.	4	1,67 ^a	0,63 ^a	0,78 ^a
			1,77 ^b	0,68 ^b	0,81 ^b
Situatives Interesse	Das Thema der heutigen Stunde ist sehr nützlich für mich.	7	3,00 ^a	0,61 ^a	0,84 ^a
			2,98 ^b	0,64 ^b	0,87 ^b

Antwortskala für alle Items: Likertskala mit 1 = stimmt gar nicht bis 4 = stimmt genau

^aBezieht sich auf Messzeitpunkt 2

^bBezieht sich auf Messzeitpunkt 3

cken könnten: Da die Leistungsentwicklung in vergleichbaren Studien neben den Elaborationsstrategien auch mit dem situationalen Interesse der Lernenden zusammenhängt (Lipowsky et al. 2009; Rakoczy et al. 2010; Seidel et al. 2002) sowie motivationale Aspekte und Elaborationsstrategien deutlich kovariieren (Baumert 1993; Rakoczy et al. 2007; Seidel 2003), wurden drei diesbezügliche Skalen aus der Studie „Unterrichtsqualität und mathematisches Verständnis in verschiedenen Unterrichtskulturen“ (Rakoczy et al. 2005) aufgenommen und stundenspezifisch adaptiert (Tab. 1).

3.2.3 Videoanalyse

Für einen ganzheitlichen Blick auf die Unterrichtssituation wurden in der vorliegenden Studie durch Videoerhebungen unterstützte Fremdbeobachtungen gewählt (Clausen 2002). Zur Sequenzierung des Unterrichtsverlaufes wurden didaktische Zielentscheidungen berücksichtigt, aus der eine sukzessive Abfolge der Lehr-Lern-Prozess-Struktur resultiert (Klafki 1991). Typischerweise wird Unterricht dabei in aufgabendominierte selbstständige Schülerarbeitsphasen sowie deren Vorbereitung im Plenumsunterricht (Stein et al. 1996) erweitert um Besprechungs- und Reflexionsphasen (Blömeke et al. 2006) eingeteilt, wobei Kleinknecht (2010) davon ausgeht, dass es je Phase unterschiedliche Zielsetzungen gibt, was eine phasenspezifische Unterrichtsanalyse nahelegt. Daher wurden alle Unterrichtsvideos zunächst mit einem niedrig-inferenten Kodierverfahren, das in Anlehnung an Hugener et al. (2006), Kleinknecht (2010) sowie Seidel et al. (2003) ausgearbeitet worden war, in 10-Sekunden-Intervallen bezüglich didaktischer Phasen (Erarbeitung, eigenständige Schülerarbeit, Besprechung) analysiert. Die unabhängige Doppelkodierung wurde durchgehend von einer Projektmitarbeiterin und alternierend von drei Raterinnen, die zuvor mithilfe videographierter Unterrichtsstunden aus anderen Forschungsprojekten und der Pilotstudie im Umgang mit dem Analysemanual geschult wurden, durchgeführt. Die Doppelkodierung erbrachte eine gute Beobachterübereinstimmung (Krippendorff's $\alpha = 0,86$). Voneinander abweichende Urteile wurden in ein Konsensurteil überführt. Auf dieser Grundlage wurden die Videos phasenspezifisch mit einem hoch-inferenten Ratingverfahren hinsichtlich kognitiv aktivierender Merkmale beurteilt. Das dafür genutzte Ratingmanual greift dabei auf die bei Lipowsky et al. (2009) beschriebenen Merkmale zurück. Das Rating wurde auf einer vierstufigen Likertskala vorgenommen, deren Skalenbezeichnung von 1 („trifft nicht zu“) bis 4 („trifft zu“) reicht. Es wurde von vier Raterinnen durchgeführt, von denen zwei Projektmitarbeiterinnen mit mindestens dreijähriger Unterrichtserfahrung und zwei Lehramtsstudierende mit dem Unterrichtsfach Mathematik waren. In einer dreitägigen Schulung wurden grundlegende Begriffe und Konzepte des hoch-inferenten Ratings theoretisch erarbeitet, der Umgang mit dem Analysemanual anhand videographierter Unterrichtsstunden aus anderen Forschungsprojekten und der Pilotstudie trainiert sowie die Übereinstimmung deskriptiv überprüft. Nach dieser Schulung schätzten je Unterrichtsstunde zwei Raterinnen (eine Projektmitarbeiterin und eine Lehramtsstudierende) die kognitive Aktivierung unabhängig voneinander ein. Zur Überprüfung der Interraterreliabilität wurden aufgrund wechselnder Raterinnenpaare unjustierte Intraklassenkorrelationskoeffizienten (ICC; Bliese 2000)

in einfaktoriellen Varianzanalysen berechnet (Wirtz und Caspar 2002). Diese liegen für die ursprünglichen Einschätzungen der Raterinnen zwischen 0,43 und 0,83. Bei Nichtübereinstimmung wurde die Einschätzung mittels Konsensverfahren festgelegt. Für jede didaktische Phase wurde eine separate Skala gebildet und deren faktorielle Struktur separat konfirmatorisch geprüft, da sich die Phasen gegenseitig ausschließen. Aufgrund konzeptioneller Überlegungen wurde jeweils ein ein-dimensionales Modell spezifiziert, dessen Anpassung anhand klassischer Cut-Off-Werte von Fit-Indices überprüft wurde, die nach Hu und Bentler (1999) folgende Grenzen nicht über- bzw. unterschreiten sollen: Comparative Fit Index (CFI) > 0,95; Root Mean Square Error of Approximation (RMSEA) < 0,06 und Standardized Root Mean Square Residual (SRMR) < 0,08. Das Modell für die Erarbeitungsphasen wies zunächst einen Heywood-Fall auf, weswegen die negative, nicht signifikant von Null verschiedene Residualvarianz auf Null fixiert wurde (Chen et al. 2001). Die Abhängigkeiten in den Daten (mehrere Phasen je Stunde, mehrere Stunden je Lehrkraft) wurden mit cluster-robusten Schätzern berücksichtigt (Browne und Arminger 1995). Auf der Grundlage dieser Kriterien ergeben die Analysen der theoretisch angenommenen Modelle sehr gute Modellanpassungen (Kognitive Aktivierung in Erarbeitungsphasen: $\chi^2 = 2,952$; $df = 3$; CFI = 1,000; RMSEA = 0,000; SRMR = 0,054; Kognitive Aktivierung in Schülerarbeitsphasen: $\chi^2 = 10,925$; $df = 9$; CFI = 0,988; RMSEA = 0,051; SRMR = 0,047; Kognitive Aktivierung in Besprechungsphasen: $\chi^2 = 1,712$; $df = 2$; CFI = 1,000; RMSEA = 0,000; SRMR = 0,073). Die Skalen (Erarbeitungsphase: $M = 2,24$; $SD = 0,75$; Schülerarbeitsphasen: $M = 2,20$; $SD = 0,55$; Besprechungsphasen: $M = 2,37$; $SD = 0,67$) ergaben in zwei Fällen gute Werte für die interne Konsistenz (Schülerarbeitsphasen: McDonald's $\omega = 0,72$; Besprechungsphasen: McDonald's $\omega = 0,71$), in einem Fall ist die interne Konsistenz gering (Erarbeitungsphasen: McDonald's $\omega = 0,58$).

3.3 Statistische Analysen

Zur Beantwortung beider Forschungsfragen wurden schrittweise Regressionsanalysen unter Verwendung der Softwarepakete lme4 (Bates et al. 2015) und lavaan (Rosseel 2012) der Software R (R Core Team 2017) durchgeführt, so dass die hierarchische Datenstruktur (mehrere Phasen innerhalb von Unterrichtsstunden) für die Schätzung korrekter Standardfehler anhand von Mehrebenenmodellen oder cluster-robusten Standardfehlern berücksichtigt werden konnte (Rabe-Hesketh und Skrondal 2008). Spezifiziert wurden zur Modellierung von Forschungsfrage 1 jeweils Zwei-Ebenen-Random-Intercept-Modelle (Gelman und Hill 2007). Da die Anwendung von Elaborationsstrategien (abhängige Variable (AV) bei Analysen zu Forschungsfrage 1) nach jeder Unterrichtsstunde, die Nachtleistung (AV bei Analysen zu Forschungsfrage 2) nur einmal erhoben wurde, unterscheiden sich die Cluster auf Ebene 2 für die erste (Level 1: Individuen; Level 2: Unterrichtsstunden) und zweite Fragestellung (Level 1: Individuen; Level 2: Klassen). Alle Daten wurden als manifeste Skalenwerte aufgenommen. Da die identifizierten Unterrichtsphasen in unterschiedlicher Häufigkeit in den Unterrichtsstunden auftraten, wurden die Daten bei mehrfachem Vorliegen phasenspezifisch aggregiert. Zur Kontrolle von Kontexteffekten wurden alle Level-1-Prädiktoren als gruppenaggregierte Variablen auf Ebe-

ne 2 mitberücksichtigt. Um eine Konfundierung durch demografische Merkmale zu minimieren, wurde zudem für das Geschlecht der Schülerinnen und Schüler und die Schulart kontrolliert. Aufgrund der theoretischen Passung und auf Anregung der Gutachtenden wurden für Forschungsfrage 2 zudem Modelle mit Interaktionseffekten aus Vortestleistung und kognitiver Aktivierung sowie Mediationsmodelle (Average Causal Mediation Effects mit cluster-robuster quasi-bayesianischer Approximierung der Konfidenzintervalle; Tingley et al. 2014) geschätzt.

4 Ergebnisse

4.1 Zusammenhang zwischen kognitiver Aktivierung und Elaborationsstrategien

Zur Beantwortung der ersten Fragestellung wurden Regressionsmodelle spezifiziert, in denen zunächst konsekutiv und abschließend simultan die Assoziierung kognitiver Aktivierung in den unterschiedlichen Unterrichtsphasen mit den Elaborationsstrategien geschätzt wurde. Spezifiziert wurden Modelle mit den Daten der Schülerinnen und Schüler auf Ebene 1 sowie den auf die 60 Unterrichtsstunden bezogenen Daten auf Ebene 2. Tab. 2 fasst die unstandardisierten Ergebnisse der Analysen zusammen.

Um den Anteil der zwischen den Unterrichtsstunden liegenden Varianz in den von den Schülerinnen und Schülern berichteten Elaborationsstrategien zu beschreiben, wurde in einem ersten Schritt die Intraklassenkorrelation berechnet. Auch wenn der Großteil der Varianz in der Verwendung von Elaborationsstrategien individuellen Ursprungs ist, bestehen systematische Unterschiede auf Gruppenebene: 9,4 % der Gesamtvarianz liegt zwischen den Unterrichtsstunden.

In den ersten drei Modellen (Modelle 1–3) wurde zunächst jeweils eine phasenspezifische kognitive Aktivierung berücksichtigt. Alle drei untersuchten Gruppenvariablen erweisen sich entgegen der Annahme nicht als signifikante Prädiktoren für die Anwendung von Elaborationsstrategien ($B=0,01$; n.s. bzw. $B=0,02$; n.s.). Dies spiegelt sich auch in der geringen Varianzaufklärung der ersten drei Modelle wider ($0,021 < \text{marginale } R^2 < 0,036$). Da in 18 Unterrichtsstunden keine Besprechungsphase zu beobachten war, konnten in Modell 3 nur 63,6 % der Individualdaten berücksichtigt werden. Um diese hohe Ausfallquote zu umgehen, wurde im Gesamtmodell mit Kontrollvariablen (Modell 4, Tab. 3) die kognitive Aktivierung in den Besprechungsphasen nicht mehr berücksichtigt. Es zeigt sich, dass vor allem die gewählten Kovariaten auf Individualebene (Amotivation: $B=-0,08$; $p=0,028$; intrinsische Motivation: $B=0,39$; $p<0,001$; Interesse: $B=0,14$; $p=0,001$) sowie das mittlere Interesse der jeweiligen Schulklasse auf Gruppenebene ($B=0,34$; $p=0,030$) signifikant zur Vorhersage der Elaborationsstrategien beitragen. Das Gesamtmodell kann 40,1 % der Varianz erklären. Da nicht-signifikante Ergebnisse frequentistischer Inferenzstatistik nicht als Evidenz für die Nichtexistenz eines Effektes gewertet werden können (Dienes 2014), wurde anschließend ein Bayes-Faktor für den Vergleich des Modells 4 mit einem Modell (Modell 4b), das bis auf die Prädiktoren der kognitiven Aktivierung alle Prädiktoren des Modells 4 enthält, geschätzt. Das Ergebnis ($BF=60,87$; basierend auf JZS priors; siehe Rouder und Morey 2012) kann als starke

Tab. 2 Befunde (unstandardisierte Regressionskoeffizienten) aus Mehrebenenanalysen zur Vorhersage der Anwendung von Elaborationsstrategien

	Elaborationsstrategien							
	Einzelmodelle						Gesamtmodell	
	Modell 1		Modell 2		Modell 3		Modell 4	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
(Intercept)	2,76	0,10	2,79	0,12	2,70	0,13	0,79	0,65
Individualebene								
Amotivation	–	–	–	–	–	–	–0,08*	0,04
Intrinsische Motivation	–	–	–	–	–	–	0,39***	0,04
Interesse	–	–	–	–	–	–	0,14**	0,04
Gruppenebene (Stunde)								
Amotivation	–	–	–	–	–	–	0,04	0,16
Intrinsische Motivation	–	–	–	–	–	–	0,21	0,13
Interesse	–	–	–	–	–	–	0,34*	0,15
<i>Kognitive Aktivierung</i>								
Erarbeitung	0,01	0,04	–	–	–	–	0,01	0,02
Lernbegleitung	–	–	0,01	0,06	–	–	0,03	0,03
Besprechung	–	–	–	–	0,02	0,05	–	–
Varianzaufklärung								
R_m^2	0,032		0,036		0,021		0,401	
R_c^2	0,111		0,115		0,101		0,406	
AIC	1860,5		1850,4		1395,0		1325,1	
BIC	1890,1		1880,0		1422,9		1388,6	
<i>N</i> (Individualebene)	1025		1030		773		973	
<i>N</i> (Gruppenebene)	57		57		42		54	

In allen Modellen wird für die Variablen Geschlecht auf Individual- und Schulart auf Stundenebene kontrolliert

B unstandardisierter Regressionskoeffizient, *SE* Standardfehler, R_m^2 marginal R^2 , R_c^2 conditional R^2 (Nakagawa und Schielzeth 2013), *AIC* Akaike Information Criterion, *BIC* Bayesian Information Criterion
* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Auf $p < 0,05$ -Niveau statistisch signifikante Prädiktoren sind fett hervorgehoben

Evidenz für eine bessere Passung der vorliegenden Daten auf die Modell 4b entsprechende Hypothese interpretiert werden und legt somit die Schlussfolgerung nahe, dass ein Effekt kognitiver Aktivierung auf die Elaborationsstrategien der Lernenden nicht existiert.

4.2 Zusammenhang zwischen kognitiver Aktivierung, Elaborationsstrategien und der Leistungsentwicklung der Lernenden

Zur Beantwortung der zweiten Forschungsfrage wurden aufgrund der Anzahl der Prädiktoren und des geringen ICC cluster-robuste Modelle geschätzt, um Singularitäten zu vermeiden. Zunächst (Modell 5) wurden die Nachttestleistungen mit den erhobenen Kovariaten (insbesondere der grand-mean-zentrierten Vortestleistung) und

Tab. 3 Befunde (standardisierte Regressionskoeffizienten) aus cluster-robusten Regressionsmodellen zur Vorhersage der Nachttestleistung

	Abhängige Variable: Nachttestleistung								
	Modell 5			Modell 6			Modell 7		
	β	SE	p	β	SE	p	β	SE	p
Geschlecht	-0,01	0,04	0,76	-0,01	0,04	0,77	-0,01	0,04	0,75
Schulart	-0,01	0,03	0,70	-0,06	0,03	0,07	-0,06	0,03	0,06
Vortestleistung	0,72	0,03	0,00	0,74	0,03	0,00	0,75	0,02	0,00
Amotivation (ES)	0,03	0,04	0,43	0,06	0,05	0,16	0,05	0,04	0,22
Amotivation (ÜS)	0,01	0,04	0,85	0,03	0,05	0,55	0,04	0,05	0,41
Intrinsische Motivation (ES)	-0,04	0,05	0,47	0,00	0,06	0,98	-0,01	0,06	0,90
Intrinsische Motivation (ÜS)	-0,05	0,05	0,31	-0,06	0,05	0,26	-0,06	0,05	0,21
Vertiefende Elaboration (ES)	0,00	0,05	0,97	-0,01	0,05	0,89	-0,01	0,05	0,80
Vertiefende Elaboration (ÜS)	0,05	0,05	0,35	0,03	0,05	0,51	0,04	0,06	0,42
kA Erarbeitung (ES)	-	-	-	0,05	0,03	0,15	0,04	0,03	0,24
kA Erarbeitung (ÜS)	-	-	-	-0,01	0,03	0,65	-0,01	0,03	0,71
kA Begleitung (ES)	-	-	-	0,08	0,05	0,07	0,08	0,05	0,10
kA Begleitung (ÜS)	-	-	-	-0,05	0,03	0,12	-0,05	0,03	0,08
Vortest · kA Erarbeitung (ES)	-	-	-	-	-	-	-0,06	0,03	0,05
Vortest · kA Erarbeitung (ÜS)	-	-	-	-	-	-	0,00	0,03	0,99
Vortest · kA Begleitung (ES)	-	-	-	-	-	-	0,10	0,05	0,05
Vortest · kA Begleitung (ÜS)	-	-	-	-	-	-	-0,07	0,06	0,24
R^2	-	0,523	-	-	0,552	-	-	0,561	-

Auf $p < 0,05$ -Niveau statistisch signifikante Prädiktoren sind fett hervorgehoben

kA kognitive Aktivierung, ES Einführungsstunde, ÜS Übungsstunde

den Elaborationsstrategien prädictiert. Hier zeigte sich erwartungsgemäß ein starker Effekt der Vortestleistung, jedoch erwartungswidrig keine Effekte der Elaborationsstrategien. In einem weiteren Modell (Modell 6) wurden die (grand-mean-zentrierten) Variablen der kognitiven Aktivierung als Prädiktoren aufgenommen, um additive Effekte zu schätzen. Auch diese zeigten insignifikante p -Werte, wenngleich die aufgeklärte Varianz deskriptiv um 3 % anstieg. Diese erhöhte sich erneut um 1 % unter Hinzunahme von Interaktionseffekten aus Vortestleistung und den Variablen der kognitiven Aktivierung (Modell 7), wobei lediglich die Interaktion aus Vortestleistung und kognitiver Aktivierung in der Begleitungsphase der Einführungsstunde signifikant ausfiel. Das positive Vorzeichen dieses Interaktionseffektes

indiziert dabei eine stärkere Assoziation der kognitiven Aktivierung in Begleitphasen der Einführungsstunde mit der Nachtestleistung für Subgruppen mit höheren Vortestwerten. Eine Analyse von Mediationseffekten ergab vernachlässigbar kleine ($|\beta| \leq 0,00$) und nicht signifikante ($p \leq 1$) Ergebnisse.

5 Diskussion

Kognitiv aktivierender Unterricht wird im Hinblick auf seine leistungsförderliche Wirkung vor allem bei Schülerinnen und Schülern mit vergleichsweise günstigen Lernvoraussetzungen untersucht. Befunde zu differenziellen Lern- und Entwicklungsmilieus an den unterschiedlichen Schularten (Baumert et al. 2006) belegen jedoch, dass Unterricht an Haupt- und Realschulen unter anderen Rahmenbedingungen stattfindet als an Gymnasien. Dies verdeutlicht die Notwendigkeit, Unterrichtsqualitätsstudien auch an nicht-gymnasialen Schularten durchzuführen. Dieses Desiderat aufgreifend standen im Zentrum der Analysen die Fragen, ob kognitiv aktivierender Unterricht prädiktiv für die Anwendung von Elaborationsstrategien ist und ob sich beides, sowohl ein anregendes Unterrichtsangebot als auch dessen vertiefte Verarbeitung, als leistungsförderlich erweist.

5.1 Zusammenfassung und Einordnung der Befunde

Unerwartet hinsichtlich der ersten Forschungsfrage ist der Befund, dass kognitive Aktivierung und verwendete Elaborationsstrategien entgegen der Annahme keine signifikanten Assoziationen zeigen. Dies trifft vielmehr auf die selbstberichteten Elaborationsstrategien und die affektiv-motivationale Einstellungen der Lernenden zu. Dies kann als Hinweis darauf verstanden werden, dass kognitiv aktivierender Unterricht nicht direkt zur intendierten Nutzung führt und motivationale Variablen unabhängig davon mit vertieften Elaborationen assoziiert sind.

Der fehlende Zusammenhang zwischen kognitiver Aktivierung und den Elaborationsstrategien weist Parallelen zum bereits berichteten Forschungsstand auf: Hugener (2008) konnte den angenommenen Zusammenhang zwischen kognitiv aktivierendem Unterricht und vertiefend-organisierenden Lernaktivitäten nicht bestätigen und bezog sich interpretierend auf die Befunde von Struyven et al. (2006), denen zufolge schülerorientierter Unterricht vermutlich nur durch die Unterstützung der Lehrkraft zur vertieften Verarbeitung führen kann. Die Ergebnisse von Rakoczy et al. (2010) bekräftigen diese Hypothese insofern, als inhaltlich strukturierende Maßnahmen wie die sprachliche Exaktheit oder die verständliche Darstellung des Erarbeiteten bei den von ihnen untersuchten leistungsschwächeren Schülerinnen und Schülern zu mehr kognitiver Aktivität führen.

Bezüglich der zweiten Forschungsfrage erwies sich kognitive Aktivierung nicht als signifikanter Prädiktor für die Leistungsentwicklung, womit die vorliegenden Daten letztlich inkonklusiv bzgl. Forschungsfrage 2 bleiben.

Die Anwendung von Elaborationsstrategien, retrospektiv erfasst auf der Basis von Selbstaussagen der Schülerinnen und Schüler, scheint offensichtlich kurzfristig betrachtet keine Bedeutung im Angebots-Nutzungs-Wirkungsgefüge zu haben. Im An-

schluss an die internationale Diskussion um desirable difficulties (Bjork 2013), wovon verschiedene evidenzbasierte Maßnahmen subsumiert werden, die das langfristige Behalten und den Wissenstransfer unterstützen, könnte aber auch vermutet werden, dass die vertiefte Auseinandersetzung der Schülerinnen und Schüler mit dem Lerngegenstand eher langfristige Wirkungen hat, etwa wenn Lernende auch mit zeitlichem Abstand vergleichbare Aufgaben noch sicher lösen können.

Lipowsky et al. (2009) und Rakoczy et al. (2010) setzten zwei verschiedene Elaborationsskalen ein: Während sich die nachvollziehende Elaboration als leistungsförderlich erwies (Lipowsky et al. 2009), konnte die Bedeutsamkeit der vertiefenden Elaboration, nach der auch in der vorliegenden Studie gefragt wurde, nicht belegt werden (Rakoczy et al. 2010). Der dort gefundene signifikante Interaktionseffekt zwischen dem Vorwissen und der vertiefenden Elaboration deutet darauf hin, dass kognitive Lernaktivitäten erst dann wirksam werden können, wenn die Lernenden über die notwendigen kognitiven Voraussetzungen verfügen, um neue Lerninhalte elaborierend zu verarbeiten (Rakoczy et al. 2010). Dies könnte ein Grund für den nicht nachweisbaren Zusammenhang zwischen Elaborationsstrategien und Leistung in der hier untersuchten Stichprobe von Lernenden mit eher ungünstigen Lernvoraussetzungen sein; konnte vorliegend jedoch nur teilweise repliziert werden.

Kognitive Aktivierung hat sich nahezu durchgehend in allen Studien als leistungsförderlich erwiesen; somit fügen sich die Ergebnisse der vorliegenden Analysen nur teilweise in den Forschungsstand ein. Zusammenfassend bestätigen die vorliegenden Befunde nur in Interaktion mit dem Vorwissen, dass kognitive Aktivierung auch bei nicht-gymnasialen Schülerinnen und Schülern leistungsförderlich ist.

5.2 Stärken und Grenzen der Studie

Die hier aufgezeigten Ergebnisse müssen auch vor dem Hintergrund des methodischen Vorgehens betrachtet werden. Hierbei stellt sich zunächst die Frage nach der Generalisierbarkeit der gezogenen Schlussfolgerungen, die angesichts der Gelegenheitsstichprobe limitiert ist. Hinzu kommt, dass aus testökonomischen Gründen nur das Geschlecht der Schülerinnen und Schüler erfasst werden konnte. Weitere soziodemographische Daten wären hilfreich gewesen, um die vorliegenden Befunde im Vergleich zu bereits vorliegenden Ergebnissen ähnlicher Studien genauer einordnen zu können. Letztlich kann somit auch nicht belegt werden, ob es sich bei der untersuchten Stichprobe tatsächlich um leistungsschwächere Lernende mit eher ungünstigen Lernvoraussetzungen handelt.

Der nicht zu belegende Zusammenhang zwischen den genutzten Elaborationsstrategien der Schülerinnen und Schüler und dem Leistungszuwachs muss auch in methodischer Hinsicht hinterfragt werden: Baumert (1993), Spörer und Brunstein (2005) sowie Schukajlow und Leiss (2011) fanden wie in der vorliegenden Studie keinen signifikanten Zusammenhang der über Fragebögen ermittelten Tiefenlernstrategien zum jeweiligen Lernerfolgsmaß. Im Widerspruch dazu stehen Interventions- und Interviewstudien, die die leistungsrelevante Bedeutung von vertiefenden Lernstrategien belegen können (Artelt 1999; Spörer und Brunstein 2005). Diese divergenten Ergebnisse führt Artelt (1999) zusammenfassend u. a. auf die sich unterscheidenden Erhebungsmethoden zurück: Studien, in denen die Nutzung von Lern-

strategien wie vorliegend möglichst handlungsnah erfasst wird, zeigen deutlichere Zusammenhänge zum Lernerfolg auf als Untersuchungen, in denen Lernstrategien ohne konkreten Situationsbezug erhoben werden. Auch Wirth und Leutner (2008) diskutieren die Frage nach einer validen Erfassung von Lernstrategien: Neben dem Aspekt der Handlungsnähe heben sie hervor, dass die Fokussierung der Qualität von Lernstrategien, die sich in der Passung der Lernstrategien zur individuellen Lernsituation niederschlägt, von größerer Bedeutung für den Lernzuwachs ist als die Quantität. Insofern wäre es wünschenswert, die Elaborationsstrategien der Schülerinnen und Schüler nicht nur als schriftliche Selbstaussage zu erfassen, sondern beispielsweise durch den Einsatz weiterer Videokameras und Mikrofone, die eine gezieltere Beobachtung der Lernenden erlauben, einen qualitativeren Eindruck von den vertiefenden Elaborationen zu erhalten.

Nicht zu belegende Zusammenhänge zwischen Lernstrategie- und Lernerfolgsmaßen können nach Artelt (1999) auch auf die mangelnde Passung dieser beiden Maße zurückgeführt werden: Wird vor allem reproduzierbares Faktenwissen als Indikator für den Lernerfolg erhoben, lassen sich keine positiven Zusammenhänge zwischen Lernstrategien und erhobenem Leistungsmaß belegen. In der vorliegenden Studie wurden daher Items im Leistungstest eingesetzt, die das operational definierte Merkmal „Konzeptuelles Verständnis des vermehrten und verminderten Grundwertes“ erfassen. Die vergleichsweise geringe Anzahl an Testitems ist dem Umstand geschuldet, sowohl die zeitliche Belastung für die teilnehmenden Schülerinnen und Schüler als auch den Unterrichtsausfall möglichst gering zu halten. Bezüglich der Interpretation der Testergebnisse ist limitierend zu ergänzen, dass der Test keine allgemeinen mathematischen Fähigkeiten erfasst, weswegen die Übertragbarkeit der im Rahmen dieser Studie gewonnenen Erkenntnisse auf den Mathematikunterricht bei leistungsschwächeren Lernenden im Allgemeinen einer weiteren empirischen Überprüfung bedarf.

Des Weiteren soll an dieser Stelle betont werden, dass die Standardvorgehensweise der Mediationsanalyse in diesem Kontext (wie sie auch vorliegend zur Anwendung kam) kritisch gesehen werden kann: Zwar scheinen Mediationsanalysen in manchen sozialwissenschaftlichen Teildisziplinen zur Standardvorgehensweise zu gehören (Bullock et al. 2010). Um diese Modelle (für einen Überblick der diversen Ansätze siehe etwa MacKinnon et al. 2002) zu identifizieren, müssen jedoch selbst bei einer randomisierten unabhängigen Variable starke Annahmen getroffen werden. Etwa dürfen keine unbeobachteten konfundierenden Variablen von Mediator und abhängiger Variable vorliegen und die mediiierende Variable muss messfehlerfrei sein. Beide Annahmen dürften im bildungswissenschaftlichen Kontext so gut wie nie erfüllt sein, was problematisch ist, da die Verletzung dieser Annahmen drastische Verzerrungen zur Folge haben kann (Imai et al. 2010), Sensitivitätsanalysen zur Abschätzung der Robustheit der empirischen Ergebnisse aber kaum zu beobachten sind. Darüberhinaus scheinen bezüglich Mediationsanalysen grundlegende Missverständnisse (Fiedler et al. 2011) ebenso weit verbreitet wie fragwürdige Forschungspraktiken (Götz et al. 2021).

In den vorliegenden Analysen wurde ein phasenspezifischer Zugang gewählt, der dem Umstand Rechnung trägt, dass Lehrkräfte möglicherweise nicht über die gesamte Dauer einer Unterrichtsstunde in gleichem Maße kognitiv aktivierend unterrichten

(Praetorius et al. 2014). Staub (2007) befürwortet in Bezug auf die zu wählende Analyseeinheit, dass sie dieselbe Reichweite hat, in der Lehrkräfte strukturelle Unterrichtsentscheidungen treffen. Mit der Gliederung der Unterrichtsstunden in reliabel zu identifizierende didaktische Phasen scheint dies eingelöst zu sein (Klein-knecht 2010). Die im Rahmen dieser Studie entstandenen Ergebnisse weisen somit eine besondere Nähe zu den von den Lehrkräften in der Unterrichtsvorbereitung zu treffenden didaktischen Entscheidungen auf und erleichtern somit möglicherweise den Transfer der Befunde auf das konkrete Unterrichtshandeln der Lehrkräfte. Daher wird der phasenspezifische Ansatz auch bei weiteren Analysen weiterverfolgt, beispielsweise bei der geplanten Untersuchung des Zusammenhangs zwischen phasenspezifischer kognitiver Aktivierung und dem kognitiven Aktivierungspotenzial der im Unterricht genutzten Lernaufgaben.

Im Zusammenhang mit herausforderndem Unterricht bei leistungsschwächeren Schülerinnen und Schülern wird immer wieder das Ausmaß an notwendiger Führung diskutiert (Kirschner et al. 2006). Bisherige Befunde zu den Voraussetzungen kognitiv aktivierenden Unterrichts belegen, dass die rezeptiven lehr-lerntheoretischen Überzeugungen der Lehrkräfte unabhängig vom Ausmaß beobachteter kognitiver Aktivierung in den Erarbeitungs- und Lernbegleitungsphasen variieren. Somit stimmen manche kognitiv aktivierend unterrichtenden Lehrkräfte einem eher kleinschrittigen Vorgehen zu und realisieren möglicherweise auch strukturierende Instruktionselemente in ihrem Unterricht, die in der vorliegenden Studie nicht erfasst wurden. Daraus und aus den Ergebnissen der aktuellen Analysen ergibt sich die Frage, ob ein anspruchsvolles Unterrichtsangebot in den Plenumsphasen des Unterrichts, das sich gleichermaßen an die ganze Klasse richtet, bei leistungsschwächeren Schülerinnen und Schülern nur leistungsförderliche Wirkung entfalten kann, wenn es gleichzeitig von strukturierenden Maßnahmen flankiert wird (Möller et al. 2002; Decristan et al. 2015). Hier scheint noch erheblicher Forschungsbedarf zu bestehen. Ist das Wechselspiel aus herausforderndem Angebot und strukturierenden Hilfestellungen auch für leistungsschwächere Lernende präziser geklärt, können konkrete Folgerungen für die Unterrichtspraxis abgeleitet werden.

Funding Open Access funding enabled and organized by Projekt DEAL.

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

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

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

Literatur

- Alfieri, L., Brooks, P.J., Aldrich, N.J., & Tenenbaum, H.R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18.
- Artelt, C. (1999). Lernstrategien und Lernerfolg – Eine handlungsnaher Studie. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 31*(2), 86–96.
- Artelt, C., Baumert, J., & Julius-McElvany, N. (2003). Selbstreguliertes Lernen: Motivation und Strategien in den Ländern der Bundesrepublik Deutschland. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000 – ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 131–164). Opladen: Leske + Budrich.
- Bates, D., Mächler, M., Bolker, B.M., & Walker, S.C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
- Baumert, J. (1993). Lernstrategien, motivationale Orientierung und Selbstwirksamkeitsüberzeugungen im Kontext schulischen Lernens. *Unterrichtswissenschaft, 21*(4), 327–354.
- Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III, dritte internationale Mathematik- und Naturwissenschaftsstudie: mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 2, S. 271–315). Opladen: Leske + Budrich.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 314–354). Münster: Waxmann.
- Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen* (S. 95–188). Wiesbaden: VS.
- Bjork, R.A. (2013). Desirable difficulties perspective on learning. In H. Pashler (Hrsg.), *Encyclopedia of the mind* (S. 307–308). Los Angeles: SAGE Reference.
- Bliese, P.D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Hrsg.), *Multilevel theory, research, and methods in organizations* (S. 349–381). San Francisco: Jossey-Bass.
- Blömeke, S., Risse, J., Müller, C., Eichler, D., & Schulz, W. (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht: Ein allgemeines Modell und seine exemplarische Umsetzung im Unterrichtsfach Mathematik. *Unterrichtswissenschaft, 34*(4), 330–357.
- Bobrow, S.A., & Bower, G.H. (1969). Comprehension and recall of sentences. *Journal of Experimental Psychology, 80*(3), 455–461.
- Browne, M.W., & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg & M. E. Sobel (Hrsg.), *Handbook of statistical modeling for the social and behavioral sciences* (S. 311–359). New York: Plenum Press.
- Bullock, J.G., Green, D.P., & Ha, S.E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology, 98*(4), 550–558.
- Chen, F., Bollen, K.A., Paxton, P., & Kirby, J.B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research, 29*(4), 469–508.
- Clausen, M. (2002). *Unterrichtsqualität: eine Frage der Perspektive?* Münster: Waxmann.
- Decristan, J., Hondrich, A.L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., Lühken, A., Adl-Amini, K., Djakovic, S.-K., Mannel, S., & Hardy, I. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *The Journal of Educational Research, 108*(5), 358–370.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 1–17.
- Ericsson, K.A., & Simon, H.A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251.
- Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology, 47*(6), 1231–1236.
- Friedrich, H.F., & Mandl, H. (1992). Lern- und Denkstrategien – ein Problemaufriss. In H. Mandl & H.F. Friedrich (Hrsg.), *Lern- und Denkstrategien: Analyse und Intervention* (S. 3–54). Göttingen: Hogrefe.

- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Götz, M., O'Boyle, E. H., Gonzalez-Mulé, E., Banks, G. C., & Bollmann, S. S. (2021). The "Goldilocks Zone": (Too) many confidence intervals in tests of mediation just exclude zero. *Psychological Bulletin*, 147(1), 95–114.
- Helm, C. (2016). Zentrale Qualitätsdimensionen von Unterricht und ihre Effekte auf Schüleroutcomes im Fach Rechnungswesen. *Zeitschrift für Bildungsforschung*, 6(2), 101–119.
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Hugener, I. (2008). *Inszenierungsmuster im Unterricht und Lernqualität: Sichtstrukturen schweizerischen und deutschen Mathematikunterrichts in ihrer Beziehung zu Schülerwahrnehmung und Lernleistung – eine Videoanalyse*. Münster: Waxmann.
- Hugener, I., Pauli, C., & Reusser, K. (2006). Videoanalysen. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Video-studie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Frankfurt am Main: DIPF.
- Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1), 51–71.
- Kirschner, P.A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.
- Klafki, W. (1991). *Neue Studien zur Bildungstheorie und Didaktik: zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik*. Weinheim: Beltz.
- Kleinknecht, M. (2010). *Aufgabenkultur im Unterricht: eine empirisch-didaktische Video- und Interviewstudie an Hauptschulen*. Baltmannsweiler: Schneider Verl. Hohengehren.
- Klieme, E., & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur“ und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMSS-Impulse für Schule und Unterricht* (S. 43–57). Bonn: Bundesministerium für Bildung u. Forschung.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.
- Kunter, M., & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 9–31). Münster: Waxmann.
- Kunter, M., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A., Klusmann, U., Krauss, S., Löwen, K., Neubrand, M., & Tsai, Y.-M. (2006). Mathematikunterricht in den PISA-Klassen 2004: Rahmenbedingungen, Formen und Lehr-Lernprozesse. In M. Prenzel, et al. (Hrsg.), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 161–194). Münster: Waxmann.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: effects on instructional quality and student development. *The Journal of Educational Psychology*, 105(3), 805–820.
- Lehtinen, E. (1992). Lern- und Bewältigungsstrategien im Unterricht. In H. Mandl & H.F. Friedrich (Hrsg.), *Lern- und Denkstrategien: Analyse und Intervention* (S. 125–149). Göttingen: Hogrefe.
- Leopold, C., & Leutner, D. (2002). Der Einsatz von Lernstrategien in einer konkreten Lernsituation bei Schülern unterschiedlicher Jahrgangsstufen. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 240–258). Weinheim: Beltz.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19(6), 527–537.
- MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83–104.

- Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20(9), 1–20.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43.
- Möller, K., Jonen, A., Hardy, I., & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 176–191). Weinheim: Beltz.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of three basic dimensions. *ZDM*, 50(3), 407–426.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- R Core Team (2017). R: a language and environment for statistical computing. <http://www.r-project.org/>. Zugegriffen: 17. Okt. 2018.
- Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using stata*. College Station: Stata Press.
- Rakoczy, K., Buff, A., & Lipowsky, L. (2005). Befragungsinstrumente. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Frankfurt am Main: DIPF.
- Rakoczy, K., Klieme, E., Drollinger-Vetter, B., Lipowsky, F., Pauli, C., & Reusser, K. (2007). Structure as a quality feature in mathematics instruction: cognitive and motivational effects of a structured organisation of the learning environment vs. a structured presentation of learning content. In M. Prenzel (Hrsg.), *Studies on the educational quality of schools. The final report on the DFG Priority Program* (S. 102–121). Münster: Waxmann.
- Rakoczy, K., Klieme, E., Lipowsky, F., & Drollinger-Vetter, B. (2010). Strukturierung, kognitive Aktivität und Leistungsentwicklung im Mathematikunterricht. *Unterrichtswissenschaft*, 38(3), 229–246.
- Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47(6), 877–903.
- Schukajlow, S., & Leiss, D. (2011). Selbstberichtete Strategienutzung und mathematische Modellierungskompetenz. *Journal für Mathematik-Didaktik*, 32(1), 53–77.
- Seidel, T. (2003). *Lehr-Lernskripts im Unterricht: Freiräume und Einschränkungen für kognitive und motivationale Lernprozesse – eine Videostudie im Physikunterricht*. Münster: Waxmann.
- Seidel, T., Prenzel, M., Duit, R., & Lehrke, M. (Hrsg.). (2003). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“*. Kiel: IPN.
- Seidel, T., Prenzel, M., Reinders, D., Euler, M., Geiser, H., Hoffmann, L., Lehrke, M., Müller, C. T., & Rimmele, R. (2002). „Jetzt bitte alle nach vorne schauen!“ – Lehr-Lernskripts im Physikunterricht und damit verbundene Bedingungen für individuelle Lernprozesse. *Unterrichtswissenschaft*, 30(1), 52–77.
- Seidel, T., Prenzel, M., Rimmele, R., Dalehefte, I. M., Herweg, C., Kobarg, M., & Schwindt, K. (2006). Blicke auf den Physikunterricht. Ergebnisse der IPN Videostudie. *Zeitschrift für Pädagogik*, 52(6), 799–821.
- Spörer, N. (2004). *Strategie und Lernerfolg: Validierung eines Interviews zum selbstgesteuerten Lernen*. Dissertation: Universität Potsdam.
- Spörer, N., & Brunstein, J. C. (2005). Diagnostik von selbstgesteuertem Lernen: Ein Vergleich zwischen Fragebogen- und Interviewverfahren. In C. Artelt & B. Moschner (Hrsg.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (S. 43–63). Münster: Waxmann.
- Staub, F. (2007). Mathematics classroom cultures: methodological and theoretical issues. *International Journal of Educational Research*, 46(5), 319–326.
- Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: an analysis of mathematical tasks used in reform classroom. *American Educational Research Journal*, 33(2), 455–488.

- Struyven, K., Dochy, F., Janssens, S., & Gielen, S. (2006). On the dynamics of students' approaches to learning: the effects of the teaching/learning environment. *Learning and Instruction, 16*(4), 279–294.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(1), 1–38.
- Waldis, M., Grob, U., Pauli, C., & Reusser, K. (2010). Der Einfluss der Unterrichtsgestaltung auf Fachinteresse und Mathematikleistung. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 209–251). Münster: Waxmann.
- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching: a project of the American Educational Research Association* (S. 315–327). New York: Macmillan.
- Wirth, J., & Leutner, D. (2008). Self-regulated learning as a competence: implications of theoretical models for assessment methods. *Zeitschrift für Psychologie, 216*(2), 102–110.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*(1), 51–59.