

Zweitveröffentlichung



Hess, Miriam; Denn, Ann-Katrin

"Ich habe den Schüler viel zu wenig gelobt." : Evaluation des Seminarerfolgs über videobasierte Reflexionen von Lehramtsstudierenden über ihr eigenes Feedbackverhalten

Datum der Zweitveröffentlichung: 05.09.2025

Verlagsversion (Version of Record), Zeitschriftenartikel

Persistenter Identifikator: urn:nbn:de:bvb:473-irb-93608x

Erstveröffentlichung

Hess, Miriam; Denn, Ann-Katrin (2023): "Ich habe den Schüler viel zu wenig gelobt." : Evaluation des Seminarerfolgs über videobasierte Reflexionen von Lehramtsstudierenden über ihr eigenes Feedbackverhalten, in: *Lehrerbildung auf dem Prüfstand, Landau/Pfalz: Empirische Pädagogik e.V.*, Jg. 16, Nr. 2, S. 204–228, doi: 10.62350/ghbh7658.

Rechtehinweis

Dieses Werk ist durch das Urheberrecht und/oder die Angabe einer Lizenz geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die für Sie geltende Gesetzgebung zum Urheberrecht und/oder durch die Lizenz erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

Für dieses Dokument gilt eine Creative-Commons-Lizenz.



Die Lizenzinformationen sind online verfügbar:

<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>

Miriam Hess & Ann-Katrin Denn

"Ich habe den Schüler viel zu wenig gelobt." Evaluation des Seminarerfolgs über videobasierte Reflexionen von Lehramtsstudierenden über ihr eigenes Feedbackverhalten

Obwohl Videoanalysen ein großes Potenzial für die Entwicklung berufsbezogener Kompetenzen zugeschrieben wird, stehen selten Wirkungen von (videobasierten) Seminaren auf die Handlungs- und Reflexionskompetenz von Lehramtsstudierenden im Mittelpunkt von Evaluationen. In der quasi-experimentellen Interventionsstudie ProFee wurde die Wirksamkeit verschiedener Seminarkonzepte auf die Kompetenzen von Grundschullehramtsstudierenden im Bereich Feedback untersucht. Im Fokus stand die Frage, wie sich unterschiedlich stark strukturierte Beobachtungsmethoden auf die Studierendenkompetenzen auswirken. Die nach dem Seminarbesuch videobasiert erfassten Handlungskompetenzen der N = 226 untersuchten Grundschullehramtsstudierenden unterscheiden sich unter Kontrolle des Vorwissens nicht signifikant zwischen den vier Kursen. Die Selbsteinschätzungen der Studierenden sowie die Zusammenhänge zwischen Selbst- und Fremdeinschätzung sind aber in derjenigen Seminargruppe enger, in der die Videos fremder Studierender sehr strukturiert, niedrig inferent analysiert wurden.

Schlagwörter: Feedback – Micro-Teaching – professionelle Wahrnehmung – Reflexion – Videoanalysen

"I have praised the student far too little." Evaluating seminar success through video-based reflections from student teachers on their own feedback behavior

Although video analyses have a great potential for the development of professional competencies, the effects of (video-based) seminars on teaching activities and reflection competencies of student teachers have rarely been the focus of evaluations. In the quasi-experimental intervention study ProFee, the effectiveness of different seminar concepts on the competencies of elementary school student teachers in giving feedback was investigated. The evaluation focused on the question of how observation assignments and methods with varying degrees of structure affect students' competencies. After attending the seminar, the teaching activity of the N = 226 students did not differ significantly between courses when controlling for prior knowledge. But the students' self-assessments as well as the correlations between self-assessment and external-assessment are closer in the group in which the videos were analyzed in a very structured, low inferential manner.

Keywords: feedback – microteaching – professional vision – reflection – video-analyses

1 Förderung von Handlungs- und Reflexionskompetenzen im Bereich Feedback durch die Analyse von Videos

Videoanalysen wird für die Lehrkräftebildung ein hohes Potenzial für die Entwicklung professioneller Kompetenzen zugeschrieben (Blomberg, 2011; Sherin, 2004; Syring, 2021). Allerdings fokussieren bisher nur wenige Evaluationen auf die Frage, welche Art der Arbeit mit Videos die höchsten Lernerfolge verspricht. Um den Erfolg von spezifischen Lehrkonzepten zur Arbeit mit Videos systematisch zu evaluieren, bieten insbesondere (quasi-)experimentelle Studien mit einem Prä-Post-Kontrollgruppendesign besondere Chancen, um spezifische Erfolgsbedingungen identifizieren zu können. Insbesondere für die Lehrkräftebildung sind auch Evaluationen interessant, die berufsbezogene Kompetenzen direkt in den Blick nehmen, wie beispielsweise die Planung, Durchführung und Reflexion eigener Unterrichtsversuche. Bisher stehen allerdings nur selten Auswirkungen von (videobasierten) Seminaren auf die tatsächliche Handlungs- und Reflexionskompetenz im Mittelpunkt von Seminarevaluationen.

Diese Desiderata werden im ProFee-Projekt (Professionell Feedback geben – Lernen mit Videos; Hess, 2019b) angegangen, einer quasi-experimentell angelegten Interventionsstudie, in der die Wirksamkeit verschiedener Seminarkonzepte auf die professionellen Kompetenzen von Grundschullehramtsstudierenden im Bereich Feedback – einer Kernpraktik im Unterricht (Grossman, 2018) – untersucht wurden.

1.1 Feedback

Das Feedback der Lehrperson wurde als Thema der Seminare ausgewählt, da Rückmeldungen von Lehrpersonen nachweislich einen bedeutsamen Einfluss auf die Leistungs- und Persönlichkeitsentwicklung von Schüler*innen ausüben (Hattie, 2023; Wisniewski, Zierer & Hattie, 2020). Auch die in den Standards für die Lehrkräftebildung formulierten Kompetenzen für die Bildungswissenschaften geben die Fähigkeit von Lehrpersonen, gute Rückmeldungen zu geben und sich der Wirkungen von Feedback bewusst zu sein, als wichtige Kompetenz an (KMK, 2019). Feedback nimmt in der Hattie-Studie (2009) unter den 138 beobachteten Merkmalen mit einer Effektstärke von $d = 0.73$ den zehnten Platz ein und zählt somit zu den stärksten Einflussfaktoren auf die Lernleistung. Auch in der Neuauflage seines Werks mit über 2 100 Metaanalysen stellt Hattie (2023) Feedback mit einer Effektstärke von $d = 0.53$ als ein wichtiges Merkmal für schulisches Lernen heraus. Neben den Zusammenhängen mit der Lernleistung zeigen sich auch Wirkungen von Feedback auf die Motivation und die Metakognition von Schüler*innen (zsf. Narciss, 2004).

Nach Hattie und Timperley (2007) besteht das Hauptziel von Feedback in der Verringerung der Diskrepanz zwischen der momentanen Leistung und dem gewünsch-

ten Lernziel. Gutes Feedback zeichnet sich daher durch seinen Informationsgehalt und seine Motivationsförderlichkeit aus und sollte den Lernenden die drei Fragen nach dem Ziel (Feed-Up), nach den nächsten Schritten (Feed-Forward) und nach dem bisherigen Vorankommen (Feed-Back) beantworten. Dabei sollte es sich nicht nur auf die Aufgabe oder die Lösung, sondern auf den Prozess und die Selbstregulation beziehen, wohingegen Feedback zur Person vermieden werden sollte (Hattie & Timperley, 2007). Feedback sollte somit elaboriert sein und auch zu Lösungsstrategien anregen (Narciss, 2006), wozu bedeutsam ist, dass Lehrkräfte auch die Gedankengänge der Lernenden nachzuvollziehen versuchen, um einschätzen zu können, welche Hilfestellungen gerade benötigt werden. In Anlehnung an Narciss (2006) sowie Hattie und Timperley (2007) systematisieren wir die verschiedenen Formen von Feedback in vier Bereiche: Feedback als Informationsquelle, Feedback als Motivationsquelle, Feedback als Hilfestellung sowie Feedback als Verstehen des Lernenden (für genauere Informationen vgl. Hess, 2019a; Lotz, 2015).

Insbesondere Beobachtungsstudien zeigen aber, dass diese Ansprüche an Feedback im tatsächlichen Unterricht leider kaum umgesetzt werden und Lehrpersonen selten wirklich differenziertes, lernförderliches Feedback erteilen (zsf. Hattie & Timperley, 2007; Lotz, 2015). Um diese Diskrepanz zwischen der hohen Bedeutsamkeit von Feedback für den Lernprozess und der meist oberflächlichen Rückmeldepraxis im tatsächlichen Unterricht zu verringern, ist es wichtig, bereits im Studium die Feedbackkompetenz angehender Lehrpersonen zu fördern.

1.2 Förderung der unterrichtlichen Handlungskompetenz durch die Arbeit mit Videos

Ein wichtiger Aspekt der professionellen Kompetenz von Lehrpersonen ist deren Handeln im Unterricht. Allerdings gibt es bisher nur wenige Studien, die Wirkungen videobasierter Lernumgebungen auf die Handlungskompetenz untersuchen. In einem Überblicksartikel stellen Steffensky und Kleinknecht (2016) dar, dass sich in Studien vereinzelte Effekte videobasierter Trainings auf das Unterrichtshandeln und zum Teil auch auf die Leistungen der Schüler*innen sowie auf deren Engagement im Unterricht zeigen. In einer neueren Literaturübersicht stellen Sablić, Miroslavljević und Škugor (2020) Studien im Bereich des videobasierten Lernens vor, die zwischen 2008 und 2019 Effekte auf Schüler*innen und auf die professionelle Kompetenz von Lehrenden untersucht haben. Im dort aufgeführten STeLLA-Projekt (Roth et al., 2011) zeigte sich beispielsweise, dass Lehrpersonen der Interventionsgruppe ihre unterrichtlichen Fähigkeiten von Prä- zu Posttest signifikant verbesserten (z. B. Einfordern von Begründungen, Stellen herausfordernder Fragen).

In einer Pilotierungsstudie zum ProFee-Projekt untersuchten Hess und Lipowsky (2019), ob sich videobasierte Lernumgebungen auf die Kompetenzen von Lehramts-

studierenden im Bereich Feedback auswirken. Das unterrichtliche Handeln der Studierenden verbesserte sich von Prä- zu Post-Test in allen Bedingungen, am deutlichsten in der am stärksten strukturierten Lernumgebung. Es handelt sich hierbei jedoch um eine deutlich kleinere Stichprobe als im vorliegenden Beitrag, die im Seminar außerdem mit eigenen statt fremden Videos arbeitete und nur in drei statt vier Gruppen aufgeteilt war.

1.3 Förderung der Reflexionsfähigkeit über Unterricht durch die Arbeit mit Videos

Für Lehrpersonen stellt die kritische Wahrnehmung und Reflexion eigenen und fremden Unterrichts eine wichtige Fähigkeit dar, mit der sie ihren Unterricht weiterentwickeln können. Die professionelle Unterrichtswahrnehmung wird meist als eine Fähigkeit dargestellt, die aus den Komponenten *noticing* und *knowledge-based reasoning* besteht (Seidel & Stürmer, 2014; Sherin & van Es, 2009). Die Komponente des *noticing* beschreibt die selektive Wahrnehmung relevanter Ereignisse oder Situationen im Unterricht, während die Komponente des *knowledge-based reasoning* die weitere Beschreibung und Interpretation dieser Situationen umfasst (zsf. Huang, Miller, Cortina & Richter, 2023). Die für diese Interpretation notwendigen Schritte werden häufig in Beschreiben, Erklären und Vorhersagen differenziert (z. B. Seidel & Stürmer, 2014). Zur Messung der professionellen Wahrnehmung gibt es mittlerweile eine Vielzahl an meist videobasierten Instrumenten, die allerdings spezifisch für verschiedene Qualitätsmerkmale des Unterrichts sind. Im Rahmen des ProFee-Projekts wurde ein Instrument zur Erfassung der professionellen Wahrnehmung von Feedbacksituationen entwickelt (Hess & Deistler, 2022).

Bisherige Forschungen zur Förderung der professionellen Wahrnehmung mit Unterrichtsvideos zeigen, dass relevante Situationen besser erkannt werden (Brunvand, 2010), Gründe für Unterrichtsentscheidungen besser nachvollzogen werden (Santagata & Guarino, 2011) und sich der Fokus von oberflächlichen Beschreibungen auf Tiefenstrukturen des Unterrichts verschiebt (Star & Strickland, 2008).

Effekte auf die professionelle Wahrnehmung werden bisher meist über die Analyse von Unterrichtsvideos untersucht, wobei ein Großteil der Studien mit Fremdvideos arbeitet (Holodynski & Meschede, 2022). Grundsätzlich werden der Arbeit mit fremden und eigenen Videos in der Lehrkräftebildung spezifische Vor- und Nachteile zugeschrieben (zsf. Holodynski & Meschede, 2022). Fremde Videos werden oft mit einem kritischeren, analytischen Blick betrachtet (Kleinknecht & Poschinski, 2014), während die Analyse von eigenen Videos besonders motivierend sein kann (Gaudin & Chaliès, 2015) und zu einer tieferen Auseinandersetzung mit dem Inhalt führt („immersion“ und „resonance“, vgl. Seidel, Stürmer, Blomberg, Kobarg & Schwindt, 2011). Besonders förderlich dürfte daher auch eine Kombination sein, in deren Rah-

men zunächst fremde und dann eigene Videos analysiert werden. Hier kann vermutet werden, dass durch die Auseinandersetzung mit fremdem Videomaterial ein distanzierterer Blick auf den Unterricht möglich ist und der Fokus stärker auf Tiefenstrukturen gelegt werden kann (vgl. Kücholl & Lazarides, 2021), welche anschließend für die Analyse von eigenen Videos als Kriterien herangezogen werden können. So könnten unter Umständen selbstwertbezogene Emotionen vermindert werden, die bei der Betrachtung von eigenen Videos die vertiefte Reflexion beeinträchtigen können (Kleinknecht & Poschinski, 2014), was dazu führen könnte, dass die Performanz in eigenen Videos weniger „objektiv“ bewertet wird. Dabei stellt sich die Frage, ob bestimmte strukturelle Merkmale einer Lernumgebung die Reflexion des eigenen Handelns begünstigen können. Da bisherige Studien noch keine systematischen Vergleiche zwischen der Selbst- und Fremdeinschätzung studentischer Handlungskompetenzen im Bereich Feedback vorgenommen haben, wird in der vorliegenden Studie auf die Analyse dieser Zusammenhänge fokussiert.

1.4 Die Rolle der Beobachtungsmethode für die Wirksamkeit von Videoanalysen in der Lehrkräftebildung

Da mit Videos authentische Situationen präsentiert werden können (Sherin, 2004), beurteilen Lehramtsstudierende videobasierte Seminarumgebungen häufig positiver als beispielsweise Erzählungen oder textbasierte Berichte (Choi & Johnson, 2007; Moreno, Abercrombie & Hushman, 2009). Dennoch sind allein aufgrund des Einsatzes von Unterrichtsvideos per se noch keine Lernerfolge zu erwarten, denn „video is but a tool“ (Seago, 2003, S. 263). Für den lernwirksamen Einsatz von Unterrichtsvideos ist also die Lernumgebung eine entscheidende Größe, in der unterschiedliche Variationsmöglichkeiten realisiert werden können (zsf. Syring, 2021): z. B. eigene vs. fremde Videos, instruktionale vs. problembasierte Lehr-Lernmodelle, offene vs. strukturierte Formen von Lernaufgaben oder individuelle vs. kooperative Arbeitsweisen. Um die spezifischen Wirksamkeitsbedingungen des Videoeinsatzes in der Lehrkräftebildung zu ermitteln, wird daher v. a. quasi-experimentelle Forschung gefordert (Steffensky & Kleinknecht, 2016).

Im ProFee-Projekt wurde daher eine systematische Variation in der konkreten Beobachtungsmethode der Videoanalyse und damit einhergehend in der Strukturiertheit der Lernumgebung umgesetzt. Eine frühere Studie mit einem ähnlichen Fokus stammt von Seidel, Blomberg und Renkl (2013), die die Wirksamkeit einer videobasierten deduktiven, durch die Dozentin gelenkten (rule-example) und stark strukturierten Instruktion mit einer videobasierten induktiven Lernumgebung, die stärker durch die Diskussion der Lernenden geprägt war (example-rule), verglichen. Während der stärker gelenkte Ansatz Vorteile für die Entwicklung von Wissen und professioneller Wahrnehmung aufwies, war bei einer Transferaufgabe, der Planung ei-

ner Unterrichtsstunde, der induktive Ansatz im Vorteil. Barth et al. (2019) untersuchen zwei problemorientierte Settings im Vergleich: Die Studierenden erhielten entweder Vorlesungen zum Thema Klassenführung (instruktional problemorientiert) oder erarbeiteten sich das Thema selbst anhand eines Videobeispiels und Literatur (traditionell problemorientiert). Im Post-Test gab es keine Unterschiede zwischen den Gruppen im Erkennen von relevanten Situationen, aber ein leichter Effekt zugunsten des instruktionalen Settings in der Interpretation der relevanten Situationen mit Bezug zu theoretischen Konstrukten.

2 Fragestellungen

Im Fokus der systematischen Seminarevaluation des ProFee-Projekts steht die Frage, wie sich unterschiedlich stark strukturierte Beobachtungsmethoden bei Videoanalysen auf die professionellen Kompetenzen der Studierenden im Bereich Feedback auswirken. Im vorliegenden Beitrag wird auf die Handlungs- und Reflexionskompetenzen der Studierenden als ausgewählte Aspekte ihrer Kompetenzen im Bereich Feedback fokussiert und es werden folgende Fragestellungen untersucht:

- (1) Evaluation der Qualität des Feedbacks von Studierenden an Schüler*innen auf Basis videobasierter Fremdeinschätzungen
 - ▶ Wie geben die Studierenden nach Besuch unterschiedlicher Interventionen einem Kind in einer Interaktionssituation Feedback? Inwiefern gelingt ihnen die Umsetzung guten Feedbacks in der Praxis?
 - ▶ Unterscheidet sich die Qualität des Feedbacks von Studierenden an Schüler*innen auf Basis videobasierter Fremdeinschätzungen in Abhängigkeit von der Strukturiertheit der Beobachtungsmethode, mit der die Studierenden im Seminar vorab Videos analysiert haben?
- (2) Evaluation der Qualität des Feedbacks von Studierenden an Schüler*innen auf Basis videobasierter Selbsteinschätzungen
 - ▶ Wie differenziert und umfassend reflektieren die Studierenden ihr eigenes Feedbackverhalten?
 - ▶ Wie bewerten die Studierenden selbst die Qualität ihres eigenen Feedbackverhaltens?
 - ▶ Unterscheidet sich die Qualität des Feedbacks von Studierenden an Schüler*innen auf Basis videobasierter Selbsteinschätzungen in Abhängigkeit von der Strukturiertheit der Beobachtungsmethode, mit der die Studierenden im Seminar vorab Videos analysiert haben?

- (3) Zusammenhänge zwischen der Selbst- und Fremdeinschätzung der Feedbackqualität
- ▶ Hängen die Reflexionen (Selbsteinschätzungen) mit einem systematischen Rating der Videos durch trainierte Beobachterinnen (Fremdeinschätzungen) zusammen?
 - ▶ Unterscheidet sich die Stärke des Zusammenhangs zwischen der Selbst- und Fremdeinschätzung der Feedbackqualität in Abhängigkeit von der Strukturiertheit der Beobachtungsmethode, mit der die Studierenden im Seminar vorab Videos analysiert haben?

3 Methode

3.1 Datengrundlage: Das Projekt ProFee

Alle Seminare wurden an der Universität Erlangen-Nürnberg (Bayern, Deutschland) durchgeführt. Es handelte sich um Hauptseminare im Fach Grundschulpädagogik, die im Wahlpflichtbereich besucht werden konnten. Das bedeutet, dass alle Grundschullehramtsstudierenden ein Hauptseminar besuchen müssen und dabei das Seminar zum Thema „Feedback“ auswählen konnten. Die Plätze wurden im Rahmen einer zentralen Seminarplatzvergabe zugeteilt, wobei sich Studierende höherer Semester bevorzugt anmelden konnten. Eine weitere Selektion oder systematische Zuteilung durch die Dozierenden fand nicht statt. Die Seminare wurden über mehrere Semester angeboten, wobei jeweils zwei Seminare pro Semester stattfanden. Im ProFee-Projekt (Professionell Feedback geben – Lernen mit Videos) wurden insgesamt vier verschiedene Seminarkonzepte systematisch in ihrer Wirksamkeit verglichen. Eingebettet wurde dies in ein Evaluationsdesign in Form einer quasi-experimentellen Prä-Post-Erhebung mit drei Interventionsgruppen sowie einer Kontrollgruppe. Abb. 1 stellt das Seminar- und Evaluationsdesign im Überblick dar (für genauere Erläuterungen vgl. auch Hess, 2019b). Alle Kurse wurden von derselben Lehrperson gehalten, wobei möglichst alle Aspekte – bis auf die Variation der konkreten Beobachtungsmethode – konstant gehalten wurden.

Zunächst wurden im Rahmen der Seminarvorbesprechung zu den Blockseminaren bereits erste Befragungen (z. B. zu demografischen Merkmalen) sowie erste Prä-Tests durchgeführt. Die Intervention selbst bestand aus jeweils einem dreitägigen Blockseminar (drei direkt aufeinanderfolgende Tage), wobei sich nur Tag 2 zwischen den Interventionsgruppen unterschied. Am ersten Tag erhielten alle Gruppen die gleiche Einführung in Theorie und Forschungsstand zum Thema Feedback und es wurde ein Vortest zur professionellen Wahrnehmung durchgeführt. Am Ende des dritten Tags erfolgten in allen Gruppen die Nacherhebungen.

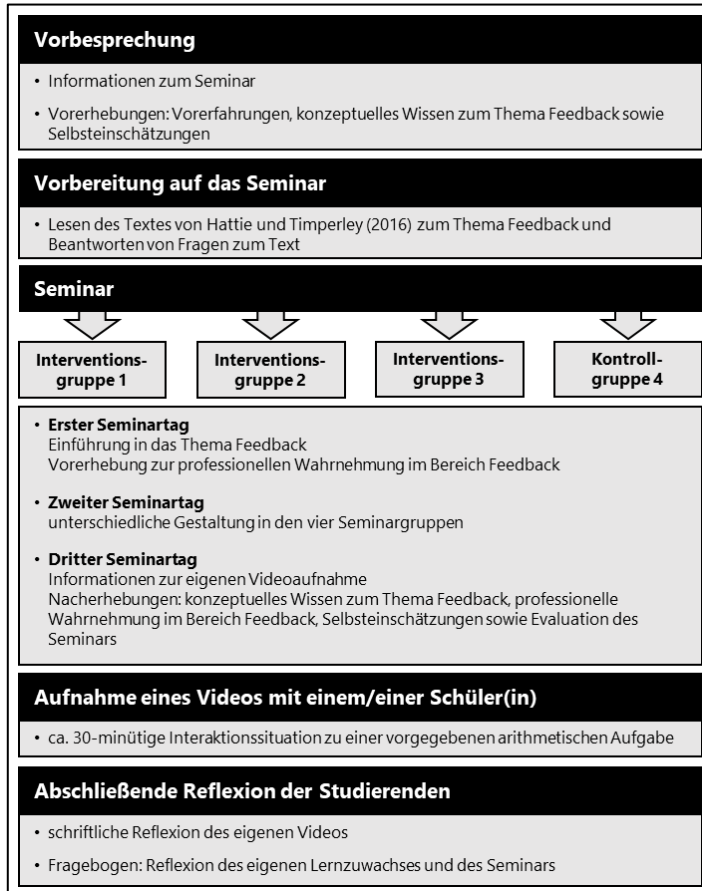


Abbildung 1: Beschreibung des Versuchsablaufs im Projekt ProFee (vgl. Hess, 2019a, S. 473)

Abb. 2 stellt die Unterschiede der vier Gruppen (Seminare) im Überblick dar. In den drei Versuchsgruppen (Kurs 1, 2 und 3) analysierten die Grundschullehramtsstudierenden jeweils Videos von Interaktionssituationen (Lehrer*in und Schüler*in) anhand unterschiedlich stark strukturierter Beobachtungsaufträge. Die vierte Seminargruppe (Kontrollgruppe) analysierte keine Videos (für genauere Informationen vgl. Abb. 2).

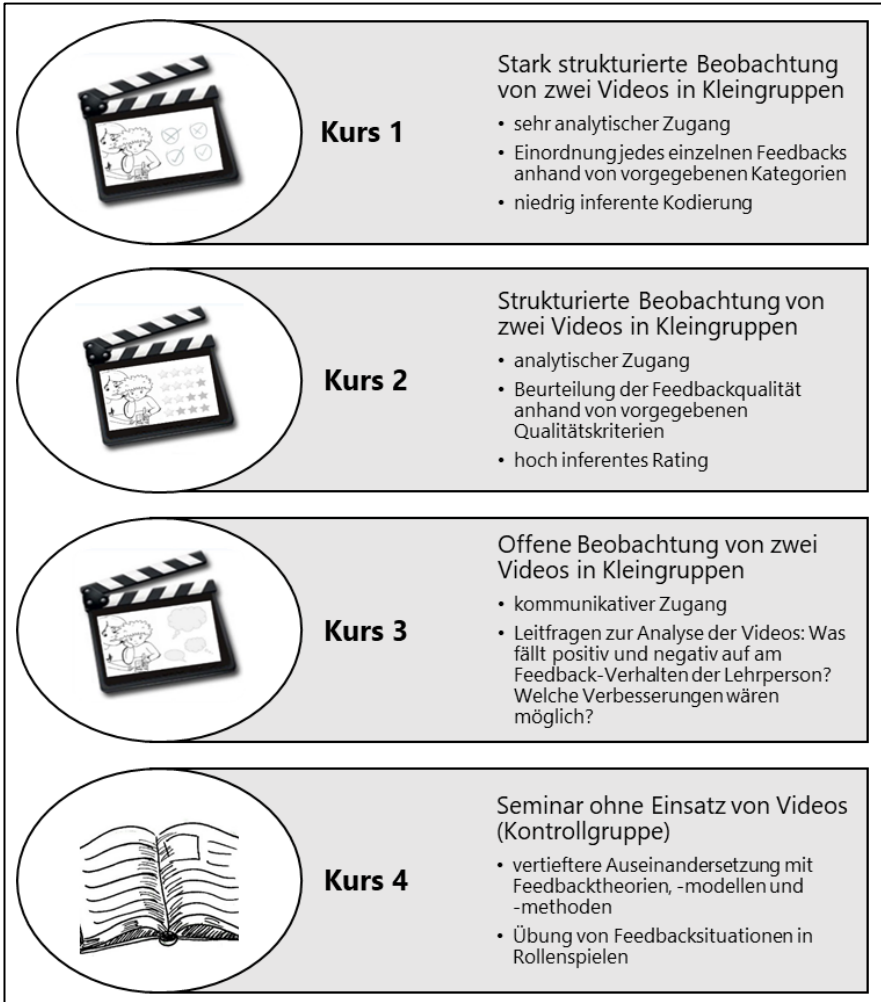


Abbildung 2: Die Interventionsgruppen (Kurs 1, 2 und 3) und die Kontrollgruppe (Kurs 4) im Projekt ProFee

Im Anschluss an das Seminar videografierten sich die Studierenden in einer etwa 30-minütigen Interaktionssituation mit einem einzelnen Kind bei der Arbeit an einer vorgegeben Mathematikaufgabe. Dieses Video sollten die Studierenden anschließend selbst schriftlich hinsichtlich der Qualität des von ihnen erteilten Feedbacks reflektieren.

3.2 Kontrollvariablen

Zur Anzahl der Fachsemester im Studiengang Grundschullehramt sowie zum Geschlecht wurden die Studierenden mittels Fragebogen im Rahmen der Seminarvorbereitung befragt.

Die Fähigkeit der Studierenden, die Qualität von Feedback fremder Lehrpersonen adäquat zu erkennen und zu beurteilen wurde zu Beginn des Seminars mit einem selbst entwickelten videobasierten Instrument erhoben (für genauere Informationen vgl. Hess & Deistler, 2022). Die Studierenden sollten dazu im Vor- und Nachtest das Feedbackverhalten einer Lehrerin in einer Interaktionssituation mit einem Drittklässler bei der Bearbeitung einer Mathematikaufgabe zum Thema Zahlenmauern beurteilen und dazu das zugehörige Transkript systematisch kommentieren. Die Kommentare der Studierenden wurden von geschulten Beurteilerinnen ausgewertet und hoch inferent in ihrer Qualität eingeschätzt. Auf diese Weise wurden der Gesamteindruck, die Qualität der Beobachtungen, die Qualität der Bewertungen und Begründungen, die Qualität der Verbesserungsvorschläge und die Qualität der theoretischen Bezüge eingeschätzt. Als Kontrollvariable für die Analysen dieses Beitrags wird der Gesamteindruck (Skala von 1 bis 6) der geschulten Beurteilerinnen verwendet (für genauere Informationen vgl. Hess & Deistler, 2022).

3.3 Videobasierte Fremdeinschätzung der Handlungskompetenz im Bereich Feedback

Im Anschluss an den Besuch des Seminars zeichneten die 226 Studierenden je ein Video von sich auf, in dem sie im Rahmen einer Microteaching-Situation mit einem einzelnen Kind an einer komplexen Mathematikaufgabe (Subtraktion von Spiegelzahlen inklusive Entdecken und Begründen des Ergebnismusters) arbeiteten und dabei möglichst gutes Feedback geben sollten.

Zur Auswertung dieser Videos wurde auf Basis theoretischer und empirischer Grundlagen zur Qualität von Feedback ein Beobachtungsmanual (Hess, 2019a) entwickelt, mit dessen Hilfe trainierte Beobachterinnen jedes Video unabhängig voneinander auswerteten. Das hoch inferente Ratingsystem besteht aus insgesamt 21 Items zur Qualität des Feedbacks, die jeweils 6-stufig bewertet wurden (vgl. Abb. 3). Je höher der Wert, desto positiver ist das Item ausgeprägt. Dabei wird in Anlehnung an Angebots-Nutzungsmodelle der Unterrichtsqualität (vgl. z. B. Seidel, 2014) mit den Items 1 bis 16 die Qualität des Feedbacks auf Angebotsseite differenziert betrachtet, wobei die Items auf Basis theoretischer Modelle (z. B. Hattie & Timperley, 2007; Narciss, 2006, ausführlich vgl. Hess, 2019a) thematisch den vier Bereichen Feedback als A) Motivationsquelle, B) Informationsquelle, C) Hilfestellung und D) Verstehen des Lernenden zugeordnet wurden. Die Items 17 und 18 fokussieren hingegen auf die beobachtete Wirkung des Feedbacks auf das Kind, indem der Nutzen

des Feedbacks sowie die Förderung des (mathematischen) Verständnisses beurteilt wurden. Zusätzlich zur differenzierten Beschreibung des Feedbacks durch die Items 1 bis 18 wurden drei Globalurteile durch die Beobachterinnen abgegeben: zur Lernförderlichkeit (Item 19) und zur Motivationsförderlichkeit (Item 20) des Feedbacks sowie zur generellen Qualität des Feedbacks (Item 21).

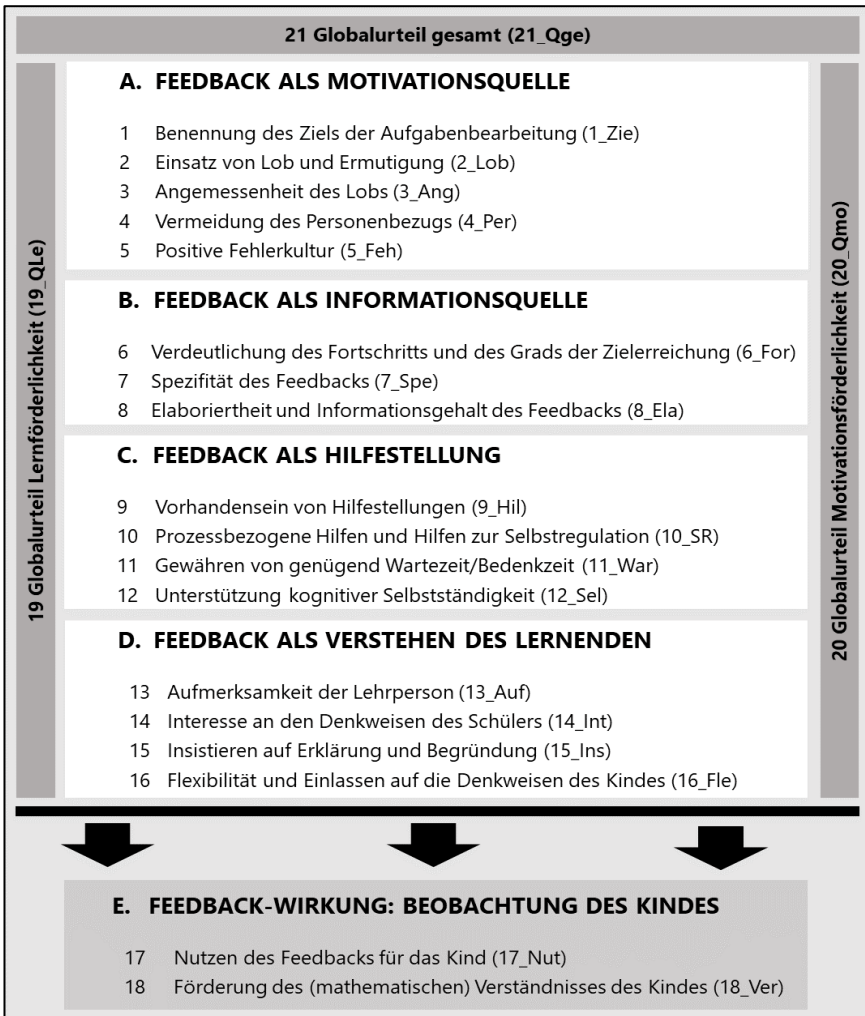


Abbildung 3: Items des hoch inferenten Beobachtungssystems zur videobasierten Fremdeinschätzung

Damit liegen die einzelnen Items auf durchaus unterschiedlichen Ebenen der Differenziertheit und Aggregation, was bei der Interpretation der Ergebnisse berücksichtigt werden muss. Für die Analysen dieses Beitrags werden gezielt die einzelnen Items für die Analysen genutzt anstatt faktorenanalytisch zusammengefasster Skalen, da so ein direkter Vergleich zwischen Fremd- und Selbsteinschätzung auf Itemebene möglich ist.

Die Interrater-Reliabilität wurde fortlaufend überprüft und der relative Generalisierbarkeitskoeffizient erreichte mit $g_M = .83$ ($g_{Min} = .73$; $g_{Max} = .92$) zufriedenstellende Werte. Für die weiteren Analysen wurde der Mittelwert der Beobachterinnen genutzt.

3.4 Videobasierte Selbsteinschätzung der Handlungskompetenz im Bereich Feedback

Alle 226 Studierenden hatten im Anschluss an die Videografie die Aufgabe, ihr Feedbackverhalten im Video schriftlich zu reflektieren: „Reflektieren Sie Ihr Feedbackverhalten und schreiben Sie Ihre Überlegungen auf. Was gelang gut? Was war schwierig? Was hätten Sie gerne anders gemacht? Verknüpfen Sie Ihre Überlegungen mit den Theorien zu gutem Feedback und analysieren Sie auch einzelne Szenen.“ Den Studierenden wurden die Kriterien zur Analyse ihrer eigenen Videos bewusst nicht vorgegeben, da es bei der Auswertung auch darum gehen sollte, wie differenziert und umfassend sie selbst auf ihr Feedback eingehen. Vorab bekannt waren den Studierenden aber die Bewertungskriterien für die Reflexionen: Herstellung von Theoriebezügen, Eingehen auf Stärken und Schwächen des eigenen Feedbackverhaltens, Analyse einzelner Feedbacksituationen, Formulierung konkreter Verbesserungsvorschläge. Der Mindestumfang sollte drei Textseiten betragen. Abb. 4 stellt einen exemplarischen Ausschnitt einer studentischen Reflexion dar.

Zur Auswertung der Reflexionen wurde ein Manual entwickelt, das eng an das Manual zur Fremdeinschätzung der Handlungskompetenz angelehnt war und durch das erhoben werden sollte, auf wie viele und welche Merkmale guten Feedbacks die Studierenden in ihren Reflexionen eingehen und wie sie ihr eigenes Feedback in diesen Merkmalen einschätzen. Daher wurde für jedes Item, das auch von den trainierten Beobachterinnen beurteilt wurde, anhand von vier Ratingstufen bestimmt, ob von den Studierenden auf das jeweilige Item eingegangen wurde und wenn ja, ob es (eher) positiv oder (eher) negativ oder aber nicht oder uneindeutig in seiner Qualität bewertet wurde (vgl. Tab. 1). Durch diese aufeinander abgestimmte Auswertung von Videos und Reflexionen ist ein systematischer Vergleich zwischen den Selbst- und Fremdeinschätzungen der Handlungskompetenzen im Bereich Feedback möglich.

Während meiner Lernstunde wurde meinerseits allerdings nicht ausreichend und vor allem nicht begründet gelobt. An einigen Stellen wäre ein motivierenderes Lob angebracht gewesen als „genau“ oder „gut“. Das generell authentischere Lob „Perfekt“ in Minute (4:38) gab es jedoch an der falschen Stelle. Die Schülerin wurde dafür gelobt, Minusaufgaben aufgeschrieben zu haben, was keine sehr anspruchsvolle Aufgabe ist. Besser wäre dieses Feedback zu geben, nachdem die Schülerin zum Beispiel herausgefunden hat, dass die Ergebnisse alle aus der Neunerreihe kommen. Hier wurde die Schülerantwort nur mit einem „genau“ (Minute 9:15) bestätigt. Außerdem wäre es erst ein gutes Feedback, wenn man dazu sagt, was genau „perfekt“ war und somit das Lob begründet. Ein solches begründetes Lob fand in keinem der Fälle statt. Motivierend war das Feedbackverhalten jedoch in dem Sinne, dass die Schülerin viel selbst handeln und überlegen durfte. Wobei auch hier noch mehr die Schülerin übernehmen hätte können. In Minute 36:00

Abbildung 4: Ausschnitt aus einer exemplarischen Reflexion einer Studentin

Ähnlich wie die Videos wurde auch jede Reflexion unabhängig von zwei geschulten Beurteilerinnen (andere Personen als die Beobachterinnen der Videos) ausgewertet, die Beobachterübereinstimmungen wurden fortlaufend geprüft (prozentuale Übereinstimmung PÜ \geq 85 %) und im Falle einzelner Nicht-Übereinstimmungen wurde ein konsensuales Urteil für die weiteren Analysen verwendet.

Tabelle 1: Ratingstufen für ein exemplarisches Item („Positive Fehlerkultur“) zur Auswertung der studentischen Reflexionen

Ratingstufe	Beschreibung der Ratingstufe
0	Der/Die Student*in geht in der Reflexion nicht darauf ein, ob er/sie implizit oder explizit deutlich gemacht hat, dass Fehler nicht schlimm sind oder sogar hilfreich für das Lernen sein können.
1	Der/Die Student*in geht in der Reflexion darauf ein, ob er/sie implizit oder explizit deutlich gemacht hat, dass Fehler nicht schlimm sind oder sogar hilfreich für das Lernen sein können. Dabei ist keine eindeutig positive oder negative Wertung erkennbar.
2	Der/Die Student*in geht in der Reflexion darauf ein, ob er/sie implizit oder explizit deutlich gemacht hat, dass Fehler nicht schlimm sind oder sogar hilfreich für das Lernen sein können. Der/Die Student*in bewertet diesen Aspekt eher positiv.
3	Der/Die Student*in geht in der Reflexion darauf ein, ob er/sie implizit oder explizit deutlich gemacht hat, dass Fehler nicht schlimm sind oder sogar hilfreich für das Lernen sein können. Der/Die Student*in bewertet diesen Aspekt eher negativ.

3.5 Stichprobe und Analysemethoden

Die Stichprobe besteht aus insgesamt 226 Studierenden des Grundschullehramts, zu denen in Tab. 2 nähere Informationen zu finden sind, sowohl insgesamt als auch spezifisch für die vier Gruppen. Alle Studierenden haben zum Zeitpunkt des Seminarbesuchs bereits mindestens ein Praktikum besucht. Signifikante Unterschiede zwischen den Gruppen bestehen nur in der Semesterhöhe der teilnehmenden Studierenden ($F = 5.98$; $p \leq .001^{***}$; $\eta^2 = .07$) und in den Prä-Test-Werten zur professionellen Wahrnehmung ($F = 2.96$; $p = .033^*$; $\eta^2 = .04$).

Zur Beantwortung der Fragestellungen werden deskriptive Analysen sowie multivariate Varianzanalysen und partielle Korrelationen jeweils unter Kontrolle der drei Kovariaten Geschlecht, Semesterhöhe sowie den Vortestwerten zur professionellen Wahrnehmung genutzt, da sich diese Variablen in vorangehenden Analysen als bedeutsam für die Handlungs- und Reflexionskompetenz der Studierenden erwiesen haben. Fehlende Werte liegen nicht vor.

Tabelle 2: Stichprobenbeschreibung

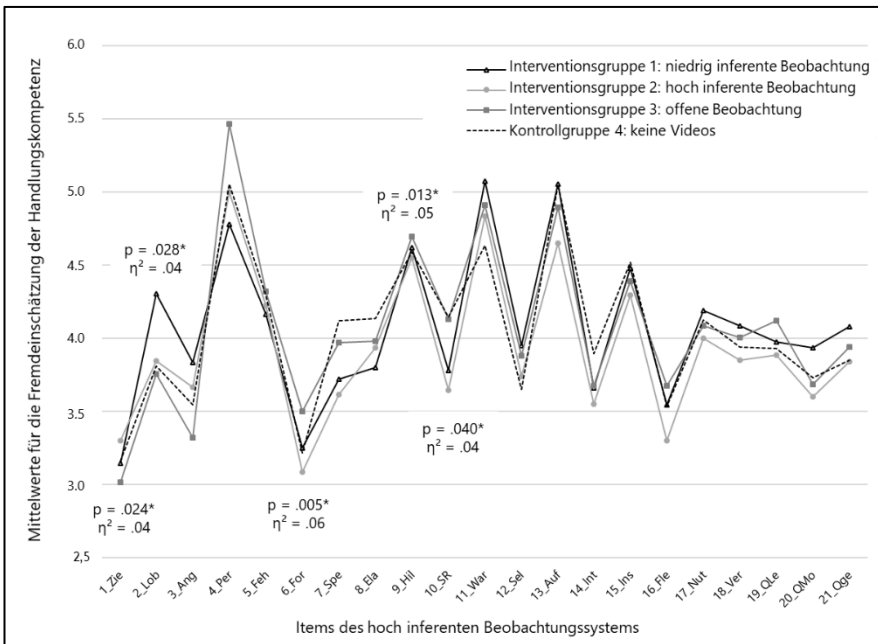
	Studierende in Kursen				
	gesamt	Kurs 1	Kurs 2	Kurs 3	Kurs 4
Anzahl der Studierenden	226	58	58	55	55
prozentualer Anteil an Studentinnen	88	91	84	84	93
durchschnittliche Anzahl Semester	4.04	3.54	3.68	4.65	4.34
Anteil Studierender, die regelmäßig mit Kindern zusammenarbeiten	75	83	78	69	69
Anteil Studierender, die vorab noch nie mit Unterrichtsvideos gearbeitet hatten	79	78	76	87	74
durchschnittlicher Prä-Test-Wert: professionelle Wahrnehmung (Skala: 1–6)	2.30	2.43	2.05	2.48	2.27

4 Ergebnisse

4.1 Videobasierte Fremdeinschätzung der Qualität des Feedbacks

In Abb. 5 sind die Mittelwerte für die einzelnen Items in den vier Gruppen abgebildet. Es fällt auf, dass sich das Muster zwischen den Gruppen insgesamt nur wenig

unterscheidet: Die Studierenden setzen also eher gruppenübergreifend bestimmte Merkmale guten Feedbacks gut um, während andere Merkmale kritischer beurteilt werden. Am besten bewertet werden die Items 4 (Vermeidung des Personenbezugs), 9 (Vorhandensein von Hilfestellungen), 11 (Gewähren von genügend Wartezeit/Bedenkzeit) und 13 (Aufmerksamkeit der Lehrperson). Im Mittel eher negativ (allerdings immer noch mit vergleichsweise positiven Werten auf der sechsstufigen Skala) schneiden hingegen die Items 1 (Benennung des Ziels der Aufgabenbearbeitung), 6 (Verdeutlichung des Fortschritts und des Grads der Zielerreichung) und 16 (Flexibilität und Einlassen auf die Denkweisen des Kindes) ab.



Anmerkungen: * $p \leq .05$; η^2 = Eta-Quadrat; Aus dem sechsstufigen Ratingsystem wird nur der obere Bereich abgebildet, da alle Items im Mittel mit mindestens mit $M = 3$ bewertet wurden

Abbildung 5: Mittelwerte für die Fremdeinschätzung der Handlungskompetenz (vollständige Itembezeichnungen vgl. Abb. 3)

Bei einer multivariaten Varianzanalyse mit den einzelnen Items der Videoanalyse als abhängige Variablen ergeben sich zunächst signifikant positive Effekte der drei Kovariaten: Studentinnen ($F = 1.65$; $p = .041^*$; $\eta^2 = .15$), Studierende in höheren Semestern ($F = 1.64$; $p = .044^*$; $\eta^2 = .15$) sowie Studierende mit höheren Werten in der professionellen Wahrnehmung im Vortest ($F = 2.71$; $p \leq .001^{***}$; $\eta^2 = .23$) erzielen

in der fremdeingeschätzten Handlungskompetenz bessere Werte. Die Handlungskompetenz unterscheidet sich zwar signifikant zwischen den Gruppen ($F = 1.77$; $p \leq .001^{***}$; $\eta^2 = .16$), allerdings ist die Effektstärke eher gering und die Unterschiede gehen auf nur fünf der insgesamt 21 Items zurück, wobei, wie Abb. 5 zeigt, auch bei diesen fünf Items nicht ein Kurs den anderen systematisch überlegen ist.

4.2 Videobasierte Selbsteinschätzung der Qualität des Feedbacks

In Tab. 3 werden zunächst die deskriptiven Ergebnisse zur Auswertung der studentischen Reflexionen dargestellt. Die Studierenden gehen in ihren Reflexionen im Mittel auf 14 von 18 möglichen Merkmalen guten Feedbacks ein, wobei es auch Studierende gibt, die nur acht Merkmale erwähnt haben, während andere auf alle 18 möglichen Merkmale eingehen.

Von den erwähnten Items wurden durchschnittlich 15 % nicht eindeutig als positiv oder negativ bewertet, 60 % als positiv und 25 % als negativ. Auch die Studierenden selbst beurteilen ihr Feedback also – ähnlich wie die externen Beurteilerinnen – insgesamt eher positiv. Allerdings gibt es auch hier deutliche interindividuelle Unterschiede. Während einige Studierende jedes von ihnen erwähnte Merkmal positiv oder eher positiv bewerten, bewerten andere bis zu 77 % der erwähnten Merkmale (eher) negativ (vgl. Tab. 3).

Für die Selbsteinschätzungen zeigen sich unter Kontrolle der Kovariaten Geschlecht, Semester und professionelle Wahrnehmung signifikante Unterschiede zwischen den Gruppen in der Anzahl der in den Reflexionen genannten Merkmale ($F = 4.12$; $p = .007^{**}$; $\eta^2 = .06$) sowie im prozentualen Anteil der positiv bewerteten Merkmale ($F = 2.95$; $p = .034^*$; $\eta^2 = .03$), wobei die Effektstärken hier sehr gering ausfallen (vgl. Tab. 3). Dabei ist auffällig, dass der Anteil positiv bewerteter Merkmale in der Kontrollgruppe am geringsten ausfällt, während in den Interventionsgruppen 1 (niedrig inferente Kodierung) und 2 (hoch inferentes Rating) die studentischen Reflexionen auf etwas mehr Merkmale guten Feedbacks eingehen als die Interventionsgruppe 3 (offene Beobachtung) und die Kontrollgruppe.

Tabelle 3: Deskriptive Statistiken zu den Reflexionen und Vergleich der Gruppen

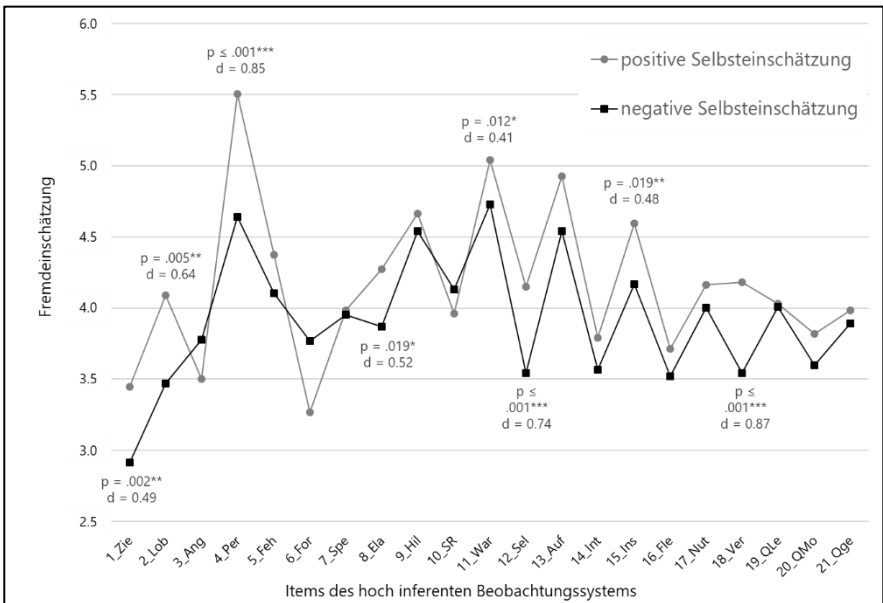
	Min	Max	M	SD	F	p	η^2	
Wie viele Merkmale guten Feedbacks haben die Studierenden in der Reflexion erwähnt? (Anzahl)	gesamt N = 8	N = 18	13.95	2.43				
	Kurs 1 N = 8	N = 18	14.10	2.43				
	Kurs 2 N = 8	N = 18	14.66	2.62	4.12	.007**	.06	
	Kurs 3 N = 8	N = 18	13.47	2.29				
	Kurs 4 N = 9	N = 18	13.51	2.20				
Wieviel Prozent der erwähnten Items enthalten keine eindeutige Bewertung? (prozentualer Anteil)	gesamt	0.00 %	60.00 %	14.47 %	10.72 %			
	Kurs 1	0.00 %	44.44 %	13.51 %	10.01 %			
	Kurs 2	0.00 %	60.00 %	15.66 %	12.72 %	1.38	.249	.02
	Kurs 3	0.00 %	35.29 %	13.23 %	10.18 %			
	Kurs 4	0.00 %	36.84 %	15.48 %	9.65 %			
Wieviel Prozent der erwähnten Items werden eher positiv bewertet? (prozentualer Anteil)	gesamt	6.67 %	100.00 %	60.21 %	18.47 %			
	Kurs 1	28.57 %	93.75 %	64.01 %	16.83 %			
	Kurs 2	6.67 %	100.00 %	58.75 %	18.96 %	2.95	.034*	.04
	Kurs 3	7.69 %	100.00 %	62.50 %	19.62 %			
	Kurs 4	12.50 %	83.33 %	55.45 %	17.65 %			
Wieviel Prozent der erwähnten Items werden eher negativ bewertet? (prozentualer Anteil)	gesamt	0.00 %	76.92 %	25.32 %	14.73 %			
	Kurs 1	0.00 %	61.90 %	22.48 %	12.48 %			
	Kurs 2	0.00 %	70.59 %	25.58 %	14.49 %	2.13	.097	.03
	Kurs 3	0.00 %	76.92 %	24.27 %	16.86 %			
	Kurs 4	0.00 %	64.71 %	29.07 %	14.48 %			

Anmerkungen: Min = Minimalwert; Max = Maximalwert; M = arithmetisches Mittel; SD = Standardabweichung; F = F-Wert; η^2 = Eta-Quadrat; ** $p \leq .01$; * $p \leq .05$

4.3 Zusammenhänge zwischen der Selbst- und Fremdeinschätzung

Abb. 6 stellt auf der y-Achse die Mittelwerte in der Fremdeinschätzung dar. Die graue Linie zeigt die Mittelwerte der externen Beobachterinnen für diejenigen Studierenden an, die sich selbst im jeweiligen Item (eher) positiv eingeschätzt haben (vgl. Tabelle 3: Wert 2 = (eher) positive Bewertung des Items in der Selbsteinschätzung), die schwarze Linie für Studierende, die das jeweilige Item in ihrer Reflexion selbst als (eher) negativ beurteilt haben (vgl. Tabelle 3: Wert 3 = (eher) negative Bewertung des Items in der Selbsteinschätzung). Items, die von den Studierenden nicht

in die Reflexion einbezogen (Wert 0; vgl. Tabelle 3) oder uneindeutig bewertet wurden (Wert 1; vgl. Tabelle 3), werden hier nicht berücksichtigt. In einigen Items sind auf deskriptiver Ebene deutliche Unterschiede in der erwarteten Richtung erkennbar. Beispielsweise erreichen Studierende, die sich selbst im Item 4 (Vermeidung von Personenbezug) positiv einschätzen, auch in der Fremdbewertung sehr gute Werte von knapp $M = 5.6$. Hingegen werden die Studierenden, die sich hier selbst negativ einschätzen auch von den externen Beobachterinnen in diesem Item als schlechter bewertet, wobei auch der Wert von $M = 4.6$ noch im oberen Bereich der Skala liegt. Signifikant sind die Unterschiede bei insgesamt 8 der 21 Items (vgl. Abb. 6). Es gibt allerdings auch viele Items, in denen sich Studierende mit positiver Selbsteinschätzung in der Fremdeinschätzung kaum von Studierenden mit negativer Selbsteinschätzung unterscheiden: Besonders deutlich ist dies beispielsweise bei Item 7 (Spezifität des Feedbacks) und Item 19 (Globalurteil: Lernförderlichkeit des Feedbacks).



Anmerkungen: $*** p \leq .001$; $** p \leq .01$; $* p \leq .05$; $d =$ Effektstärke; Aus dem sechsstufigen Ratingsystem wird nur der derjenige Ausschnitt abgebildet, in dem sich die Mittelwerte bewegen

Abbildung 6: Mittelwerte für die Fremdeinschätzung der Handlungskompetenz getrennt für Studierende mit positiver und negativer Selbsteinschätzung (vollständige Itembezeichnungen vgl. Abb. 3)

Tabelle 4: Partielle Korrelationen zwischen Selbst- und Fremdeinschätzung unter Kontrolle von Geschlecht, Semester und professioneller Wahrnehmung

		Fremdeinschätzung			
		Lernförderlichkeit (Item 19)	Motivationsqualität (Item 20)	Globalurteil (Item 21)	Summenwert aller Items
Anzahl genannter Merkmale	gesamt	r = .13+ p = .062	r = .19** p = .006	r = .19** p = .004	r = .20** p = .003
	Kurs 1	r = .21 p = .120	r = .27** p = .046	r = .25+ p = .073	r = .26 p = .058
	Kurs 2	r = .07 p = .638	r = .21 p = .137	r = .14 p = .323	r = .25+ p = .072
	Kurs 3	r = -.12 p = .418	r = -.02 p = .875	r = .07 p = .611	r = .04 p = .764
	Kurs 4	r = .25+ p = .082	r = .19 p = .177	r = .16 p = .262	r = .16 p = .254
Selbsteinschätzung	gesamt	r = .18** p = .007	r = .19** p = .004	r = .23*** p ≤ .001	r = .27*** p ≤ .001
	Kurs 1	r = .32* p = .018	r = .28* p = .042	r = .33* p = .016	r = .35** p = .009
	Kurs 2	r = -.01 p = .963	r = .14 p = .314	r = .12 p = .382	r = .21 p = .130
	Kurs 3	r = .14 p = .332	r = .17 p = .231	r = .24+ p = .087	r = .22 p = .115
	Kurs 4	r = .20 p = .155	r = .12 p = .401	r = .14 p = .320	r = .27+ p = .054
Anteil (eher) positiv bewerteter Merkmale	gesamt	r = .13+ p = .060	r = .11+ p = .100	r = .15* p = .024	r = .19** p = .005
	Kurs 1	r = .24+ p = .077	r = .17 p = .228	r = .23 p = .102	r = .25+ p = .067
	Kurs 2	r = -.06 p = .676	r = .04 p = .758	r = .06 p = .655	r = .11 p = .416
	Kurs 3	r = .17 p = .242	r = .17 p = .231	r = .20 p = .163	r = .20 p = .165
	Kurs 4	r = .08 p = .601	r = .02 p = .904	r = .07 p = .633	r = .19 p = .181

Anmerkungen: r = Koeffizient für die partielle Korrelation; *** p ≤ .001; ** p ≤ .01; * p ≤ .05; + p ≤ .1

Zusätzlich zum itemspezifischen Vergleich in Abb. 6 wurden partielle Korrelationen zwischen der Selbst- und Fremdeinschätzung unter Kontrolle der drei Kovariaten Geschlecht, Semesterhöhe und professionelle Wahrnehmung berechnet, sowohl für die Gesamtstichprobe als auch für die einzelnen Gruppen (vgl. Tab. 4). Anhand der partiellen Korrelationen für die Gesamtstichprobe (abgebildet jeweils in der oberen Zeile) wird deutlich: Je positiver das Feedback von den externen Beobachterinnen bewertet wurde, desto mehr Merkmale guten Feedbacks nennen auch die Studierenden selbst, desto mehr Merkmale beurteilen die Studierenden selbst positiv und desto höher ist auch der prozentuale Anteil der selbst als positiv beurteilten Merkmale – auch wenn hier die Korrelationen etwas niedriger ausfallen als bei den beiden oberen Items. Dabei sind die Zusammenhänge zwischen Selbst- und Fremdeinschätzung in Gruppe 1 jeweils enger als in den übrigen Gruppen: Studierende, die im Seminar die Videos fremder Studierender sehr strukturiert, niedrig inferent analysiert haben, beurteilen ihre eigenen Videos also differenzierter und weisen stärkere Ähnlichkeiten zur Fremdeinschätzung auf.

5 Zusammenfassung und Diskussion

Evaluationen stellen eine Form anwendungsbezogener Forschung dar, deren Ergebnisse in möglichst kurzer Zeit handlungsrelevant und verwertbar sein sollen (Balzer & Beywl, 2018). Obwohl Unterrichtsvideoanalysen in der Lehrkräftebildung häufig eingesetzt werden, wird bei der Evaluation videobasierter Lernumgebungen häufig auf die Entwicklung des Wissens und die professionelle Wahrnehmung fokussiert, selten zusätzlich auf die eigenen unterrichtlichen Handlungskompetenzen. Auch Vergleiche zwischen Selbst- und Fremdeinschätzungen werden selten systematisch vorgenommen.

Hier setzte die vorliegende Studie an und zeigte zunächst, dass die Handlungskompetenz der Studierenden nach Besuch der Seminare von den geschulten Beobachterinnen insgesamt eher positiv eingeschätzt wird. Dies passt zu den Ergebnissen von Roth et al. (2011), die in ihrer Interventionsstudie ebenfalls zeigen konnten, dass sich die teilnehmenden Lehrpersonen in ihrem Unterrichtshandeln generell verbesserten und von den externen Beobachter*innen hohe Werte in der kognitiven Aktivierung erhielten.

In den Ergebnissen der vorliegenden Studie weisen die Profillinien der mittleren Stärken und Schwächen zwischen den Gruppen insgesamt recht ähnliche Verläufe auf. Dies deutet darauf hin, dass es Facetten guten Feedbacks gibt, die sich leichter/schneller erlernen lassen als andere oder die grundsätzlich einfacher umsetzbar sind. Es könnte sich um Merkmale handeln, deren Bedeutung den Studierenden vor Besuch des Seminars gegebenenfalls nicht bewusst war und die gleichzeitig wenig Übung oder Routinen erfordern. Beim konkreten Blick auf die sehr positiv beurteil-

ten Aspekte 4) „Vermeidung des Personenbezugs“, 9) „Vorhandensein von Hilfestellungen“, 11) „Gewähren von genügend Wartezeit/Bedenkzeit“ und 13) „Aufmerksamkeit der Lehrperson“ erscheint diese Vermutung plausibel. Alle Aspekte dürften eines grundlegenden Bewusstseins über deren grundsätzliche Bedeutung bedürfen, können dann aber vermutlich vergleichsweise einfach umgesetzt werden. Allerdings wäre es ebenso möglich, dass diese Merkmale auch ohne die Intervention von den Studierenden bereits vergleichsweise gut umgesetzt worden wären. Dafür spricht, dass das Thema Feedback seit der Hattie-Studie (2009) eine breite Anwendung in der Lehrkräftebildung findet und die Kategorien des Feedbackmodells nach Hattie und Timperley (2007; vgl. Abschnitt 1.1) vielfach eingesetzt werden (vgl. auch Hattie, 2023). Es könnte demnach sein, dass die Studierenden in anderen Seminaren bereits über die Merkmale guten Feedbacks informiert wurden, diese aber vielleicht noch nicht selbst konkret umsetzen mussten. Zur Absicherung der Ergebnisse müsste daher die vorliegende Evaluation um einen Prä-Test zur Handlungskompetenz ergänzt werden. In der bereits in Abschnitt 3 dargestellten Pilotierungsstudie zum ProFee-Projekt (Hess & Lipowsky, 2019), in der die Handlungskompetenz im Vor- und Nachtest erhoben wurde (allerdings mit einer kleineren Stichprobe von $N = 54$ Lehramtsstudierenden), ergaben sich signifikante Verbesserungen von Vor- zu Nachtest für die Items 4 (Vermeidung des Personenbezugs) und 11 (Unterstützung kognitiver Selbstständigkeit), nicht aber für 9 (Vorhandensein von Hilfestellungen) und 13 (Aufmerksamkeit der Lehrperson). Ergebnisse aus einer Vorstudie des ProFee-Projekts zum konzeptuellen Vorwissen der Studierenden scheinen allerdings die oben angeführte Interpretation zu stützen. In einer offenen Erhebung zum Vorwissen von Studierenden zu Merkmalen guten Feedbacks wurde das Gewähren von genügend Wartezeit nur von 2 % und die Vermeidung des Personenbezugs nur von 19 % der Studierenden genannt (Hess, Werker & Lipowsky, 2017), also tatsächlich war die Bedeutsamkeit dieser Merkmale nur wenigen Studierenden bewusst.

Die Auswertung der studentischen Reflexionen zeigt, dass die Studierenden ihr Feedbackverhalten facettenreich analysieren und sich im Mittel auch eher positiv einschätzen, wobei hier deutliche interindividuelle Unterschiede auftreten. Dabei fallen die Unterschiede zwischen den Gruppen zwar signifikant aus, weisen aber nur sehr geringe Effektstärken auf. Es deuten sich hier Vorteile der Interventionsgruppe 1 an, in der die Videoanalysen niedrig inferent durchgeführt wurden. Der Vergleich von Selbst- und Fremdeinschätzung verdeutlicht, dass Studierende, deren Feedback von den externen Beobachterinnen positiv bewertet wurde, auch selbst mehr Merkmale ihres eigenen Feedbackverhaltens positiv beurteilen. Besonders aufschlussreich ist hier, dass die Zusammenhänge zwischen Selbst- und Fremdeinschätzung in Gruppe 1 enger ausfallen als in den übrigen Gruppen. Dies zeigt, dass Studierende, die im Seminar die Videos fremder Studierender sehr strukturiert mit

einer niedrig inferenten Beobachtungsmethode analysiert haben, ihre eigenen Videos im Anschluss an das Seminar differenzierter und umfassender beurteilen und stärkere Ähnlichkeiten zur Fremdeinschätzung aufweisen. Auch in den Studien, die in Abschnitt 5 dargestellt wurden, deuteten sich Vorteile für eher strukturierte, gelenkte Settings bei der Videoanalyse an (vgl. Seidel et al., 2013).

Anhand der Zusammenhänge zwischen Selbst- und Fremdeinschätzung zeigt sich, dass die Reflexionsfähigkeit bezogen auf eigene Unterrichtsversuche nach Besuch der Seminare insgesamt positiv ausgeprägt ist. Angesichts der insgesamt lediglich mittelhohen Korrelationen zwischen Selbst- und Fremdeinschätzung (vgl. Tabelle 4) könnte man allerdings fragen, warum die Zusammenhänge nicht generell höher ausfallen. Eine Möglichkeit zur Erhöhung der Kongruenz wäre das Angebot weiterer Vergleichsmöglichkeiten, z. B. durch die systematische Reflexion mehrerer eigener oder fremder Videos. Zusätzlich aufschlussreich könnte die Untersuchung des Einflusses persönlicher Merkmale auf die Selbsteinschätzung und den Zusammenhang zwischen Selbst- und Fremdeinschätzung sein. Zudem könnten Beobachtungstendenzen bei der Bewertung eine Rolle spielen und in den Fokus weiterer Studien gerückt werden. Folgestudien könnten außerdem die Reflexionsrichtlinien vorgeben, damit Studierende ihre Selbsteinschätzung anhand derselben Kriterien wie die trainierten Beobachterinnen vornehmen können. Und auch für den Vergleich zwischen Selbst- und Fremdeinschätzung wäre vor allem die Evaluation der Fortschritte im Rahmen eines Prä-Post-Designs hilfreich, um zu analysieren, ob durch die (Art der) Intervention der Zusammenhang zwischen Fremd- und Selbsteinschätzung enger wird. Auch Auswertungen auf Skalenebene könnten die vorliegenden Ergebnisse sinnvoll ergänzen.

Die itemspezifischen Analysen verdeutlichen, dass die Einschätzungen der Studierenden mit den Fremdeinschätzungen durch die geschulten Beobachterinnen in einigen Items enger zusammenhängen als in anderen Items. Dies deutet darauf hin, dass entweder die Einschätzung einiger Aspekte guten Feedbacks insgesamt leichter fällt und damit objektiver möglich ist oder dass die professionelle Wahrnehmung von Feedbacksituationen der Studierenden in einigen Bereichen (durch die Intervention?) höher ausgeprägt ist als in anderen. Insgesamt muss berücksichtigt werden, dass auch die Fremdbeurteilung letztlich eine Einschätzung bleibt, die nicht fehlerfrei ist. Zwar wurde durch ein Manual, eine umfangreiche Schulung, die fortlaufende Prüfung der Reliabilität und das Rating des gesamten Materials durch mehrere Personen versucht, die Objektivität der Daten zu erhöhen. Es handelt sich jedoch weiterhin um ein hoch inferentes Schätzverfahren, das seinerseits Messfehler beinhalten kann (vgl. Praetorius, 2014).

Gerade aufgrund der fortschreitenden Digitalisierung in der Lehrkräftebildung wird das Aufzeichnen und Analysieren von Unterrichtsvideos zunehmend technisch ein-

facher. Dass Videoanalysen grundsätzlich wirksam sein können, belegen inzwischen zahlreiche Studien (vgl. zusammenfassend Steffensky & Kleinknecht 2016). Die hier berichtete Studie leistet einen Beitrag zur Frage, welche Art der videobasierten Lernumgebung besonders lernwirksam ist (vgl. Seago, 2003; Syring, 2021). Obwohl die Effekte und Unterschiede in dieser Studie teilweise nur gering ausfallen, deuten sich dennoch Vorteile für diejenige Interventionsgruppe an, die im Seminar fremde Videos sehr strukturiert mit dem niedrig inferenten Beobachtungsverfahren ausgewertet hatte. Dies unterstreicht die Bedeutsamkeit strukturierter, differenzierter Beobachtungsmethoden für Videoanalysen in der Lehrkräftebildung. Ob diese Ergebnisse auch auf andere Phasen der Lehrkräftebildung oder andere Ausbildungssysteme übertragbar sind, müsste in weiteren Studien überprüft werden.

Literatur

- Balzer, L. & Beywl, W. (2018). *evaluiert – erweitertes Planungsbuch für Evaluationen im Bildungsbereich*. Bern: hep verlag.
- Barth, V., Piwowar, V., Kumschick, I., Ophardt, D. & Thiel, F. (2019). The impact of direct instruction in a problem-based learning setting. Effects of a video-based training program to foster preservice teachers' professional vision of critical incidents in the classroom. *International Journal of Educational Resources*, 95, 1-12.
- Blomberg, G. (2011). *Der Einsatz von Unterrichtsvideos in der universitären Lehrerbildung*. Dissertation: Technische Universität München.
- Brunvand, S. (2010). Best practices for producing video content for teacher education. *Contemporary Issues in Technology and Teacher Education*, 10 (2), 247-256.
- Choi, H. & Jonson, S. (2007). The effect of problem-based video instruction on learner satisfaction, comprehension and retention in college courses. *British Journal of Educational Technology*, 38 (5), 885-895.
- Gaudin, C. & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, 16, 41-67.
- Grossman, P. (Ed.) (2018). *Teaching core practices in teacher education*. Cambridge, MA: Harvard Education Press.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J. (2023). *Visible learning: The sequel. A synthesis of over 2100 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77 (1), 81-112.
- Hess, M. (2019a). Die Handlungskompetenz von Lehrpersonen beim Erteilen von Feedback beurteilen: Ein Beurteilungssystem für unterrichtliche Interaktionssituationen. In K. Schöppe & F. Schulz (Hrsg.), *Kreativität & Bildung: Nachhaltiges Lernen* (S. 465-520). München: kopaed.
- Hess, M. (2019b). Professionell Feedback geben: Lernen mit Videos in der ersten Phase der Grundschullehrerbildung. In C. Donie et al. (Hrsg.), *Grundschulpädagogik zwischen Wissenschaft und Transfer* (S. 102-107). Wiesbaden: Springer.
- Hess, M. & Deistler, A. (2022). Zur videobasierten Erfassung der professionellen Wahrnehmung von Feedbacksituationen: Vorstellung eines Diagnoseinstruments. In K. Kramer & B. Hoyer (Hrsg.), *„Individuell fördern“: Wissenschaftlicher Hintergrund sowie Ansatzpunkte aus und für die Praxis* (S. 149-179). Erlangen: FAU University Press.

- Hess, M. & Lipowsky, F. (2019). Zur Handlungskompetenz von Lehramtsstudierenden beim Erteilen von Feedback: Effekte der Strukturiertheit bei der Analyse eigener Videoaufnahmen. In M.-C. Vierbuchen & F. Bartels (Hrsg.), *Feedback in der Unterrichtspraxis: Schülerinnen und Schüler beim Lernen wirksam unterstützen* (S. 143-160). Stuttgart: Kohlhammer.
- Hess, M., Werker, K. & Lipowsky, F. (2017). Was wissen Lehramtsstudierende über gutes Feedback? Zur Erfassung konzeptuellen Wissens und zu dessen Zusammenhang mit der Selbsteinschätzung der Studierenden. *Jahrbuch für Allgemeine Didaktik*, 11-29.
- Holodynski, M. & Meschede, N. (2022). Videobasierte Lehre und Forschung in der Lehrkräftebildung. *Quo vadis?* In R. Junker, V. Zucker, M. Oellers, T. Rauterberg, S. Konjer, N. Meschede & M. Holodynski (Hrsg.), *Lehren und Forschen mit Videos in der Lehrkräftebildung* (S. 197-217). Münster: Waxmann.
- Huang, Y., Miller, K., Cortina, K. & Richter, D. (2023). Teachers' professional vision in action. Comparing expert and novice teacher's real life eye movements in the classroom. *Zeitschrift für Pädagogische Psychologie*, 37, 122-139.
- Kleinknecht, M. & Poschinski, N. (2014). Eigene und fremde Videos in der Lehrerfortbildung. Eine Fallanalyse zu kognitiven und emotionalen Prozessen beim Beobachten zweier unterschiedlicher Videotypen. *Zeitschrift für Pädagogik*, 60 (3), 471-490.
- Kücholl, D. & Lazarides, R. (2021). Video- und protokollbasierte Reflexionen eigener praktischer Unterrichtserfahrungen im Lehramtsstudium. *Zeitschrift für Erziehungswissenschaft*, 24, 985-1006. doi.org/10.1007/s11618-021-01021-8
- Kultusministerkonferenz (2019). Standards für die Lehrerbildung: Bildungswissenschaften. Bonn und Berlin: Sekretariat der Kultusministerkonferenz. Verfügbar unter: https://www.kmk.org/fileadmin/veroeffentlichungen_beschlusse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf [08.08.23].
- Lotz, M. (2015). *Kognitive Aktivierung im Leseunterricht der Grundschule*. Wiesbaden: Springer.
- Moreno, R., Abercrombie, S. & Hushman, C. (2009). Using virtual classroom cases as thinking tools in teacher education. In I. Gibson et al. (Eds.), *Proceedings of society for information technology & teacher education international conference 2009* (pp. 2615-2622). Chesapeake, VA: AACE.
- Narciss, S. (2004). The impact of informative tutoring feedback and self-efficacy on motivation and achievement in concept learning. In *Experimental Psychology*, 51 (3), 214-228.
- Narciss, S. (2006). Informatives tutorielles Feedback. *Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse*. Münster: Waxmann.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Roth, K., Garnier, H., Chen, C., Lemmens, M., Schwille, K. & Wickler, N. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, 48 (2), 117-148.
- Sablić, M., Miroslavljević, A. & Škugor, A. (2020). Video-based learning (VBL)—Past, present and future: An overview of the research published from 2008 to 2019. *Technology, Knowledge and Learning*, 26, 1061-1077.
- Santagata, R., & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *The International Journal of Mathematics Education*, 43 (1), 133-145.
- Seago, N. (2003). Using video as an object of inquiry mathematics teaching and learning. In J. Brophy (Ed.), *Using video in teacher education* (pp. 259-285). Amsterdam: *Advances in Research on Teaching*.
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. *Integration von Struktur- und Prozessparadigma*. *Zeitschrift für Pädagogik*, 60 (6), 850-866.
- Seidel, T., Blomberg, G. & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, 34, 56-65.
- Seidel, T. & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, 51 (4), 739-771.
- Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M. & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education*, 27, 259-267.

- Sherin, M. G. (2004). New perspectives on the role of video in teacher education. In J. Brophy (Ed.), *Using video in teacher education* (pp. 1-28). Oxford, UK: Elsevier.
- Sherin, M. G. & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60 (1), 20-37.
- Star, J. & Strickland, S. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education*, 11 (2), 107-125.
- Steffensky, M. & Kleinknecht, M. (2016): Wirkungen videobasierter Lernumgebungen auf die professionelle Kompetenz und das Handeln (angehender) Lehrpersonen. Ein Überblick zu Ergebnissen aus aktuellen (quasi-)experimentellen Studien. *Unterrichtswissenschaft*, 44, 305-321.
- Syring, M. (2021). Videobasierte Kasuistik in der Lehre. In D. Wittek, T. Rabe & M. Ritter (Hrsg.), *Kasuistik in Forschung und Lehre* (S. 230-244). Bad Heilbrunn: Klinkhardt.
- Wisniewski, B., Zierer, K. & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087. doi.org/10.3389/fpsyg.2019.03087

Autorinnen

Prof. Dr. Miriam Hess, Universität Bamberg, Lehrstuhl für Grundschulpädagogik und -didaktik

Dr. Ann-Katrin Denn, Hochschule für öffentliches Management und Sicherheit, Kassel

Korrespondenz an: miriam.hess@uni-bamberg.de