



# Förderung der adaptiven Erklärkompetenz angehender Chemielehrkräfte durch Reflexion videografiert adaptiver Erklärhandlungen

Stefanie Reimer<sup>1</sup> · Oliver Tepner<sup>1</sup>

Eingegangen: 11. Juni 2024 / Angenommen: 5. März 2025  
© The Author(s) 2025

## Zusammenfassung

Gutes adaptives Erklären wirkt sich positiv auf den Lernzuwachs der Lernenden aus. Bisherige Studien haben die Erklärqualität anhand simulierter Experten-Novizen-Dialoge untersucht. Desiderate bestehen in der Übertragbarkeit bisheriger Erkenntnisse auf unterrichtliche Settings sowie in der Förderung adaptiver Erklärkompetenz im Rahmen des RCM (Refined Consensus Model). Ziel der Arbeit ist die Entwicklung eines universitären Seminars zur Förderung adaptiver Erklär- und Reflexionskompetenz. Da Reflexion sowohl durch eigenes als auch durch beobachtetes Handeln ausgelöst werden kann, wird ergänzend in einem Vergleichsgruppen-Design mit drei Messzeitpunkten untersucht, ob die Beobachtung unterrichtlicher Erklärungen sowie die Anzahl der Zyklen (plan, teach, reflect) einen Einfluss auf den Zuwachs an adaptiver Erklärungscompetenz in Bezug auf Planung und Handlung haben. Dazu planen, halten und reflektieren Studierende Erklärungen und erklären entweder dreimal (*Erklärende-3×*), zweimal (*Erklärende-2×*) oder nie (*Beobachtende*) adaptiv. Signifikante Lernzuwächse lassen sich bei Subfacetten wie Adaptivität (Planung und Handlung) feststellen. Die Ergebnisse zeigen, dass die Reflexion adaptiver Erklärungen die adaptive Erklärkompetenz in der Planung in einzelnen Subkategorien fördern kann. Ebenso gibt es Zusammenhänge zwischen den Facetten des ePCK (plan und reflect).

**Schlüsselwörter** Adaptive Erklärkompetenz · Refined Consensus Model of PCK · Reflexionskompetenz · Videografie

## Promoting Adaptive Explaining Skills of Preservice Chemistry Teachers by Reflecting on Videotaped Explanations

### Abstract

Good adaptive explanations have a positive effect on students' learning. Previous studies have investigated the quality of explanations using simulated expert-novice dialogues. Desiderata exist in the transferability of previous findings to classroom settings and in the promotion of adaptive explanatory competence within the framework of the RCM (Refined Consensus Model). This study aims to develop an university seminar to promote adaptive explanation and reflection skills. As reflection can be triggered both by one's actions and by observed actions, a comparison group design with three measurement points is also used to investigate whether the observation of teaching explanations and the number of cycles (plan, teach, reflect) influence the increase in adaptive explanation skills about planning and action. To this end, students plan, hold, and reflect on explanations and explain adaptively either three times (explainer-3×), twice (explainer-2×), or never (observer). Significant learning gains can be observed in sub-facets such as adaptivity (planning and action). The results show that reflecting on adaptive explanations can promote adaptive explanatory competence in planning in individual subcategories. There are also correlations between the facets of the ePCK (plan and reflect).

**Keywords** Adaptive explanatory skills · Refined Consensus Model of PCK · Reflection skills · Videography

✉ Stefanie Reimer  
stefanie.reimer@ur.de

Oliver Tepner  
oliver.tepner@ur.de

<sup>1</sup> Universität Regensburg, Regensburg, Deutschland

## Einleitung

Ältere Studien zeigen, dass gute Erklärungen von Lehrkräften positiv mit dem Lernerfolg von Lernenden korrelieren (Evans und Guyson 1978), jedoch fällt es angehenden Lehrkräften schwer, verständliche Erklärungen zu formulieren. Insofern ist es erstaunlich, dass der Förderung der Erklärungskompetenz angehender Lehrkräfte bisher wenig Aufmerksamkeit geschenkt wurde (Merzyn 2005) und erst in jüngerer Literatur gefordert wird, diese explizit zum Bestandteil der Lehrkräfteausbildung zu machen (Kulgemeyer und Geelan 2024). Die Forschung hat sich bereits mit dem Thema auseinandergesetzt z. B. durch die Analyse schriftlicher Erklärungen (Guler und Celik 2016) oder die Untersuchung in standardisierten Laborsituationen (Findeisen 2017; Kulgemeyer und Tomczyszyn 2015). Die Erforschung und Förderung adaptiver Erklärungskompetenz im Unterrichtskontext stellen Desiderate dar.

Da die Teilnahme an Unterrichtszyklen (Planen, Unterrichten, Reflektieren) die Reflexion bezüglich der fachlichen Darstellung fördern kann (Zemba-Saul et al. 2000), ist zu erwarten, dass auch die Förderung der Erklärungskompetenz durch wiederholte Erklärsituationen in Kombination mit Reflexion gelingt. Reflexion kann dabei durch selbst gehaltenen oder beobachteten Unterricht ausgelöst werden (Kulgemeyer et al. 2021).

## Theoretischer Hintergrund

### Definition: adaptive Erklärungskompetenz

Pädagogische Erklärungen sollen das Verstehen der Rezipienten fördern (Barbieri et al. 1989). Dieser Prozess ist interaktiv (Kiel 1999), mit asymmetrischer Wissensverteilung (Barbieri et al. 1989) und erfordert eine adaptive Anpassung an Schülerbeiträge (Parsons 2012). Schülerfeedback oder abschließende Anwendungsaufgaben dienen dazu, bestehende Verständnisprobleme zu erkennen (Kulgemeyer 2019) und durch Anpassung von Kontext, Darstellung, Sprache und Sachinhalt zu beheben (Kulgemeyer und Schecker 2013). Die Anpassung an die Lernenden kann in der Planung oder Handlung erfolgen (Beck et al. 2008). Geplante Erklärungen ermöglichen eine sorgfältige und methodische Analyse. Trotz Planung ist eine ausreichende Flexibilität wichtig, um bei Bedarf Anpassungen vornehmen zu können (z. B. Vorbereitung alternativer Erklärungen) (Wagner und Wörn 2011). Erklärungen im naturwissenschaftlichen Unterricht sind als konstruktivistische Interaktion zwischen Lehrkräften und Schülerinnen und Schülern zu verstehen (Kulgemeyer und Geelan 2024; Treagust und Harrison 1999).

Vor diesem Hintergrund wird *adaptives* Erklären in dieser Forschungsarbeit ebenfalls konstruktivistisch interpretiert und wie folgt definiert: „Unterrichtliches Erklären einer Lehrkraft ist ein komplexer, vorbereiteter oder sich situativ ergebender sowie interaktiver Kommunikationsprozess, der auf eine Fähig- bzw. Fertigkeitsvermittlung sowie die Initiierung eines Verstehensprozesses bei Adressatinnen und Adressaten abzielt.“ (Lindl et al. 2019, S. 131).

### Messung und Bewertung der adaptiven Erklärungskompetenz

Für die Messung der Erklärungskompetenz werden verschiedene Ansätze genutzt. Einerseits kann Erklärungswissen als Teilfacette des fachdidaktischen Wissens erfasst werden (Kunter und Voss 2011) oder separat gemessen werden, z. B. durch die Bewertung schriftlicher (Guler und Celik 2016) oder mündlicher Befragungen (Wheeldon 2011). Darüber hinaus können sowohl schriftliche Vorbereitungen als auch videografierte Erklärungen (Sevian und Gonsalves 2008) oder Erklärungen in kontrollierten und standardisierten Laborsituationen (Findeisen 2017) untersucht werden. Auch die Bewertung einer Erklärung durch die Adressaten ist möglich (Odora 2014).

Zur Bewertung der adaptiven Erklärungsqualität werden empirisch entwickelte Kriterien herangezogen. Kulgemeyer (2019) nennt neun Faktoren einer guten adaptiven Erklärung: Struktur, Anpassung an die Adressatinnen und Adressaten, Beschränkung auf das Nötigste, reparierende Erklärungen bei Missverständnissen oder Fehlkonzepten, weitere Lernaufgaben, Fokussierung zentraler Konzepte, Hervorheben der Relevanz, Konzentration auf Grundsätze und Anwendung möglicher Instrumente zur Anpassung an das Vorwissen. Beispiele und aktive Einbindung der Lernenden sind für die Adressatenorientierung wichtig (Kulgemeyer und Schecker 2013). Da Lernende mit einem vorunterrichtlichen Verständnis in den Unterricht kommen, besteht die größte Herausforderung beim Erklären darin, diese Modelle zu modifizieren oder zu verändern (Sevian und Gonsalves 2008). Um Verständnis bei Lernenden zu erreichen, sollte eine Lehrkraft die Probleme der Lernenden erkennen und beheben (Kulgemeyer 2019). Dabei können auch Visualisierungen eine wesentliche Rolle spielen (Schopf und Zwischenbrugger 2015), um mentale Bilder zu entwickeln (Sevian und Gonsalves 2008). Darüber hinaus zeichnet sich eine verständliche Sprache durch Begriffsbewusstheit und die Verwendung von Aufmerksamkeitsmarkern aus (Lindl et al. 2019). Bezogen auf das Fach Chemie spielt die Trennung von Stoff- und Teilchenebene beim Erklären chemischer Phänomene eine besondere Rolle (Sevian und Gonsalves 2008).

## Befunde zur Erklärkompetenz

Gute Lehrkräfteerklärungen wirken sich positiv auf den Lernerfolg aus (Evans und Guyson 1978). In Studien zur Erklärkompetenz von angehenden Lehrkräften zeigen sich Probleme bei der Kohärenz einer Erklärung (Borko et al. 1992). Auch gelingen die Einschätzung von Lernvoraussetzungen und typischen Lernschwierigkeiten (Inoue 2009) sowie das adaptive Reagieren darauf nur unzureichend (Borko und Livingston 1989). Andere Studien befassen sich mit Experten-Novizen-Unterschieden, wobei Novizen eine geringere adaptive Erklärkompetenz aufweisen (Findeisen 2017) und weniger Zugänge zu einem Thema haben als Experten (Housner und Griffey 1985). Novizen sind nicht in der Lage, spontan auf Fragen der Lernenden zu reagieren (Borko und Livingston 1989) oder ihre Erklärungen flexibel anzupassen. Ebenso wurde versucht, die Erklärkompetenz eines Lehrerpaares durch Unterrichtszyklen (Planung, Unterricht, Reflexion) zu fördern. Es zeigt sich, dass die Studierenden ihren Plänen folgen und ungeplante, spontane Erklärungen oft fehlerhaft sind (Leinhardt und Greeno 1986). Erfahrene Lehrkräfte sind besser in der Lage, Unterrichtsprobleme zu antizipieren und verfügen über eine größere Anzahl von Lösungsmöglichkeiten (Housner und Griffey 1985). Miltz (1972) zeigt, dass gezieltes Training signifikante Verbesserungen in der Erklärkompetenz bewirken kann und gibt so Hinweise, auf die Erlernbarkeit der Erklärkompetenz. Diese korreliert insbesondere mit dem PCK und CK von Lehramtsstudierenden. Dabei scheint der Einfluss des CK auf die Erklärleistung vom PCK mediert zu werden, da CK nur dann einen positiven Einfluss hat, wenn auch das PCK entsprechend steigt (Kulgemeyer und Riese 2018).

## Adaptives Erklären und Reflektieren als Merkmale fachdidaktischen Wissens

Erklärungswissen ist Teil des fachdidaktischen Wissens (pedagogical content knowledge, PCK) von Lehrkräften und umfasst alternative Formulierungen von Sachverhalten sowie ein reichhaltiges Repertoire an Repräsentationen (Shulman 1987). Im Rahmen des Refined Consensus Model of PCK (RCM) wird letzteres in drei Bereiche untergliedert: collective PCK (cPCK), personal PCK (pPCK) und enacted PCK (ePCK). Das cPCK bezieht sich auf das kollektive Wissen in der Lehrkräftebildung und -forschung, das pPCK ist das Wissen einer einzelnen Lehrkraft und das ePCK beschreibt das verfügbare und genutzte Wissen in einer bestimmten Anwendungssituation (Carlson et al. 2019). Für die Entwicklung fachdidaktischen Wissens spielt die wechselseitige Transformation von pPCK und ePCK eine besondere Rolle. Diese kann durch die intendierte Reflexion des Unterrichtshandelns erfolgen, wobei die gezogenen Konse-

quenzen als Teil des pPCK gespeichert werden. Dies wirkt sich auf zukünftige Planungsprozesse weiterer Unterrichtsstunden sowie Erklärungen aus, in denen auf die neu gewonnenen Kompetenzen zurückgegriffen wird (Alonzo und Nilsson 2019). Das ePCK umfasst Planen, Unterrichten und Reflektieren als zyklische Elemente, die als Grundlage für Trainingsprozesse in der Lehrkräfteaus- und -fortbildung dienen können (Carlson et al. 2019). Die Unterrichtsplanung gilt als Voraussetzung für erfolgreichen Unterricht (Blömeke et al. 2008), wobei Lehrkräfte geeignete Aktivitäten für nachfolgende Unterrichtssituationen vorbereiten. Unter Unterrichten wird der Prozess verstanden, in welchem (angehende) Lehrkräfte u. a. auf Basis ihres Wissens und ihrer Einstellungen Entscheidungen treffen und Wissenserwerb aktiv gestalten. Sie reagieren auf ihre Schülerinnen und Schüler z. B. mit Feedback, Erklärungen und Fragen, ohne die Gründe für diese Entscheidungen zu formulieren (Alonzo et al. 2019). Zusammenhänge zwischen der Qualität von Unterrichtsplänen und deren Umsetzung durch angehende Lehrkräfte wurden nachgewiesen (Kademian und Davis 2018).

Reflexion ist in der Lehrkräftebildung sowohl für die Professionalisierung (Baumert und Kunter 2006) als auch für die Erweiterung des Strategierepertoires von Schülerinnen und Schülern wichtig. Gut angeleitetes „self-assessment“ verbessert die Leistungen insbesondere leistungsschwacher Lernenden (Zulliger et al. 2022). Das Verständnis von Reflexion variiert jedoch (von Aufschnaiter et al. 2019). Im vorliegenden Beitrag wird Reflexion als zielgerichtetes Nachdenken über Handlungen verstanden, aus dem begründete Konsequenzen abgeleitet und umgesetzt werden (Wyss 2013). Reflexion wird als zyklischer Prozess konzeptualisiert: (1) Handlung, (2) Rückblick auf die Handlung, (3) Bewusstwerden wesentlicher Aspekte, (4) Formulierung alternativer Handlungen, (5) Ausführung der alternativen Handlung (Korthagen und Vasalos 2005). Da sich die Reflexionskompetenz (schriftlicher Reflexionen) von Lehramtsstudierenden signifikant zwischen dem zweiten und dritten Reflexionszyklus (Planen, Unterrichten und Reflektieren) verbessert (Kobl 2021), wird die Anzahl der Unterrichtszyklen in dieser Studie beforscht. Nicht nur aktives Handeln, sondern auch das Beobachten einer Handlung kann eine Reflexion auslösen (Kulgemeyer et al. 2021), sodass sich die im Folgenden dargestellten Ziele und Forschungsfragen ergeben.

## Ziele und Forschungsfragen

Zur Förderung der adaptiven Erklärkompetenz von Lehramtsstudierenden im Fach Chemie ist ein universitäres Seminar entwickelt worden. Entsprechend dem ePCK (Planen, Unterrichten, Reflektieren) des RCM (Carlson et al.

2019) kann die adaptive Erklärkompetenz durch Reflexion unterstützt werden. Die Ebenen Planen und Unterrichten des RCM sollen auf das adaptive Erklären nach Beck et al. (2008) angewendet werden. Da sowohl aktives Handeln als auch das Beobachten einer Handlung Reflexion auslösen können (Kulgemeyer et al. 2021), werden in dieser Studie beide Gruppen miteinander verglichen. Dabei wird zudem das Fachwissen berücksichtigt, weil es mit ePCK, pPCK und cPCK interagiert (Carlson et al. 2019). Untersucht wird, ob das Halten eigener oder das Beobachten fremder Unterrichtserklärungen zu Unterschieden in der Entwicklung von Wissen (F1.1), Planungskompetenz (F1.2) und Handlungskompetenz (F1.3) von angehenden Chemielehrkräften in Bezug auf adaptives Erklären führt. Da sich die Reflexionskompetenz bei Lehramtsstudierenden erst zwischen dem zweiten und dritten Messzeitpunkt signifikant zu verbessern scheint (Kobl 2021) und diese die Erklärkompetenz beeinflusst (Zemba-Saul et al. 2000), wird die Anzahl der Unterrichtszyklen untersucht. Weil auch ein Zusammenhang zwischen der Reflexionskompetenz und dem pädagogischen Wissen sowie dem Fachwissen einer Lehrperson nachgewiesen werden konnte (Kulgemeyer et al. 2021), wird darüber hinaus beforscht, ob es Zusammenhänge zwischen der Reflexionsfähigkeit und dem Planen adaptiver Erklärungen (F2a) bzw. zwischen der Reflexionsfähigkeit und dem Halten adaptiver Erklärungen (F2b) gibt. Ergänzend sollen der Kompetenzzuwachs der Studierenden am Ende des Seminars hinsichtlich des Schüler-Wissenszuwachses (F3a) und der wahrgenommenen Erklärqualität (F3b) auf Schülerebene belegt werden.

Die Ziele und Forschungsfragen dieser Studie lauten entsprechend:

#### **F1.1: Lassen sich Gruppenunterschiede im Wissenszuwachs im adaptiven Erklären nachweisen?**

- H1.1a: Der Zuwachs des Wissens zum adaptiven Erklären von Studierenden, die drei Mal adaptiv handeln, unterscheidet sich nicht von Studierenden, die zwei Mal adaptiv handeln.
- H1.1b: Der Zuwachs des Wissens zum adaptiven Erklären von Studierenden, die drei Mal adaptiv handeln, unterscheidet sich nicht von Studierenden, die drei Mal adaptives Erklären beobachten.

Diese Forschungsfrage und Hypothesen beziehen sich auf den Wissenstest zum adaptiven Erklären, der in einem Pre-Post-Follow-up-Design (Multiple-Choice-Single-Select) durchgeführt wurde. Die Intervention findet zwischen der Pre- und Post-Testung statt und ist für alle Gruppen gleich, weshalb an dieser Stelle keine Unterschiede zwischen den Gruppen zu erwarten sind.

#### **F1.2: Lassen sich Gruppenunterschiede in der Entwicklung der Kompetenz zum adaptiven Erklären in der Planung nachweisen?**

- H1.2a: Studierende, die drei Mal adaptiv handeln, haben einen größeren Zuwachs an *adaptiver Erklärungskompetenz in der Planung* als Studierende, die zwei Mal adaptiv handeln.
- H1.2b: Die Entwicklung der *adaptiven Erklärkompetenz in der Planung* bei Studierenden, die drei Mal adaptiv handeln, unterscheidet sich nicht von Studierenden, die drei Mal adaptives Handeln beobachten.

Diese Forschungsfrage und Hypothesen beziehen sich auf die vor jedem Erklärtag schriftlich ausformulierten adaptiven Erklärungen (Planungen).

#### **F1.3: Kann die adaptive Erklärungskompetenz in der Handlung bei Studierenden, die drei Mal adaptiv handeln, stärker gefördert werden als bei Studierenden, die zwei Mal adaptiv handeln?**

- H1.3a.: Studierende, die drei Mal adaptiv handeln, haben einen größeren Zuwachs an *adaptiver Erklärungskompetenz in der Handlung* als Studierende, die zwei Mal adaptiv handeln.

Diese Forschungsfrage und Hypothese beziehen sich auf die adaptiven Erklärungen (Handlungen), die einige Studierende an den Erklärtagen adaptiv durchgeführt haben.

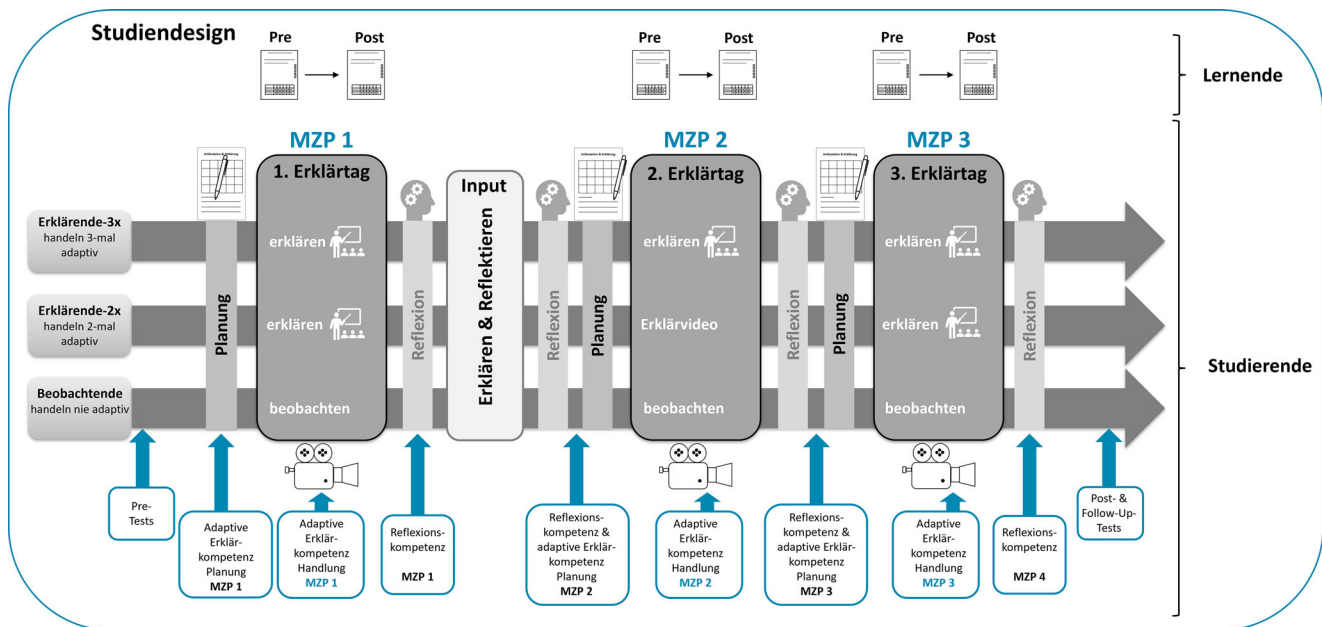
#### **F2: Welche Zusammenhänge bestehen zwischen Planen, Unterrichten und Reflektieren adaptiver Erklärungen?**

- H2a: Es besteht ein positiver Zusammenhang zwischen der Reflexionskompetenz der Studierenden und ihrem Planen adaptiver Erklärungen.
- H2b: Es besteht ein positiver Zusammenhang zwischen der Reflexionskompetenz der Studierenden und ihrem Halten adaptiver Erklärungen.
- H2c: Es besteht ein positiver Zusammenhang zwischen der adaptiven Erklärkompetenz in der Planung und Handlung.

Diese Forschungsfrage und Hypothesen beziehen sich auf die im RCM vorgesehene Aufteilung von ePCK in Planen, Lehren und Reflektieren sowie auf mögliche Verbindungen zwischen diesen Bereichen.

#### **F3: Lässt sich der Kompetenzzuwachs der Studierenden auf Schülerebene nachweisen?**

- H3a: Der Wissenszuwachs der Schülerinnen und Schüler ist am dritten Erklärtag größer als am ersten Schülertag



**Abb. 1** Das Studiendesign wird im Rahmen eines Universitätsseminars mit Chemielehramtsstudierenden angewandt, die in drei Gruppen aufgeteilt werden: *Erklärende-3x*, *Erklärende-2x* und *Beobachtende*

und in der Gruppe der Erklärenden-3x größer als in der Gruppe der Erklärenden-2x.

- H3b: Die Schülerinnen und Schüler bewerten die adaptiven Erklärungen des dritten Erklärtags besser als die des ersten Erklärtags und die der Erklärenden-3x besser als die der Erklärenden-2x.

Die Forschungsfrage und Hypothesen beziehen sich auf den Einfluss der Erklärerfahrungen der Studierenden auf das Verständnis der Rezipienten (Odora 2014) und auf die Bewertung der Adressaten (Barbieri et al. 1989).

## Methoden

### Untersuchungsdesign und Stichprobe

#### Design

Das Seminar folgte einem dreistufigen Zyklus, basierend auf dem ALACT-Modell (Korthagen und Vasalos 2005) und Kobl (2021).

Das Studiendesign (Abb. 1) umfasst drei Interventionsgruppen, wobei für die vorliegenden Fragestellungen von einer Interventionsgruppe und zwei Kontrollgruppen gesprochen werden kann: *Erklärende-3x* (IG, drei adaptive Erklärungen), *Erklärende-2x* (KG1, zwei adaptive Erklärungen) und *Beobachtende* (KG2, keine adaptive Handlung).

Alle Studierenden absolvieren vor Seminarbeginn einen Wissenstest (Multiple-Choice-Single-Select, Pre) zum adaptiven Erklären. Nach der Intervention und den drei Reflexionszyklen folgt der Wissenstest erneut am Seminarende (Post) und 12 Wochen nach dem Seminar (Follow-up).

Nach dem Pre-Test bilden jeweils ein Erklärender-3x und ein Beobachtender ein Tandem und erhalten dasselbe Thema. Jede Person plant eigenständig eine 20-minütige Erkläreinheit zum Thema „Chemische Reaktion und Energetik“. *Erklärende-2x* und *Erklärende-3x* führen diese mit Schülerinnen und Schülern einer 8./9. Klasse am ersten Erklärtag durch, während sie videografiert und anschließend schriftlich reflektiert wird. Die *Beobachtenden* führen die geplante Erklärung nicht selbst durch, sondern beobachten die Erklärung des Tandempartners und reflektieren dessen Unterrichtsvideo (MZIP 1). Eine Intervention zur Förderung der Reflexion und des adaptiven Erklärens findet über vier Seminareinheiten statt. Theoriebasiert reflektieren alle Studierenden den ersten Erklärtag erneut und überarbeiten die Erklärungen auf dieser Basis (MZIP 2).

Am zweiten Erklärtag verfahren *Erklärenden-3x* und die *Beobachtenden* wie zuvor. Die *Erklärenden-2x* reflektieren und überarbeiten ihre Erkläreinheit, setzen diese aber nicht adaptiv mit den Lernenden um, sondern videografieren ihre überarbeitete Erkläreinheit in einem Erklärvideo. So bleibt Time-on-Task der *Erklärenden-2x* und -3x gleich. Am dritten Erklärtag wiederholt sich der Ablauf des ersten. Damit kombiniert die Studie ein Vergleichsgruppen- mit einem Pre-Post-Follow-up-Testdesign (Wissenstest).



Die Schülerinnen und Schüler absolvieren zu Beginn einen Wissenstest über alle Lerninhalte und nach jeder Erkläreinheit einen themenspezifischen Wissenstest. Zusätzlich bewerten sie jede Erkläreinheit mit Hilfe eines Bewertungsbogens.

### Stichprobe

Die Haupterhebung erstreckte sich über drei Semester und umfasste 59 Lehramtsstudierende der Chemie (viertes–zehntes Semester): 22 *Erklärende-3x*, 18 *Erklärende-2x* und 20 *Beobachtende*.

Insgesamt nahmen 1858 Lernende (Realschule und Gymnasium) an der Studie teil: 591 Schülerinnen und Schüler wurden von *Erklärenden-2x* und 1267 von *Erklärenden-3x* unterrichtet. Alle Lernenden (8./9. Klasse) hatten das Thema „Chemische Reaktion und Energetik“ bereits im Vorhinein/Vorjahr gelernt (Tab. 1).

Sowohl die Heterogenität bei den Studierenden (Wissen zum Erklären, adaptives Erklären in Planung und Handlung sowie Reflexion) als auch bei den Schülerinnen und Schülern (Wissen) wurde berücksichtigt, indem das Vorwissen jeweils als Kovariate bzw. über den residualen Lernzuwachs in die Berechnungen einbezogen wurde. Zudem hatten die Studierenden einen substanziellen Teil ihrer fachlichen und fachdidaktischen Ausbildung erfolgreich absolvieren müssen, um am Seminar teilnehmen zu können.

## Instrumente

### 1. Wissenstest zum adaptiven Erklären

Für das adaptive Erklärungswissen wurde ein 21 Item umfassender Multiple-Choice-Single-Select-Test (fünf Antwortmöglichkeiten, Cronbach's  $\alpha=0,70$ ) entwickelt, der sowohl deklaratives als auch prozedurales Wissen erfasst. Die deklarativen Fragen behandeln das im Seminar vermittelte Inhaltswissen zum adaptiven Erklären basierend auf verschiedenen Autoren (u. a. Findeisen 2017; Kulgemeyer 2019; Lindl et al. 2019). Bei den prozeduralen Items wählen die Studierenden die beste Erklärung aus fünf Optionen, die aus Lehrbüchern und seriösen Internetquellen stammen und an die Erklärkriterien angepasst wurden. Der Test misst die systematische Anwendung der Erklärungskriterien (z. B. sie sind unvollständig oder fördern falsche Vorstellungen, Beispielitems siehe Anhang A). Somit kann der entwickelte Wissenstest eine gewisse Handlungsqualität in Bezug auf das Anwenden der Qualitätskriterien voraussagen, jedoch nur in planbaren Situationen (Planungen). Die Anpassungen der Erklärungen an die Studierenden (Handlung) kann mit dem Wissenstest gemessen werden. Hierbei wurde eine ähnliche Vorgehensweise wie beim ersten Teil des Testinstruments der Studie von Bartels und Kulgemeyer (2019)

verwendet. Zur genauen Definition der richtigen Antworten wurde ein Konsensverfahren mit vier Forschenden aus dem Bereich der naturwissenschaftlichen Bildung durchgeführt (analog Bartels und Kulgemeyer 2019), um die inhaltliche Validität der Items zu erreichen.

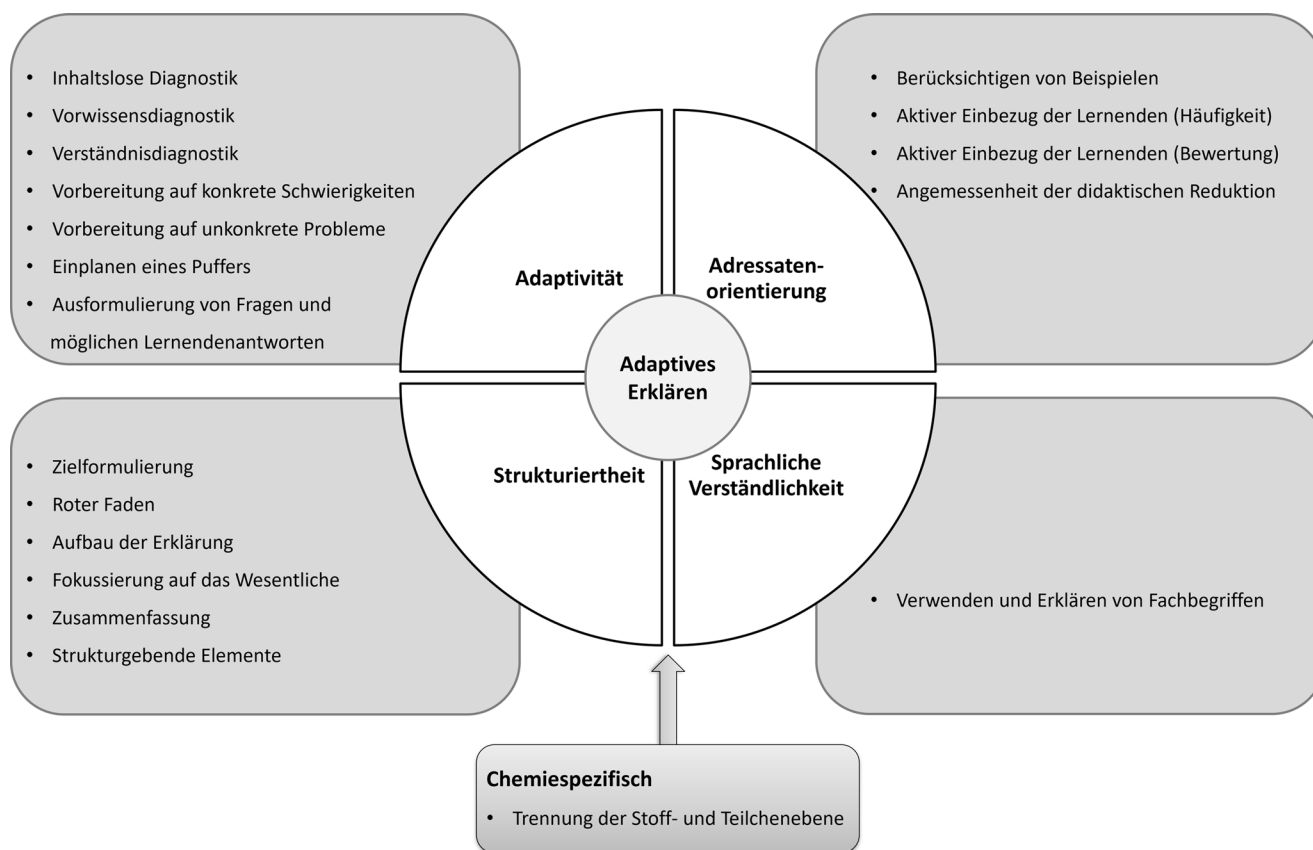
### 2. Adaptives Erklären – Planung

Zunächst wurden Kategorien für gute adaptive Erklärungen deduktiv entwickelt<sup>1</sup> (Findeisen 2017; Kulgemeyer 2019; Lindl et al. 2019). Im interdisziplinären Projekt FALKE-q wurden literaturbasierte Qualitätskriterien für das Erklären (Strukturiertheit, Adressatenorientierung, Sprech- und Körperausdruck, sprachliche Verständlichkeit, fachspezifische Aspekte) zusammengestellt und Erklärvideos aus verschiedenen Perspektiven bewertet, um diese Kriterien zu validieren (Lindl et al. 2019). Die Kategorien und Items des Projekts dienten als Grundlage für das Kodierhandbuch. Da die Adaptivität von Erklärungen fehlte, wurden zusätzliche Items entwickelt (Findeisen 2017; Kulgemeyer 2019) und die Ratingskalen induktiv erweitert. Das Kodierhandbuch umfasst final die folgenden Konstrukte: Strukturiertheit (6 Items), Adressatenorientierung (4 Items), Adaptivität (7 Items), sprachliche Verständlichkeit (31 Item) und chemiespezifische Aspekte (1 Item). Abb. 2 gibt eine Übersicht über die Items.

Jedes Item hat eine Ratingskala von 0 bis 3 bzw. 0 bis 2, analog zum Kodiermanual von Findeisen (2017) (Beispielitem Anhang B). Zur Bewertung der Planung wurde ein Score-Wert gebildet. Bei einigen Adaptivitäts-Items gab es keinen Maximalwert, da absolute Häufigkeiten erfasst wurden. Die höchste Punktzahl des besten Studierenden wurde als Maximum verwendet. Die prozentuale Übereinstimmung wurde als Maß für die Reliabilität aller Kodiermanuals verwendet (keine Normalverteilung, drei Rater: 89 %, ICC=0,99).

Drei chemiespezifische Kriterien wurden vorab identifiziert: fachliche Korrektheit (einschließlich Fachsprache), Schülervorstellungen und Fehlvorstellungen sowie das Stoff-Teilchen-Konzept. Letztlich wurde nur das Stoff-Teilchen-Konzept in die Bewertung aufgenommen. Fachliche Korrektheit blieb als grundlegende Voraussetzung in anderen Kategorien bestehen und wurde bei der Punktevergabe dort zum Teil (z. B. bei Roter Faden und Stoff-Teilchen-Konzept) berücksichtigt. Um eine doppelte Wertung zu vermeiden, wurde die Kategorie nicht gesondert gewertet. Zudem gibt es unterschiedliche Auffassungen darüber, was in spezifischen Kontexten als fachlich korrekt oder angemessen gilt. Während Schülervorstellungen und

<sup>1</sup> Dies geschah teilweise gemeinsam mit weiteren Fachdisziplinen im Rahmen eines Domänen übergreifenden Projekts. (Informationen werden nach Annahme des Manuskriptes ergänzt).



**Abb. 2** Überblick über die Subfacetten und Items zum adaptiven Erklären in der Planung

Fehlvorstellungen zunächst kodiert worden sind, wurde sich gegen ihre Berücksichtigung bei der Bepunktung entschieden, weil einzelne ungenaue Formulierungen keine validen Rückschlüsse auf dahinterliegendes Wissen erlauben.

### 3. Adaptives Erklären – Handlung

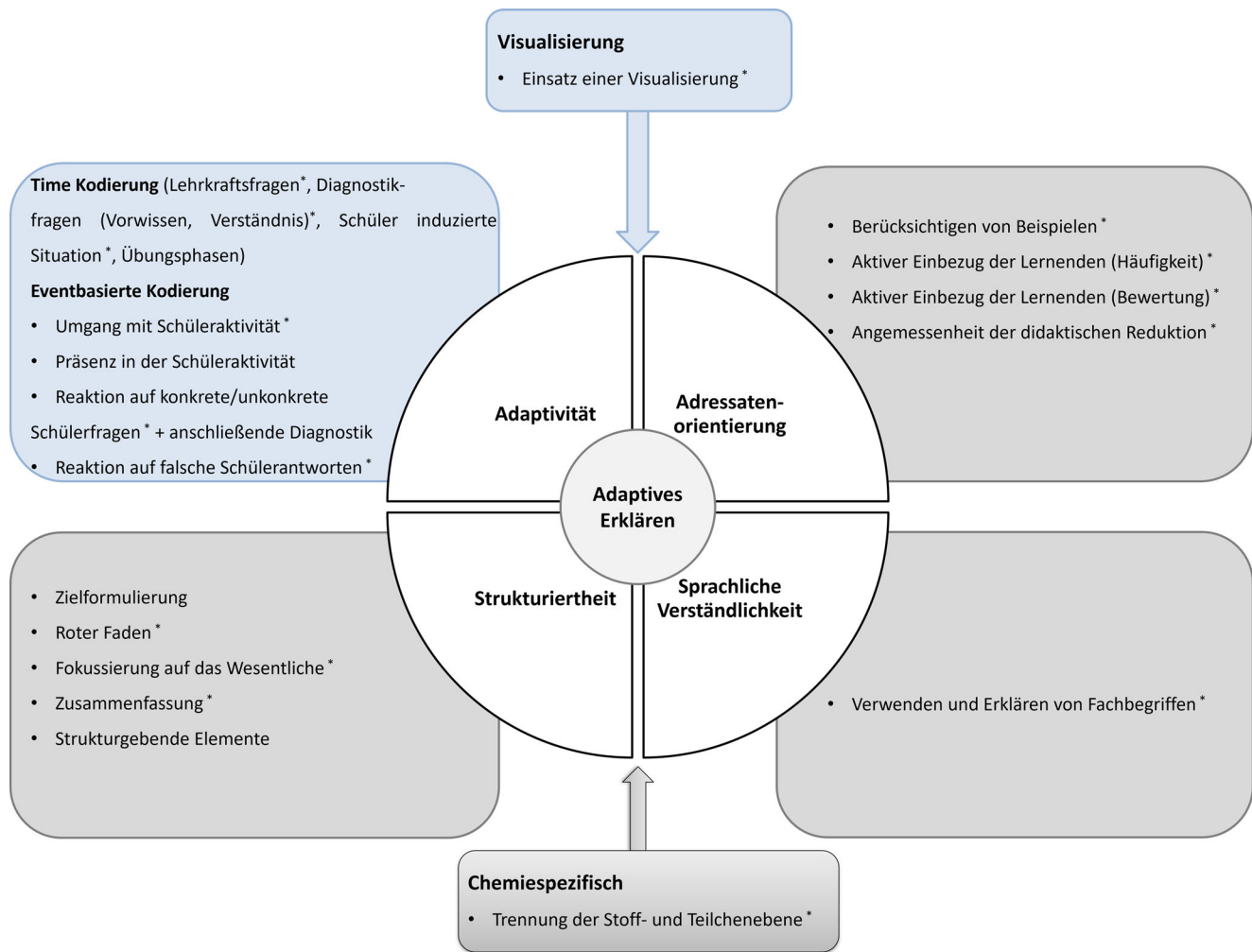
Das Kodiermanual für die Planungen wurde für die Videos adaptiert (Abb. 3). Ein Item zur Visualisierung wurde hinzugefügt und die Items zur Adaptivität wurden angepasst und in zwei Schritten kodiert. Im ersten Schritt erfolgte die zeitbasierte Kodierung in 15-Sekunden-Intervallen zur Interaktion der Lehrperson mit den Lernenden (Fragen, Übungen oder Diagnosen). Im zweiten Schritt wurde ereignisbasiert kodiert, wie die Lehrperson auf Fragen oder falsche Aussagen der Lernenden reagiert und wie sie die Übungsphase gestaltet (Anhang B). Die prozentuale Übereinstimmung aller Items betrug 70 % (zwei Rater, ICC=0,86), und die Zeitkodierung ergab eine Übereinstimmung von 80 % (Cohen's Kappa=0,616).

Auch bei den Unterrichtsvideos wurde ein Score-Wert gebildet. Zur Validierung der Kodiermanuale (Planung, Handlung) bewerteten zehn Lehrkräfte im Rahmen einer Fortbildung jeweils ein Video vom ersten und dritten Er-

klärtag. Sie notierten positive und negative Aspekte, die den Kategorien zugeordnet werden. Unklare Aussagen wurden nicht berücksichtigt. Die meisten Kategorien wurden validiert (Abb. 3, mit Stern markiert). In acht von zehn Fällen wurde das Video des dritten Schülertags besser bewertet als vom ersten (Anzahl positiver und negativer Aspekte). Bei zwei Videos führte die persönliche Gewichtung der Fehler/Monita, welche auch im dritten Video zu finden waren, zu einer schlechteren Einschätzung.

### 4. Schriftliche Reflexionen

Das Kodiermanual unterteilt die Reflexionskompetenz in Reflexionsbreite (inhaltliche Aspekte) und Reflexionstiefe (Kobl 2021). Die Reflexionsbreite beschreibt die Kategorien der Kodiermanuale zum adaptiven Erklären (1–6) und wurde um zusätzliche Aspekte der Unterrichtsqualität erweitert, die von den Studierenden benannt wurden. Jedes Breiten-Item wurde in zwei weiteren Durchgängen zur Tiefe kodiert. Letztere gliedert sich in fünf Überkategorien (Abb. 4, modifiziert nach Kobl 2021).



**Abb. 3** Konzeptualisierung und Validierung adaptiver Erklärkompetenz in der Handlung. Übereinstimmungen von theoretisch abgeleiteten Kategorien und von Experten genannten Kategorien sind durch Sternchen kenntlich gemacht

## 5. Lernenden-Wissenstest und Bewertungsbogen

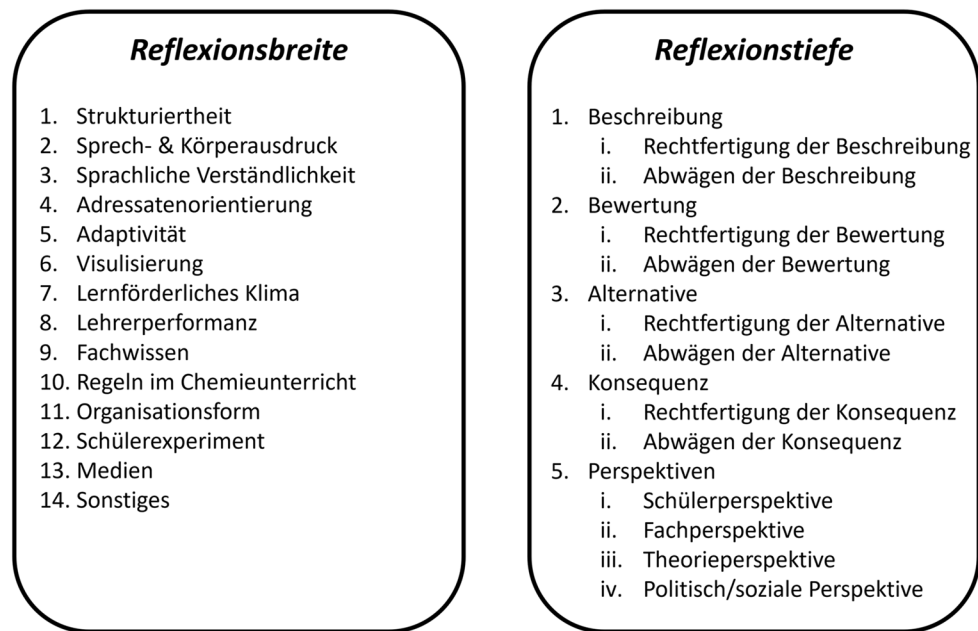
Um den Lernfortschritt zu messen, wurde ein Wissenstest mit 21 Items im Multiple-Choice-Single-Select-Format entwickelt. Der Test umfasst drei bis vier Fragen pro Thema zu Experimenten, Phänomenen und Transferfragen (Anhang D, Cronbachs  $\alpha$  (gesamter Test) = 0,64). (Die Werte zu den jeweiligen Erkläreinheiten siehe Anhang D). Die Daten sind weder im Pre- noch im Posttest normalverteilt. Dies ist möglicherweise auf die kleine und heterogene Stichprobe zurückzuführen: Die Schülerinnen und Schüler kommen von unterschiedlichen Schulen mit unterschiedlichen Lehrkräften. Trotz der formal fehlenden Voraussetzungen werden die Ergebnisse des t-Tests bzw. der ANOVA aufgrund der Stabilität bei Verletzungen der Annahmen berichtet.

Ein ergänzender Bewertungsbogen (Cronbachs  $\alpha$  = 0,91) bewertet Adressatenorientierung, Strukturiertheit, Verständlichkeit und Persönlichkeit (Lindl et al. 2019). Zusätzlich wurden Fragen zur Visualisierung, Adaptivität und sprachli-

chen Verständlichkeit neu entwickelt (Findeisen 2017; Kulgemeyer 2019) sowie drei Items zur Wirkung auf den Lernerfolg und zur Wissenschaftlichkeit aufgenommen (Elmer, in Druck). Die Bewertungen erfolgten auf einer sechsstufigen Likert-Skala von „trifft nicht zu“ bis „trifft voll zu“ (34 Items). Für die Bewertungsbögen wurde eine 6-stufige Likert-Skala gewählt. Diese erleichtert den Bezug zu Schulnoten, da sie der typischen Notenskala von 1 (trifft voll zu) bis 6 (trifft nicht zu) entspricht. Dadurch können Ergebnisse intuitiv interpretiert und mit bekannten Bewertungssystemen verglichen werden. Die Mittelwerte der Unterkategorien und des Summenwertes des Bewertungsbogens weisen keine Normalverteilung auf (Anhang E, F). Trotz dieser Abweichung von den Voraussetzungen werden die Ergebnisse des t-Tests aufgrund seiner Stabilität bei Verletzungen der Annahmen berichtet.



**Abb. 4** Kategorien der Reflexionsbreite und -tiefe



## Ergebnisse

### Wissen zum adaptiven Erklären

Zur Beantwortung der Hypothesen H1.1a (*Wissenszuwachs Erklärende-3× vs. Erklärende-2×*) und H1.1b (*Wissenszuwachs Erklärende-3× vs. Beobachtende*) wurden zwei lineare gemischte Regressionsmodelle (LMM) für den Wissenstest zum adaptiven Erklären berechnet, hier hypothesenkonform zunächst für die Gruppen *Beobachtende* und *Erklärende-3×* (Anhang H, Referenzwerte: *Beobachtende*, Pre-Test). Gründe für die Verwendung und eine Erläuterung des LMM finden sich in Anhang G. Es zeigt sich ein signifikanter Zuwachs zwischen Pre- und Post- ( $N=42$ ;  $\beta=4,29$ ;  $p=5,25 \cdot 10^{-8}$ \*\*) sowie zwischen Pre- und Follow-up-Test ( $N=31$ ;  $\beta=3,46$ ;  $p<0,001$ \*\*) bei den *Beobachtenden* und *Erklärenden-3×*. Es gibt keine signifikanten Unterschiede in der Interaktion des Lernzuwachses zwischen den Gruppen (Post-Test:  $\beta=-0,60$ ;  $p=0,540$ ; Follow-Up-Test:  $\beta=0,60$ ;  $p=0,969$ ).

Anschließend erfolgt der Gruppenvergleich zwischen *Erklärende-3×* und *Erklärende-2×* analog (Abb. 5, Referenzwerte: *Erklärenden-2×*; Pre-Test). Es zeigen sich signifikante Lernzuwächse vom Pre- zum Post- und zum Follow-up-Test sowie keine signifikanten Unterschiede in der Interaktion des Lernzuwachses und den Gruppen (Anhang I). Damit lernen alle drei Gruppen in Bezug auf das Wissen zum adaptiven Erklären in ähnlichem Maße dazu.

### Kodiermanuale zum adaptiven Erklären

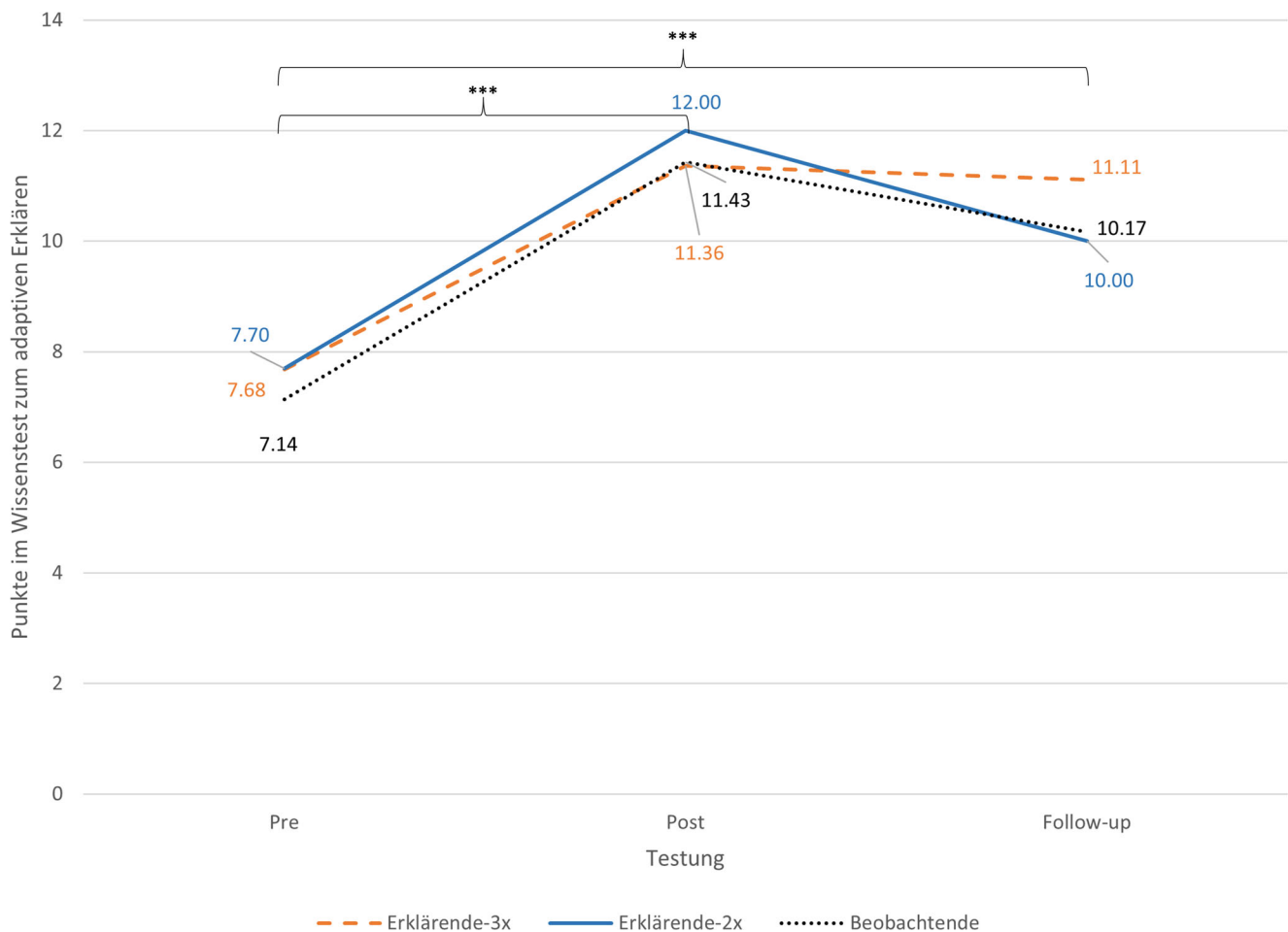
#### Adaptives Erklären – Planung

Um die Hypothesen H1.2a (*Entwicklung adaptive Erklärkompetenz in der Planung Erklärende-3× vs. Erklärende-2×*) und H1.2b (*Entwicklung adaptive Erklärkompetenz in der Planung Erklärende-3× vs. Beobachtende*) zu beantworten, wurden zwei LMM berechnet. Im ersten LMM (Referenzwerte: *Beobachtende*, erster Erklärtag; IG: *Erklärende-3×*) zeigt sich kein signifikanter Lernzuwachs bei den *Beobachtenden* (zu MZP2:  $\beta=0,06$ ;  $p=0,557$ ; zu MZP3:  $\beta=0,06$ ;  $p=0,518$ ) (Abb. 6). Die nicht signifikanten Interaktionseffekte (MZP2\**Erklärende-3*:  $\beta=0,17$ ;  $p=0,197$ ; MZP3\**Erklärende-3*:  $\beta=0,22$ ;  $p=0,091$ ) deuten darauf hin, dass es keine signifikanten Gruppenunterschiede gibt (Anhang J).

Wird ein weiteres LMM (Referenzpunkte: *Erklärende-2×*, erster Erklärtag) ohne die *Beobachtenden*-Gruppe berechnet, finden sich signifikante Lernzuwächse zwischen MZP1 und 2 sowie zwischen MZP1 und 3 in beiden Treatment-Gruppen (Anhang K).

#### Explorative Datenanalyse (Planung)

Mit den bisherigen Berechnungen wurden alle Messzeitpunkte des Testinstruments und jeweils zwei Gruppen miteinander verglichen, um die wesentlichen Hypothesen mit einem LMM zu prüfen. Im nächsten Schritt werden die Unterschiede in der Grafik sowie zwischen einzelnen Subfacetten statistisch überprüft. Besonderes Augenmerk wird dabei auf die unterschiedlichen Ergebnisse in Bezug auf



**Abb. 5** Wissenstest – Entwicklung des Wissens zum adaptiven Erklären in den Gruppen *Erklärende-3x*, *Erklärende-2x* und *Beobachtende*

die Gruppe der *Erklärenden-3x* gelegt, die im ersten LMM (Referenzwerte: *Beobachtende*, erster Erklärtag; IG: *Erklärende-3x*) keinen signifikanten Lernzuwachs, im zweiten LMM (Referenzpunkte: *Erklärende-2x*, erster Erklärtag, IG: *Erklärende-3x*) jedoch signifikante Lernzuwächse zu verzeichnen hat. Es sollen signifikante Hinweise auf Unterschiede aufgezeigt werden, ohne zuvor theoretisch abgeleitete Hypothesen zu testen oder verallgemeinern zu wollen. Daher wird auf eine Bonferroni-Korrektur des Alpha-Niveaus verzichtet. Auch wenn die Daten nicht zu jedem Messzeitpunkt normalverteilt sind (Anhang L), werden t-Tests (für gepaarte Stichproben) durchgeführt, da diese robust gegenüber Verletzungen ihrer Voraussetzungen und allgemein anerkannt sind.

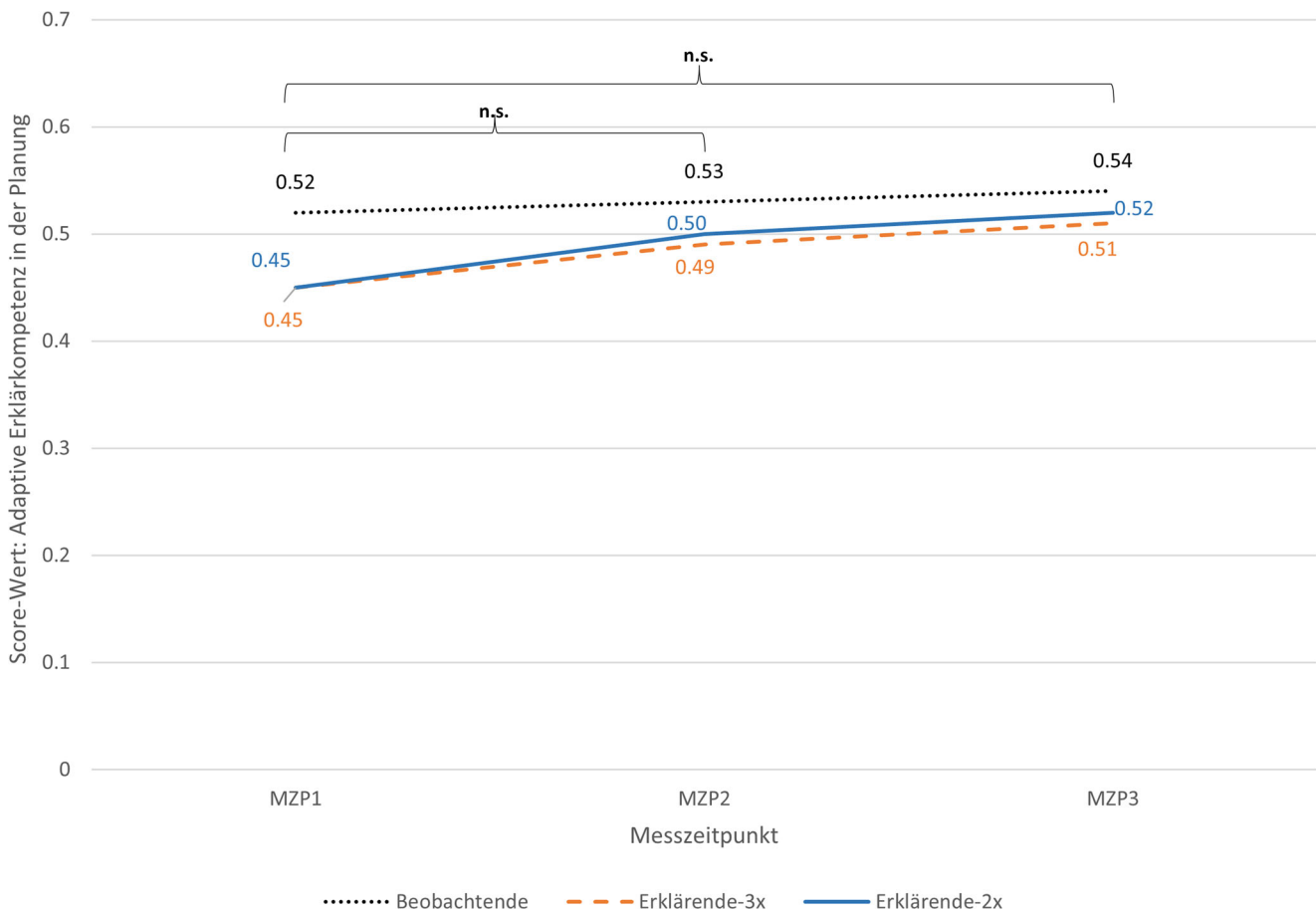
Die *Erklärenden-3x* zeigen substanzielle Lernzuwächse zwischen dem ersten und zweiten ( $t(21) = -2,29$ ;  $p = 0,032$ ;  $d_z = -0,489$ ) sowie dem ersten und dritten Erklärtag ( $t(21) = -2,96$ ;  $p = 0,007$ ;  $d_z = -0,631$ ). Sehr ähnlich verhält es sich mit den *Erklärenden-2x* zwischen dem ersten und zweiten ( $t(17) = -2,15$ ;  $p = 0,046$ ;  $d_z = -0,507$ ) sowie ersten und dritten ( $t(17) = -3,24$ ;  $p = 0,005$ ;  $d_z = -0,762$ ) Erklärtag. Dies

ist bei den *Beobachtenden* weder zwischen dem ersten und zweiten ( $t(18) = -0,625$ ;  $p = 0,540$ ;  $d_z = -0,143$ ) noch dem ersten und dritten Erklärtag ( $t(18) = -0,92$ ;  $p = 0,371$ ;  $d_z = -0,211$ ) der Fall.

### Adaptives Erklären – Handlung

Da die *Erklärenden-2x* nur am ersten und dritten Erklärtag adaptiv handeln, können über die beiden Erklärendengruppen nur diese beiden Messzeitpunkte miteinander verglichen werden. Um die Hypothese H1.3a (*Zuwachs adaptive Erklärkompetenz in der Handlung Erklärende-3x vs. Erklärende-2x*) zu beantworten, wird ein LMM (Referenzwerte: *Erklärende-2x*), Erklärtag 1 berechnet (Abb. 7).

Bei den *Erklärenden-2x* kann zwischen dem ersten und dritten Erklärtag knapp kein signifikanter Zuwachs gefunden werden ( $\beta = 0,04$ ;  $p = 0,060$ ). Es gibt keine signifikanten Interaktionseffekte mit dem dritten Erklärtag und den *Erklärenden-3x* ( $\beta = 0,01$ ;  $p = 0,795$ ) (Anhang M). Wie zuvor soll eine explorative Datenanalyse Effekte auf Subfacettenebene aufklären.



**Abb. 6** Entwicklung der adaptiven Erklärkompetenz in der Planung – *Erklärende-3x*, *Erklärende-2x* und *Beobachtende*

### Zusammenhänge – Reflektieren und adaptive Erklärkompetenz in Planung und Handlung

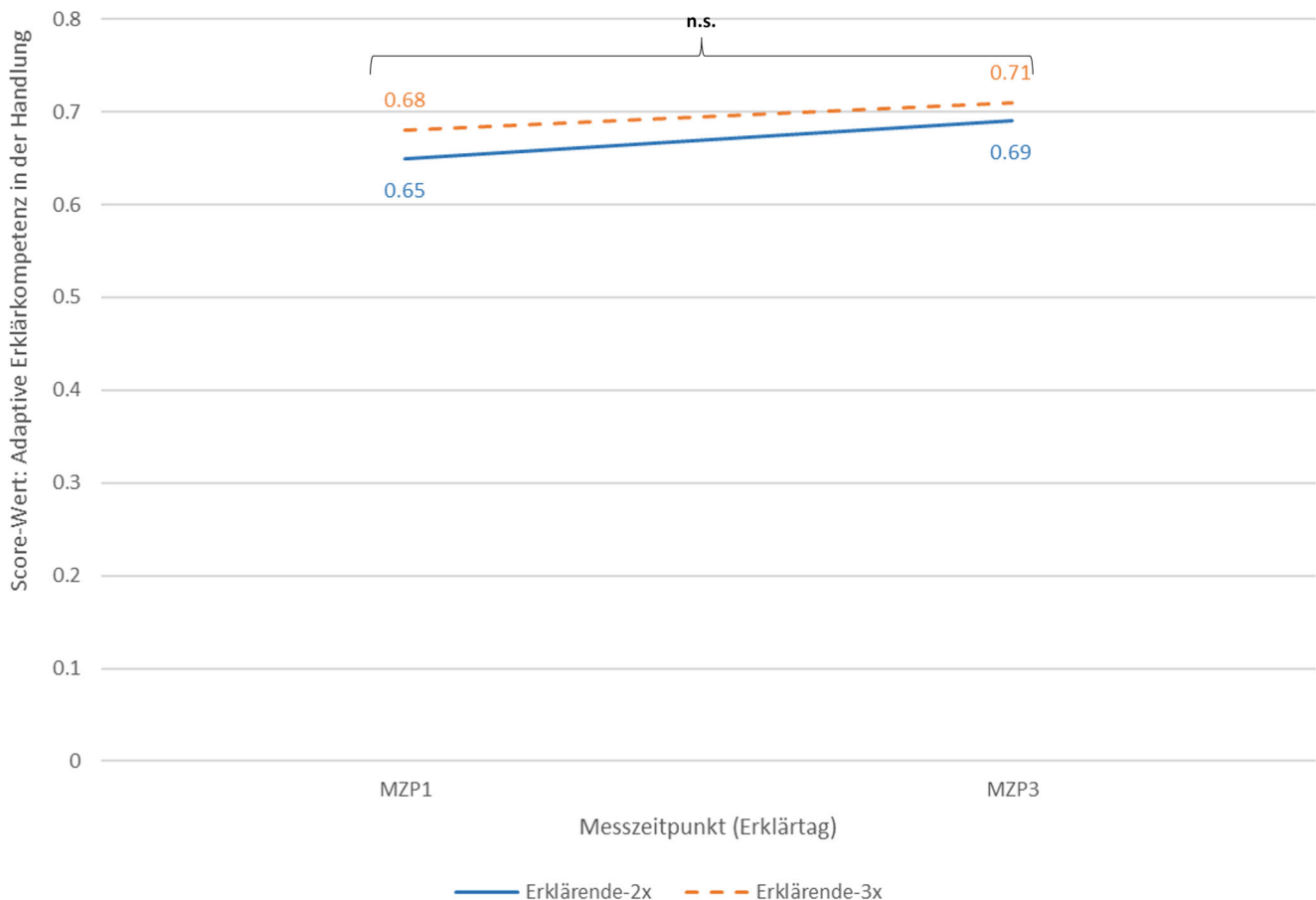
Um die Hypothesen H2a (*Zusammenhang Reflexionskompetenz und adaptive Erklärkompetenz – Planung*), H2b (*Zusammenhang Reflexionskompetenz und adaptive Erklärkompetenz – Handlung*) und H2c (*Zusammenhang der adaptiven Erklärkompetenz in Planung und Handlung*) beantworten zu können, wurden ANOVAs berechnet. Mit dem LMM wäre das zwar auch möglich, würde aber zu beta-Werten führen, die mit den bisher berechneten Werten nicht vergleichbar wären.

Wir berichten Zusammenhänge des letzten Reflexionszyklusses also dem Zyklus, bei dem der Lernerfolg beim Reflektieren, dem Planen und Handeln am größten sein sollte. Dafür wurden ANOVAs der Reflexionskompetenz (MZIP3, Kovariate) und die darauffolgende Planung (MZIP3, abhängige Variable) und Handlung (MZIP3, abhängige Variable) berechnet. Die Reflexionskompetenz zeigt keine signifikanten Zusammenhänge mit dem Scorewert der Planung ( $F=0,24$ ;  $p=0,627$ ;  $\eta^2=0,006$ ). Die ANOVA der Reflexionskompetenz und der darauffolgenden Handlung im adaptiven Erklären kann keinen signifikanten Zusammenhang nach-

weisen ( $F=0,15$ ;  $p=0,703$ ;  $\eta^2=0,008$ ). Die ANOVA zwischen dem Gesamtscore Video (abhängige Variable) und Gesamtscore Planungen (Kovariate) zeigt einen signifikanten Zusammenhang ( $F=7,45$ ;  $p=0,013$ ;  $\eta^2=0,282$ ).

### Weitere Datenanalyse

Um einer Alphafehler-Kumulierung Rechnung zu tragen, wird das Alpha-Niveau nach Bonferroni auf 0,005 korrigiert. Während bei alleiniger Betrachtung der Reflexionskompetenz zu MZIP3 keine Zusammenhänge berichtet werden können, kann der standardisierte residuale Lernzuwachs der Reflexionskompetenz über das gesamte Seminar (MZIP1–4) einen Ansatz darstellen, welcher der gemessenen Handlungs- und Planungskompetenz besser Rechnung trägt. Zunächst findet sich kein signifikanter Zusammenhang zwischen dem standardisierten residualen Lernzuwachs der Reflexionskompetenz (MZIP1–4) und dem Gesamtscore des adaptiven Erklärens in Planung und Handlung. Allerdings scheint ein starker Zusammenhang zwischen dem standardisierten residualen Lernzuwachs im Reflektieren und der Subfacette Adaptivität zu bestehen (Planen:  $F=10,49$ ;  $p=0,002$ ;  $\eta^2=0,216$ ). Zudem kann ein sig-



**Abb. 7** Entwicklung der adaptiven Erklärkompetenz in der Handlung – *Erklärende-2x* und *Erklärende-3x*

nifikanter Zusammenhang des standardisierten residualen Lernzuwachs der Reflexionsbreite und der Subfacette Adaptivität (Planung:  $F=9,23$ ;  $p=0,004$ ;  $\eta^2=0,195$ ) berichtet werden. Ebenfalls ergibt sich ein signifikanter Zusammenhang des standardisierten residualen Lernzuwachs der Reflexionstiefe und der Subfacette Adaptivität (Planung:  $F=8,69$ ;  $p=0,005$ ;  $\eta^2=0,186$ ) (Anhang N).

### Wissenstest Schülerinnen und Schüler

Um die Hypothese H3a (*Wissenszuwachs Lernende dritter vs. erster Schülertag bzw. Erklärende-3x, Erklärende-2x*) zu beantworten, wurde ein t-Test für unabhängige Stichproben mit dem standardisierten residualen Lernzuwachs der Schülerinnen und Schüler zwischen MZP1 und 3 berechnet. Das Alpha-Niveau wurde nach Bonferroni auf  $\alpha=0,025$  angepasst. Es kann kein signifikanter Unterschied des standardisierten residualen Lernzuwachses zwischen dem ersten und dritten Schülertag festgestellt werden ( $t(1354)=-0,15$ ;  $p=0,441$ ;  $d_z=-0,008$ , Mann-Whitney-U-Test siehe Anhang O). Dies gilt unabhängig davon, ob die Schülerinnen und Schüler von den *Erklärenden-2x* oder *Erklärenden-3x* unterrichtet wurden.

Um zu eruieren, ob Unterschiede im Zuwachs des standardisierten residualen Lernzuwachses (zu jedem Schülertag MZP1–3, pre-post) der Schülerinnen und Schüler in Abhängigkeit von der Gruppenzugehörigkeit (*Erklärende-2x, Erklärende-3x*) bestehen, wurde zusätzlich eine Varianzanalyse berechnet. Die Varianzanalyse zeigt, dass die Gruppenzugehörigkeit (*Erklärende-2x, Erklärende-3x*) einen signifikanten Einfluss auf den standardisierten residualen Lernzuwachs zwischen dem ersten und dem dritten Erklärtag hat ( $F=39,78$ ;  $p<0,001$ ;  $\eta^2=0,021$ ). Dabei war nur der standardisierte residuale Lernzuwachs am MZP3 derjenigen Lernenden größer als am MZP1, die von den Erklärenden-2x unterrichtet wurden. Die Erklärenden-3x erreichten bei den Schüler:innen bereits am ersten Schülertag höhere residuale Lernzuwächse als die Erklärenden-2x und keine weitere Steigerung (Anhang P).

### Bewertungsbogen Schülerinnen und Schüler

Um die Hypothese H3b (*Bewertungsbogen zum adaptiven Erklären dritter vs. erster Schülertag bzw. Erklärende-3x, Erklärende-2x*) zu beantworten, wurden für beide Gruppen jeweils zehn unabhängige t-Tests zwischen dem ersten

**Tab. 1** Charakteristika der Stichprobe

	Erklärende-3×	Erklärende-2×	Beobachtende
Anzahl	22	18	20
Semester (Durchschnitt)	7	6,94	6,65
Weiblich	11	7	12
Männlich	11	11	8
Lehramt Gymnasium	16	14	16
Lehramt Realschule	6	4	4

**Tab. 2** Zusammenfassung aller wesentlichen Ergebnisse

Signifikante Ergebnisse	Nicht signifikante Ergebnisse
<i>Wissenstest</i>	<i>Wissenstest</i>
Signifikanter Zuwachs bei Pre-Post- und Pre-Follow-up-Tests der Beobachtenden sowie Erklärenden-2× und -3×	Keine signifikanten Gruppenunterschiede im Lernzuwachs
<i>Adaptives Erklären in der Planung</i>	<i>Adaptives Erklären in der Planung</i>
Signifikante Lernzuwächse bei den Erklärenden-2× und -3×	Keine signifikanten Lernzuwächse bei Erklärenden-3× und Beobachtenden
Explorativ: Erklärenden-3× und Erklärende-2× zeigen substantielle Lernzuwächse zwischen dem ersten und zweiten sowie dem ersten und dritten Erklärtag	Explorativ: Beobachtende zeigen keine signifikanten Lernzuwächse zwischen den Erklärtagen
<i>Adaptives Erklären in der Handlung</i>	<i>Adaptives Erklären in der Handlung</i>
	Erklärende-2× und -3× zeigen zwischen dem ersten und dem dritten Erklärtag knapp keinen signifikanten Zuwachs
<i>Zusammenhänge: Reflexion, Planung &amp; Handlung</i>	<i>Zusammenhänge: Reflexion, Planung &amp; Handlung</i>
Signifikanter Zusammenhang zwischen Gesamtscore Handlung und Planungen	Kein signifikanter Zusammenhang der Reflexionskompetenz mit dem Gesamtscore Handlung oder Planung
Weitere Berechnungen: Es finden sich Zusammenhänge zwischen dem standardisierten residualen Lernzuwachs der Reflexionskompetenz (-breite und -tiefe) mit der Subfacette Adaptivität in der Planung	Weitere Berechnungen: Zwischen dem standardisierten residualen Lernzuwachs der Reflexionskompetenz (MZP1–4) und dem Gesamtscore des adaptiven Erklärens in Planung und Handlung findet sich kein signifikanter Zusammenhang
<i>Wissenstest Schülerinnen und Schüler</i>	<i>Wissenstest Schülerinnen und Schüler</i>
	Kein signifikanter Unterschied des standardisierten residualen Lernzuwachses zwischen dem ersten und dritten Schülertag, sowie keine Gruppenunterschiede
<i>Bewertungsbögen Schülerinnen und Schüler</i>	<i>Bewertungsbögen Schülerinnen und Schüler</i>
Zum dritten Erklärtag zeigen sich signifikante Gruppenunterschiede in den Subfacetten Strukturiertheit, Adressatenorientierung, sprachliche Verständlichkeit Visualisierung, Persönlichkeit und dem Mittelwert des gesamten Bewertungsbogen zugunsten der Erklärenden-2×	Keine signifikante Verbesserung des gesamten Erklärkonstrukts am dritten Erklärtag im Vergleich zum ersten (in beiden Gruppen). Die Adaptivität wird signifikant besser bewertet, während die Wirkung auf den Lernerfolg und auf die Wissenschaftlichkeit signifikant schlechter eingeschätzt wird (beide Gruppen)

und dem dritten Erklärtag durchgeführt (vergleichbare Ergebnisse des Mann-Whitney-U-Tests, Anhang Q, R). Anschließend werden zehn ANOVAs durchgeführt, um Gruppenunterschiede im Gesamtkonstrukt und in den Subskalen zu analysieren. Das  $\alpha$ -Niveau wurde mit Bonferroni-Holm auf  $\alpha=0,03$  festgelegt (Anhang S).

Das gesamte Erklärkonstrukt zeigt in beiden Gruppen keine signifikante Verbesserung am dritten Erklärtag im Vergleich zum ersten Erklärtag. Allerdings wird bei den Erklärenden-3× die Adaptivität ( $t(734)=-6,47$ ;  $p<0,001$ ;  $d_z=-0,484$ ) signifikant besser bewertet, während die Wirkung auf den Lernerfolg ( $t(738)=3,16$ ;  $p=0,002$ ;  $d_z=0,236$ ) und auf die Wissenschaftlichkeit ( $t(734)=3,01$ ;  $p=0,003$ ;  $d_z=0,226$ ) signifikant schlechter eingeschätzt wird.

Dieser Trend gilt ebenso für die Erklärende-2× Gruppe (Anhang T).

Zum dritten Erklärtag zeigen sich signifikante Gruppenunterschiede in den Subfacetten Strukturiertheit ( $F=9,96$ ;  $p=0,002$ ;  $\eta^2=0,020$ ), Adressatenorientierung ( $F=7,28$ ;  $p=0,007$ ;  $\eta^2=0,014$ ), sprachliche Verständlichkeit ( $F=10,34$ ;  $p=0,001$ ;  $\eta^2=0,020$ ), Visualisierung ( $F=9,44$ ;  $p=0,002$ ;  $\eta^2=0,019$ ), Persönlichkeit ( $F=7,48$ ;  $p=0,006$ ,  $\eta^2=0,015$ ) und dem Mittelwert des gesamten Bewertungsbogen ( $F=7,81$ ;  $p=0,005$ ;  $\eta^2=0,016$ ) zugunsten der Erklärenden-2×. In Tab. 2 werden alle wesentlichen Ergebnisse der Studie zusammengefasst.



## Diskussion

Die Hypothesen H1.1a (*Wissenszuwachs Erklärende-3× vs. Erklärende-2×*) und H1.1b (*Wissenszuwachs Erklärende-3× vs. Beobachtende*) werden angenommen, da die Unterschiede im Lernzuwachs (Wissen zum adaptiven Erklären) zwischen den *Erklärenden-3×* und *Erklärenden-2×* sowie zwischen *Erklärenden-3×* und *Beobachtenden* nicht signifikant sind. Dies deutet darauf hin, dass alle Gruppen ähnlich viel gelernt haben. Der Wissenstest bestätigt die Wirksamkeit des Seminars (Intervention) (Cauet 2016). Da alle Gruppen der Studie die gleichen Seminarinhalte vermittelt bekamen, ist das Ergebnis erwartungskonform. Bedeutsam ist der Testteil insofern, da das Wissen zum adaptiven Erklären als Teil der Handlungskompetenz verstanden werden kann und somit die Basis für die nachfolgenden Kategorien guten adaptiven Erklärens (in Planung und Handlung) darstellt (Baumert und Kunter 2006).

Die Hypothesen H1.2a (*Entwicklung adaptive Erklärkompetenz in der Planung Erklärende-3× vs. Erklärende-2×*) und H1.2b (*Entwicklung adaptive Erklärkompetenz in der Planung Erklärende-3× vs. Beobachtende*) werden aufgrund der LMM-Berechnungen angenommen. Hinsichtlich des Zuwachses an adaptiver Erklärungskompetenz in der Planung gibt es keine signifikanten Interaktionen zwischen den Gruppen *Beobachtende* und *Erklärende-3×* sowie *Erklärende-2×* und *Erklärende-2×*. Dies weist auf fehlende Gruppenunterschiede hin. Der generelle Zuwachs auch der adaptiven Erklärkompetenz kann mit der Modellierung der professionellen Handlungskompetenz nach Baumert und Kunter (2006) verstanden werden, die das Professionswissen als bedeutsamen Aspekt der Handlungskompetenz verstehen. Möglicherweise ist die in allen Gruppen gleichermaßen vermittelte Wissensbasis ursächlich für die zunächst insgesamt nicht nachweisbaren Unterschiede in der Handlungskompetenz.

Weitere explorative Analysen zeigen allerdings, dass die *Erklärenden* im Gegensatz zu den *Beobachtenden* signifikante Zuwächse vom ersten bis zum dritten Erklärtag aufweisen, was die Hypothese H1.2b widerlegt. Die widersprüchlichen Ergebnisse deuten darauf hin, dass eine profunde Wissensbasis allein nicht ausreicht, um Expertenkompetenz zu erreichen (Neuweg 2002). Die wiederholte Anwendung und Reflexion von Wissen in Erklärsituationen scheint entscheidend für die Entwicklung adaptiver Erklärkompetenz in der Planung zu sein.

Die Zuwächse bei den *Beobachtenden* können mit der allgemeinen Förderung der Reflexionskompetenz zusammenhängen, da diese mit dem Professionswissen der Lehrkräfte in Zusammenhang zu stehen scheint (Kulgemeyer et al. 2021). Es kann daher angenommen werden, dass die *Beobachtenden* ähnliche reflexive Kompetenzen erworben haben, die sich in ähnlicher Weise auf die Anpassungsfähigkeit in der Planung auswirken.

Es erscheint plausibel, dass einzelne Planungsaspekte wie das literaturbasierte Berücksichtigen des Vorwissens auch ohne aktive Anwendung im Unterricht verbessert werden können.

Die Hypothese H1.3a (*Zuwachs adaptive Erklärkompetenz in der Handlung Erklärende-3× vs. Erklärende-2×*) muss abgelehnt werden. Die Gruppen zeigen einen – wenn auch knapp – nicht signifikanten Lernzuwachs (LMM). Dies bestätigt, dass Novizen Schwierigkeiten haben, adaptiv zu handeln (Borko und Livingston 1989) und deutet darauf hin, dass angehende Lehrkräfte in Handlungssituationen möglicherweise nicht gezielt auf ihr Wissen zurückgreifen können, um Entscheidungen zu treffen (Neuweg 2002). Beide Gruppen zeigen keinen signifikanten Interaktionseffekt (LMM) mit den Messzeitpunkten. Dies weist darauf hin, dass beide Gruppen in ähnlicher Weise hinzugelernt haben.

Der dritte adaptive Erklärungsversuch scheint jedoch einen wesentlichen Kompetenzzuwachs zu bewirken. Dies deutet darauf hin, dass wiederholte Unterrichtszyklen (Planung, Handlung, Reflexion) wichtig für die Entwicklung adaptiver Erklärkompetenz sind. Die Ergebnisse bestätigen, dass Reflexion und Erfahrung einen signifikanten Einfluss auf die Erklärungskompetenz haben (Zemba-Saul et al. 2000). Die Studie von Kobl (2021) weist nach, dass die Reflexionskompetenz bei Studierenden, die selbst unterrichtet haben, erst zwischen dem 2. und 3. Messzeitpunkt signifikant zunimmt.

Die Hypothesen H2a (*Zusammenhang Reflexionskompetenz und adaptive Erklärkompetenz – Planung*) und H2b (*Zusammenhang Reflexionskompetenz und adaptive Erklärkompetenz – Handlung*) müssen abgelehnt werden, da kein Zusammenhang zwischen Reflexionskompetenz und adaptiver Erklärkompetenz in Planung und Handlung nachweisbar ist. Dies bestätigt die Studie von Vogelsang et al. (2022), die beschreibt, dass Reflexion nicht zu den Planungsfähigkeiten angehender Lehrkräfte beiträgt. Allerdings zeigen sich signifikante Effekte des residualen Lernzuwachses der Reflexionskompetenz (Breite und Tiefe) mit der Subfacette Adaptivität in der Planung. Möglicherweise werden auf dieser Ausbildungsstufe der Studierenden (wenig Erfahrung mit komplexen Unterrichtssituationen) zunächst nur bestimmte Facetten erlernt und dies ist primär vom residualen Lernzuwachs der Reflexionskompetenz abhängig.

Die Hypothese H2c (*Zusammenhang der adaptiven Erklärkompetenz in Planung und Handlung*) kann angenommen werden, da zwischen der adaptiven Erklärkompetenz in Planung und Handlung ein signifikanter Zusammenhang gefunden wird. Dies bestätigt Studien, die einen Zusammenhang zwischen Unterrichtsplanung und -handlung zeigen (Kademian und Davis 2018). Das gewählte Seminar-konzept scheint also in diesem nicht selbstverständlichen Aspekt ebenfalls förderlich zu sein.

Die Kodiermanuale für Planung und Handlung berücksichtigen insbesondere fächerübergreifende Kriterien guten Erklärens (Lindl et al. 2019). Als chemiespezifisches Kriterium wurde lediglich das Stoff-Teilchen-Konzept (Sevian und Gonsalves 2008) in die Bewertung aufgenommen. Die Definition chemiespezifischer Kriterien erwies sich als herausfordernd, da die Forschungslage im Fach Chemie gering ist bzw. die Forschungsarbeiten nicht an das adaptive Konzept dieser Studie einfach zu übertragen waren. Weitere potenziell relevante Kriterien wie Fachlichkeit (Schopf und Zwischenbrugger 2015) und Schülervorstellungen sind stark themenabhängig und nicht eindeutig zu bewerten. Chemiespezifische Aspekte guten Erklärens konnten in dieser Studie daher nicht – wie ursprünglich intendiert – noch stärker berücksichtigt werden und bedürfen weiterer Forschung. Die Hypothese H3a (*Wissenszuwachs Lernende dritter vs. erster Schülertag bzw. Erklärende-3×, Erklärende-2×*), die einen höheren Lernerfolg der Schülerinnen und Schüler am dritten im Vergleich zum ersten Schülertag postuliert, kann nur für die *Erklärenden-2×* bestätigt werden. Über beide Gruppen hinweg zeigt sich kein signifikanter Wissenszuwachs vom 1. zum 3. MPZ der (unterschiedlichen) Schulklassen. Das Ausbleiben des Zuwachses über die Messzeitpunkte kann darauf zurückzuführen sein, dass eine Verbesserung der Studierenden-erklärungen in einzelnen Subfacetten nicht generell zu einem höheren Lernzuwachs der Schülerinnen und Schüler führt, wenn es nicht effektiv genutzt werden kann. Die Varianzanalyse zeigt, dass die Gruppenzugehörigkeit (*Erklärende-2×, Erklärende-3×*) einen signifikanten Einfluss auf den standardisierten residualen Lernzuwachs zwischen dem ersten und den dritten Erklärtag hat. Es zeigt sich, dass die *Erklärenden-2×* einen besseren Lernzuwachs der Schüler:innen bewirken als die *Erklärenden-3×*. Möglicherweise lernen die *Erklärenden-2×* erst beim letzten MZP3 stärker dazu, während die *Erklärenden-3×* bereits den MZP2 für die Verbesserung ihrer Handlungskompetenz nutzen können. Die Analyse der deskriptiven Daten zeigt allerdings, dass die *Erklärenden-3×* grundsätzlich höhere Lernzuwächse bei den Schüler:innen bewirken – an MZP1 und 3. Der Lernzuwachs ist an beiden MZP ähnlich. Die *Erklärenden-2×* weisen bereits an MZP1 einen geringeren Lernzuwachs ihrer Lernenden auf als die der *Erklärenden-3×*. Sie können zum MZP 3 den Lernerfolg der SuS steigern, erreichen aber generell geringere Lernzuwächse der Schülerinnen und Schüler als die *Erklärenden-3×*. Zu vermuten sind hier Stichprobeneffekte – trotz der randomisierten Zuteilung der Studierenden auf die drei Gruppen.

Die Hypothese H3b (*Bewertungsbogen zum adaptiven Erklären dritte vs. erster Schülertag bzw. Erklärende-3×, Erklärende-2×*) muss bezogen auf das Gesamtkonstrukt abgelehnt werden, da sich am dritten Schülertag keine signifikant besseren Gesamtbewertungen zeigen. Ebenfalls sind

keine signifikanten Unterschiede in der Erklärbewertung zwischen beiden Gruppen erkennbar. Die Hypothese kann allerdings in der Subfacette Adaptivität bestätigt werden. In beiden Erklärgruppen wird die Adaptivität zum dritten Messzeitpunkt besser eingeschätzt als zu Beginn. Dies wird auch durch das Expertenrating in der Planung bestätigt, jedoch nicht in der Handlung. Hier zeigt nur die *Erklärende-3×*-Gruppe eine signifikante Verbesserung. Dies deutet darauf hin, dass die zusätzliche Erklärgelegenheit der *Erklärenden-3×*-Gruppe vorteilhaft ist.

Die subjektive Wahrnehmung des Lernerfolgs und der Wissenschaftlichkeit der Erklärungen nimmt vom Anfang zum Ende des Seminars signifikant ab. Kulgemeyer et al. (2022) postulieren, dass Lehrbuchtexte als verlässlicher und ernsthafter empfunden werden als Erklärvideos, da sie formaler wirken. Dies führt möglicherweise dazu, dass Lernende ihren Lernerfolg aus diesen Texten überschätzen. Die Ergebnisse dieser Studie deuten darauf hin, dass die adaptiven Erklärungen der Studierenden möglicherweise ebenfalls weniger verlässlich wirken und deshalb nicht als besser wahrgenommen werden. Möglicherweise werden vereinfachte und adressatengerechtere Erklärungen mit weniger Fachtermini als wenig wissenschaftlich eingeschätzt, während sie dem Verständnis und dem Lernerfolg zuträglich sind.

Die Ergebnisse der Studie sollten mit folgenden Einschränkungen betrachtet werden. Trotz hoher Lernzuwächse, Effektstärken und signifikanter Gruppendifferenzen in den t-Tests und ANOVAs können diese aufgrund der begrenzten Stichprobengröße nicht verallgemeinert werden. Durch die geringe Stichprobe können nur sehr große Effekte nachgewiesen werden. Dies lässt vermuten, dass die Zuwächse und Gruppenunterschiede in der Studie mit einer größeren Teilnehmerzahl evtl. auch in den LMM gezeigt werden können. Ein weiterer limitierender Faktor ist das Nutzen einer Gelegenheitsstichprobe. Die Studierenden sind in unterschiedlichen Semestern und besitzen unterschiedliche fachliche Voraussetzungen. Gleichzeitig können dadurch Kohorteneffekte minimiert werden (Weißbach und Kulgemeyer 2024). Um vergleichbare Ausgangsbedingungen sicherzustellen, mussten alle Studierenden die erste chemiedidaktische Vorlesung und das experimentelle Seminar erfolgreich abgeschlossen haben. Das Thema „Chemische Reaktion und Energetik“ wurde bereits im ersten Semester behandelt und stellt keine fachliche Hürde dar.

## Zusammenfassung und Ausblick

Die Studie zeigt, dass adaptive Erklärungskompetenz in einem universitären Seminar in spezifischen Planungs- und Handlungsaspekten wie Adaptivität und Adressatenorientierung gefördert werden kann. Differenzierte Entwicklun-

gen in den Gruppen und der Einfluss von Wiederholungen werden deutlich. Für die Lehrkräftebildung sind die Ergebnisse relevant, da nicht alle Studierenden regelmäßig aktive Lerngelegenheiten haben. Die Erkenntnis, dass adaptive Erklärungskompetenz auch durch die Beobachtung und Reflexion adaptiver Erklärungen anderer gestärkt wird, erleichtert die Lehrerausbildung und den Wechsel zwischen eigenem Erklären und Beobachten. Es kann gezeigt werden, dass es sinnvoll ist, den ePCK-Zyklus (plan, teach, reflect) (Carlson et al. 2019) zur Förderung professioneller Kompetenzen zu nutzen. Zukünftige Studien können die explorativ gefundenen Ergebnisse zu einzelnen Subfacetten überprüfen und ggf. in die Gestaltung bestehender Seminare zur Reflexions- und Erklärkompetenz einfließen lassen. Das komplexe Studiendesign kann als Vorlage für entsprechenden Vergleichs-Studien in unterrichtsnahen Settings dienen.

**Zusatzmaterial online** Der Anhang ist in der Online-Version dieses Artikels (<https://doi.org/10.1007/s40573-025-00183-w>) enthalten.

**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

- Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: enacted and personal PCK. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 271–286). Singapore: Springer.
- Alonzo, A. C., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: enacted and personal PCK. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 271–286). Singapore: Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_15](https://doi.org/10.1007/978-981-13-5898-2_15).
- von Aufschnaiter, C., Fraij, A., & Kost, D. (2019). Reflexion und Reflexivität in der Lehrerbildung. Herausforderung Lehrer\*innenbildung-Zeitschrift zur Konzeption. *Gestaltung und Diskussion*, 2(1), 144–159. <https://doi.org/10.4119/hlz-2439>.
- Barbieri, M. S., Colavita, F., & Scheuer, N. (1989). Explanations: A pragmatic basis for early child competence. *IPRA Papers in Pragmatics*, 3(1), 130–154. <https://doi.org/10.1075/iprapip.3.1.04bar>.
- Bartels, H., & Kulgemeyer, C. (2019). Explaining physics: an online test for self-assessment and instructor training. *European Journal of Physics*, 40(1), 15701. <https://doi.org/10.1088/1361-6404/aab55e>.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P., et al. (2008). *Adaptive Lehrkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrwissens*. Pädagogische Psychologie und Entwicklungspsychologie, Bd. 63. Münster, München, Berlin: Waxmann.
- Blömeke, S., Paine, L., Houang, R. T., Hsieh, F.-J., Schmidt, W. H., Totto, M. T., et al. (2008). Future teachers' competence to plan a lesson: first results of a six-country study on the efficiency of teacher education. *ZDM—Mathematics Education*, 40(5), 749–762. <https://doi.org/10.1007/s11858-008-0123-y>.
- Borko, H., Eisenhart, M., Brown, C., Unterhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard Mathematics—Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, 23(3), 194–222. <https://doi.org/10.5951/jresmetheduc.23.3.0194>.
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473–498. <https://doi.org/10.3102/00028312026004473>.
- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., & Borowski, A., et al. (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). Singapore: Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_15](https://doi.org/10.1007/978-981-13-5898-2_15).
- Cauet, E. (2016). *Testen wir relevantes Wissen? Studien zum Physik- und Chemielernen*, Bd. 204. Dissertation
- Evans, W. E., & Guyson, R. F. (1978). *Clarity of explanation: A powerful indicator of teacher effectiveness*
- Findeisen, S. (2017). *Fachdidaktische Kompetenzen angehender Lehrpersonen. Eine Untersuchung zum Erklären im Rechnungswesen*. Economics education and human resource management. Wiesbaden: Springer.
- Guler, M., & Celik, D. (2016). A research on future mathematics teachers instructional explanations: The sample of Algebra. *Educational Research and Reviews*, 11(16), 1500–1508. <https://doi.org/10.5897/ERR2016.2823>.
- Hilbert, S., Stadler, M., Lindl, A., Naumann, F., & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*. <https://doi.org/10.4473/TPM26.1.6>.
- Housner, L. D., & Griffey, D. C. (1985). Teacher cognition: differences in planning and interactive decision making between experienced and inexperienced teachers. *Research Quarterly for Exercise and Sport*, 56(1), 45–53. <https://doi.org/10.1080/02701367.1985.10608430>.
- Inoue, N. (2009). Rehearsing to teach: content-specific deconstruction of instructional explanations in pre-service teacher training. *Journal of Education for Teaching*, 35(1), 47–60. <https://doi.org/10.1080/02607470802587137>.
- Kademian, S. M., & Davis, E. A. (2018). Supporting beginning teacher planning of investigation-based science discussions. *Journal of Science Teacher Education*, 29(8), 712–740. <https://doi.org/10.1080/1046560X.2018.1504266>.
- Kiel, E. (1999). *Erklären als didaktisches Handeln*. Würzburg: Ergon.
- Kobl, C. (2021). *Förderung und Erfassung der Reflexionskompetenz im Fach Chemie*. Studien zum Physik- und Chemielernen, Bd. 312. Dissertation



- Korthagen, F., & Vasalos, A. (2005). Levels in reflection: core reflection as a means to enhance professional growth. *Teachers and Teaching*, 11(1), 47–71. <https://doi.org/10.1080/1354060042000337093>.
- Kulgemeyer, C. (2019). Towards a framework for effective instructional explanations in science teaching. *Studies in Science Education*, 54(2), 109–139. <https://doi.org/10.1080/03057267.2018.1598054>.
- Kulgemeyer, C., & Geelan, D. (2024). Towards a constructivist view of instructional explanations as a core practice of science teachers. *Science Education*, 108(4), 1034–1050. <https://doi.org/10.1002/sce.21863>.
- Kulgemeyer, C., Hörnlein, M., & Sterzing, F. (2022). Exploring the effects of physics explainer videos and written explanations on declarative knowledge and the illusion of understanding. *International Journal of Science Education*, 44(11), 1855–1875. <https://doi.org/10.1080/09500693.2022.2100507>.
- Kulgemeyer, C., Kempin, M., Weißbach, A., Borowski, A., Buschhüter, D., Enkrott, P., et al. (2021). Exploring the impact of pre-service science teachers' reflection skills on the development of professional knowledge during a field experience. *International Journal of Science Education*, 43(18), 3035–3057. <https://doi.org/10.1080/09500693.2021.2006820>.
- Kulgemeyer, C., & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, 55(10), 1393–1418. <https://doi.org/10.1002/tea.21457>.
- Kulgemeyer, C., & Schecker, H. (2013). Students explaining science—Assessment of science communication competence. *Research in Science Education*, 43(6), 2235–2256. <https://doi.org/10.1007/s11165-013-9354-1>.
- Kulgemeyer, C., & Tomczyszyn, E. (2015). Physik erklären – Messung der Erklärnsfähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 111–126. <https://doi.org/10.1007/s40573-015-0029-5>.
- Kunter, M., & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85–114). Münster: Waxmann.
- Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78(2), 75–95. <https://doi.org/10.1037/0022-0663.78.2.75>.
- Lindl, A., Gaier, L., Weich, M., Frei, M., Ehres, C., Gastl-Pischetsrieder, M., et al. (2019). Eine ‚gute‘ Erklärung für alle?! Gruppenspezifische Unterschiede in der Beurteilung von Erklärqualität—erste Ergebnisse aus dem interdisziplinären Forschungsprojekt FALKE. Lehrer. Bildung. Gestalten. Beiträge zur empirischen Forschung in der Lehrerbildung. (S. 128–141).
- Merzyn, G. (2005). Junge Lehrer im Referendariat. *Der mathematische und naturwissenschaftliche Unterricht*, 58(1), 4–7.
- Miltz, R. J. (1972). *Development and evaluation of a manual for improving teachers explanations*. Technical Report, Bd. 26.
- Neuweg, G. H. (2002). Lehrerhandeln und Lehrerbildung im Lichte des Konzepts des impliziten Wissens. *Zeitschrift für Pädagogik*, 48(1), 10–29. <https://doi.org/10.25656/01:3818>.
- Odora, R. J. (2014). Using explanation as a teaching method: how prepared are high school technology teachers in free state Province, South Africa? *The Social Science Journal*, 38(1), 71–81.
- Parsons, S. A. (2012). Adaptive teaching in literacy instruction. *Journal of Literacy Research*, 44(2), 149–170. <https://doi.org/10.1177/1086296X12440261>.
- Schopf, C., & Zwischenbrugger, A. (2015). Verständliche Erklärungen im Wirtschaftsunterricht. Eine Heuristik basierend auf dem Verständnis der Fachdidaktiker/innen des Wiener Lehrstuhls für Wissenschaftspädagogik. *Zeitschrift für ökonomische Bildung*, 3, 1–31.
- Sevian, H., & Gonsalves, L. (2008). Analysing how scientists explain their research: a rubric for measuring the effectiveness of scientific explanations. *International Journal of Science Education*, 30(11), 1441–1467. <https://doi.org/10.1080/09500690802267579>.
- Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Treagust, D., & Harrison, A. (1999). The genesis of effective science explanations for the classroom. In J. Loughran (Hrsg.), *Researching teaching. Methodologies and practices for understanding pedagogy* (S. 28–43). Abingdon: Routledge.
- Vogelsang, C., Kulgemeyer, C., & Riese, J. (2022). Learning to plan by learning to reflect?—Exploring relations between professional knowledge, reflection skills, and planning skills of preservice physics teachers in a one-semester field experience. *education sciences*, 12, 479–500.
- Wagner, A., & Wörn, C. (2011). *Erklären lernen—Mathematik verstehen. Ein Praxisbuch mit Lernangeboten* (1. Aufl.). Seelze: Klett.
- Weißbach, A., & Kulgemeyer, C. (2024). Reflexionsfähigkeit von Physiklehramtsstudierenden: Ein Online-Self-Assessment mit Feedback. *Unterrichtswissenschaft*, 52(1), 135–161. <https://doi.org/10.1007/s42010-024-00195-4>.
- Wheeldon, R. (2011). Examining pre-service teachers' use of atomic models in explaining subsequent Ionisation energy values. *Journal of Science Education and Technology*, 21(3), 403–422. <https://doi.org/10.1007/s10956-011-9333-0>.
- Wyss, C. (2013). *Unterricht und Reflexion*. Empirische Erziehungswissenschaft, Bd. 44. Dissertation
- Zemba-Saul, C., Blumenfeld, P., & Krajcik, J. (2000). Influence of guided cycles of planning, teaching, and reflecting on prospective elementary teachers' science content representations. *Journal of Research in Science Teaching*, 37(4), 318–339.
- Zulliger, S., Buholzer, A., & Ruelmann, M. (2022). Observed quality of formative peer and self-assessment in everyday mathematics teaching and its effects on student performance. *European Journal of Educational Research*, 11(2), 663–680.
- Hemmerich, W. (2016). StatistikGuru: Rechner zur Adjustierung des  $\alpha$ -Niveaus. <https://statistikguru.de/rechner/adjustierung-des-alphaniveaus.html>

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