

Worauf kommt es wirklich an? Eine explorative Studie zu Feedback in der zweiten Phase der Lehrerbildung

Benedikt Wisniewski, Stefan T. Siegel

Angaben zur Veröffentlichung / Publication details:

Wisniewski, Benedikt, and Stefan T. Siegel. 2020. "Worauf kommt es wirklich an? Eine explorative Studie zu Feedback in der zweiten Phase der Lehrerbildung." *LbP - Lehrerbildung auf dem Prüfstand* 13 (2): 240–60. <https://www.vep-landau.de/produkt/lehrerbildung-auf-dem-pruefstand-2020-13-2-kap-5-digital/>.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

Deutsches Urheberrecht

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



Benedikt Wisniewski & Stefan T. Siegel

Worauf kommt es wirklich an? Eine explorative Studie zu Feedback in der zweiten Phase der Lehrerbildung

Feedback spielt eine wesentliche Rolle für die Professionalisierung von Berufsanfänger*innen in der zweiten Phase der Lehrerbildung. Welche Inhalte Gegenstand dieses Feedbacks sind, ist bisher weitgehend unerforscht. In diesem Beitrag wird exploriert, welche Kriterien Lehrerbildner*innen als Grundlage für das Feedback verwenden, das sie Lehramtsanwärter*innen geben. 127 Lehrerbildner*innen (52 % weiblich, Durchschnittsalter: 52,3 Jahre) aus 10 der 16 deutschen Bundesländer nahmen an der querschnittlich angelegten Online-Befragung teil. Sowohl die Häufigkeit von Feedbackkriterien als auch die subjektiven Vorstellungen von Unterrichtsqualität und die Einstellungen zu wissenschaftlichen Theorien wurden anhand von Selbstberichten erfasst. Die Befunde weisen darauf hin, dass sich Feedback von Lehrerbildner*innen schwerpunktmäßig auf Tiefendimensionen der Unterrichtsqualität bezieht. Theorieverdrossenheit beeinflusst zum Teil, wie häufig welche Merkmale von Unterrichtsqualität im Feedback thematisiert werden. Geschlecht, Alter und Berufserfahrung von Lehrerbildner*innen haben kaum Einfluss auf die Inhalte des Feedbacks.

Schlagwörter: Feedback - Lehrerbildner*innen - Zweite Phase der Lehrerbildung - Unterrichtsqualität - Pädagogische Professionalisierung

What is really important? An explorative study on feedback in the second phase of teacher training

Feedback plays an essential role in the professional development of trainee teachers during their practical training. The content of this feedback is still largely unexplored. This paper explores what criteria teacher trainers use as a basis for the feedback they give to trainee teachers. 127 teacher trainers (52 % female, average age: 52.3 years) from 10 of the 16 German federal states took part in the cross-sectional online survey. The frequency of feedback criteria as well as subjective beliefs about teaching quality and attitudes towards scientific theories were obtained on the basis of self-reports. The findings indicate that feedback from teacher trainers focuses on the deep structures of instructional quality. Dismissal of scientific theories partly influences how often which characteristics of instructional quality are addressed in feedback. Gender, age and professional experience of teacher trainers have little influence on the content of the feedback they give.

Keywords: Feedback - teacher trainers - practical teacher training - instructional quality - professional development

1 Einleitung

Feedback als wesentliche Voraussetzung von Professionalisierung (Bastian, Combe & Reh, 2002; Reh, 2004) stellt für die zweite Phase der Lehrerbildung ein wichtiges Instrument dar: Erhaltenes Feedback dient Berufsanfänger*innen dazu, das eigene

berufliche Handeln im Abgleich mit einer Außenperspektive zu reflektieren und anzupassen (Gärtner, 2013). Eine Selbstreflexion, die sich aus einem Diskurs über eigene Erfahrungen und externe Wahrnehmungen ergibt und dabei wissenschaftliche Erkenntnisse mit einbezieht, ist der Kern pädagogischer Professionalität (Schratz, Paseka & Schrittmesser, 2011). Einen Großteil des Feedbacks in der zweiten Phase der Lehrerbildung erhalten Lehramtsanwärter*innen von den Lehrpersonen, die sie ausbilden (Behnke, 2016). Diese Lehrerbildner*innen sind für die Durchführung und Gestaltung der Ausbildung angehender Lehrpersonen am Studienseminar zuständig, während derer sie eine beratende Funktion einnehmen und gleichzeitig implizit oder explizit mit der Bewertung und Beurteilung betraut sind (Behnke, 2016). Dies führt zu einer ambivalenten Rolle dieser Personengruppe und gleichzeitig zu einer nicht klar eingegrenzten Funktion von Feedback während der Ausbildung. Lehramtsanwärter*innen erhalten Rückmeldungen zu verschiedenen Bereichen der Berufsausübung, darunter (oder schwerpunktmäßig) zur Qualität des gehaltenen Unterrichts. Es existieren für Rückmeldungen der Lehrerbildner*innen keine verbindlichen Rückmeldekriterien, weshalb unklar ist, welche Kriterien durch Feedback abgedeckt werden, anhand welcher wissenschaftlicher oder nichtwissenschaftlicher Grundlage diese Kriterien ausgewählt werden und wie einheitlich oder uneinheitlich diese angewendet werden.

Im Folgenden wird dieses Forschungsdesiderat aufgegriffen und nach einem kurzen theoretischen Überblick die Ergebnisse eines explorativen Forschungsprojekts berichtet. In dessen Rahmen wurde untersucht, welche inhaltlichen Kriterien Lehrerbildner*innen dem Feedback zu Grunde legen, das sie Lehramtsanwärter*innen geben und inwiefern die Feedback-Kriterien mit der individuellen Ausprägung von Theorieverdrossenheit (im Sinne der Einschätzung wissenschaftlicher Inhalte als unwesentlich für die zweite Phase der Lehrerbildung) zusammenhängen.

1.1 Feedback an Lehramtsanwärter*innen

Feedback ist die Information, die einer Person von einer anderen Person zu Aspekten der eigenen Leistung oder des eigenen Verständnisses gegeben wird (Hattie & Timperley, 2007). Es scheint problematisch, dass dieses von Lehramtsanwärter*innen in vielen Fällen als belastend erlebt wird, weil es in erster Linie als Kritik an der eigenen Person wahrgenommen wird (Schaefers, 2002). Dazu, welche Rückmeldungen Lehramtsanwärter*innen tatsächlich erhalten, existieren wenige Forschungsergebnisse. Als Anhaltspunkt können jedoch Kriterienkataloge verwendet werden, welche bei der Bewertung und Beurteilung im Vorbereitungsdienst eingesetzt werden, da ähnliche Kriterien, die in diesen enthalten sind, mit hoher Wahrscheinlichkeit auch die Grundlage von Feedbackgesprächen außerhalb der Beurteilungssituation bilden, beispielsweise nach Unterrichtshospitationen durch

die Lehrerbildner*innen. Diese Kriterien werden in unsystematischer und uneinheitlicher Weise verwendet (Döbrich & Abs, 2008) und spiegeln häufig weder ältere noch neuere einschlägige Forschungsbefunde der Schul- und Unterrichtsforschung wider (Strietholt & Terhart, 2009). Dies wirft die Frage auf, inwieweit Lehrerbildner*innen die Feedbackinhalte anhand ihrer subjektiven Vorstellungen definieren.

1.2 Vorstellungen von Unterrichtsqualität als Grundlage von Feedback

Wenn Lehrerbildner*innen Lehramtsanwärter*innen Feedback zu deren Unterricht geben, greifen sie auf unterschiedliche Vorstellungen dieses Konstrukts zu. Während zu den epistemologischen und pädagogischen Überzeugungen von Lehrpersonen bereits umfangreiche Befunde vorliegen (Gilleece, 2012; Hartinger, Kleickmann & Hawelka, 2006; Leuchter, Pauli, Reusser & Lipowsky, 2006; Negrini, 2016; Staub & Stern, 2002), ist wenig dazu bekannt, welche Vorstellungen Lehrpersonen davon haben, was qualitativollen Unterricht ausmacht. Bei Vorstellungen von Unterrichtsqualität handelt es sich um subjektive Theorien, welche einen personenspezifischen Grad an Elaboration und Bezug zu wissenschaftlichen Theorien aufweisen (Wahl, 2013). Auch die subjektiven Theorien von Lehrerbildner*innen zur Unterrichtsqualität beruhen nicht auf einem einheitlichen Konsens (Sykes, Bird & Kennedy, 2010).

Aus wissenschaftlich-empirischer Sicht wird Unterrichtsqualität als mehrdimensionales Konstrukt aufgefasst (Ferguson & Danielson, 2014; Helmke, 2017; Klieme, Lipowsky & Rakoczy, 2006), welches sich je nach Modellvorstellung in verschiedene Dimensionen aufgliedern lässt: Zunächst wird zwischen Tiefenstrukturen und Oberflächenmerkmalen von Unterrichtsqualität differenziert (Kunter & Voss, 2013, Abbildung 1). Während sich Oberflächenmerkmale auf direkt beobachtbare Unterrichtsmerkmale beziehen, beschreiben Tiefenstrukturen Merkmale, die sich aus der Interaktion zwischen Lehrenden und Lernenden ergeben. Tiefenstrukturen variieren innerhalb ein und derselben Lehrperson deutlich weniger als Oberflächenmerkmale (Praetorius, Pauli, Reusser, Rakoczy & Klieme, 2014). Letztere haben eine geringere Erklärungsmacht, insofern dass sie relevante Ergebnisse von Unterricht (in erster Linie den Lernerfolg von Lernenden) deutlich schlechter vorhersagen (Seidel & Shavelson, 2007).



Abbildung 1: Oberflächen- und Tiefenstrukturen der Unterrichtsqualität (eigene Abbildung)

Zur Beschreibung der Tiefenstrukturen von Unterrichtsqualität existieren theoretisch gut begründete Modelle (Ferguson & Danielson, 2014; Helmke, 2017; Klieme et al., 2006; Slavin, 1994). Die Dimensionalität dieser Modelle ist empirisch abgesichert (Fauth, Decristan, Rieser, Klieme & Büttner, 2014; Künsting, Neuber & Lipowsky, 2016; Kunter & Voss, 2013; Praetorius, Klieme, Herbert & Pinger, 2018; Wisniewski et al., 2020). Dabei kann allgemeine Unterrichtsqualität als relativ stabiles Merkmal aufgefasst werden (Wagner et al., 2016), welches fächerübergreifend konzeptualisierbar ist (Praetorius et al., 2018).

Zusätzlich zu Oberflächen- und Tiefenstrukturen der Unterrichtsqualität enthalten subjektive Vorstellungen von Lehrpersonen in Bezug auf qualitativollen Unterricht eine zusätzliche Komponente, die sich unter dem Begriff der personenbezogenen Merkmale zusammenfassen lässt. Dieser Begriff bezieht sich auf diejenigen Eigenschaften oder Persönlichkeitsmerkmale, die als Prädiktoren für erfolgreiches Unterrichten angesehen werden. Aus Sicht der Unterrichtsforschung ist der Versuch, Eigenschaften erfolgreicher Lehrkräfte zu identifizieren, die einen guten Unterricht vorhersagen, zwar überholt (Bromme & Haag, 2004), jedoch werden von Lehrpersonen bis zu 80 % der erfolgreichen Berufsausübung personenbezogenen Merkmalen zugeschrieben (Bromme & Haag, 2004; Helmke, 2017; Herrmann & Hertrampf, 1997; Lipowsky, 2003). Auch in Beurteilungsinstrumenten zum Unterricht werden Merkmale wie Höflichkeit, Humor, Äußeres, Ausstrahlung etc. miteinbezogen (Strietholt & Terhart, 2009). Als Fragestellung ergibt sich daher, zu welchem Grad die subjektiven Vorstellungen von Lehrerbildner*innen zum qualitativollen Unterricht durch wissenschaftliche Erkenntnisse beeinflusst sind bzw. zu welchem Grad wissenschaftliche Erkenntnisse (bspw. in Bezug auf die Relevanz von Persönlichkeitsmerkmalen) ignoriert oder abgelehnt werden.

1.3 Relevanz von Einstellungen zu wissenschaftlichen Erkenntnissen

Die kolportierte Standardbegrüßung von Lehramtsanwärter*innen im Studienseminar "Vergessen Sie alles, was Sie an der Uni gelernt haben" (Wenzl, Wernet & Kollmer, 2018, S. 18) wird häufig herangezogen, um einen vermeintlichen Bruch zwischen erster und zweiter Phase der Lehrerbildung und auf die Irrelevanz wissenschaftlicher Erkenntnisse für die tatsächliche Praxis zu verweisen. Inwieweit Lehrerbildner*innen wissenschaftliche Erkenntnisse tatsächlich ablehnen, ist dagegen bisher wenig untersucht. Das Konstrukt der Theorieverdrossenheit beschreibt negative theoriebezogene Einstellungen, die sich bspw. in Beurteilungen von wissenschaftlichen Theorien als dröge, nutzlos und praxisfern manifestieren (Bleck & Lipowsky, 2020). Einstellungen von Lehrpersonen erfuhren in den letzten Dekaden verstärkte Aufmerksamkeit in der pädagogisch-psychologischen Professionalisierungsforschung (Baumert & Kunter, 2006). Sembill und Seifried (2009, S. 346.) bezeichnen diese als Lehrersichtweisen, "die [...] mentale Vorgänge [steuern], wie ein Filter [wirken] und bei der Informationsaufnahme und beim Wissenserwerb aktiv und damit sui generis internal und external handlungsleitend [sind]".

Einstellungen haben großen Einfluss auf das Denken und Handeln von Lehrpersonen, können sowohl professionalisierungshemmend als auch -fördernd wirken und stellen wichtige Elemente einschlägiger Kompetenzmodelle des Lehrer*innenhandelns dar (Baumert & Kunter, 2006; 2013; Blömeke, Kaiser & Lehmann, 2008; Sembill & Seifried, 2009). Dabei sind Einstellungen von Lehrerbildner*innen insofern von besonderem Interesse, da diese nicht nur ihre eigene Professionalisierung beeinflussen, sondern auch die der weit größeren Zahl der von ihnen ausgebildeten Lehramtsanwärter*innen. Es wird angenommen, dass die Einstellungen dieser Personengruppe gegenüber wissenschaftlichen Befunden einen Einfluss darauf haben, wie und welche Art von Feedback sie Lehramtsanwärter*innen geben (Behnke, 2016; Bleck & Lipowsky, 2020), jedoch fehlen hierzu bisher empirische Befunde.

1.4 Forschungsziele, -fragen und -hypothesen

Ein grundsätzliches Desiderat der Forschung besteht dahingehend, welche Art von Feedback Lehramtsanwärter*innen in der zweiten Phase der Lehrerbildung erhalten. Die übergreifenden Forschungsziele der vorliegenden Studie sind die Exploration von Vorstellungen von Lehrerbildner*innen zur Unterrichtsqualität, welche dem Feedback zu Grunde liegen, die inhaltliche Eingrenzung der Feedback-Kriterien und die Prüfung, durch welche Variablen diese Kriterien beeinflusst werden. Aus diesen Zielen ergeben sich folgende Forschungsfragen:

F1: Welche subjektiven Vorstellungen von Unterrichtsqualität haben Lehrerbildner*innen?

F2: Welche inhaltlichen Kriterien enthält das Feedback von Lehrerbildner*innen an Lehramtsanwärter*innen?

F3: Wie hängen die subjektiven Vorstellungen von Unterrichtsqualität und die Feedback-Kriterien zusammen?

F4: Wie wirkt die Theorieverdrossenheit von Lehrerbildner*innen auf deren Feedback-Kriterien?

F5: Wie wirken Personenvariablen auf Theorieverdrossenheit und Feedback-Kriterien von Lehrerbildner*innen?

Nach der explorativen Untersuchung von F1 und F2 sollen zur Beantwortung der restlichen forschungsleitenden Fragen folgende Hypothesen geprüft werden:

H3: Die Gewichtung von Tiefenstrukturen (H3.1), Oberflächenmerkmalen (H3.2) und Persönlichkeitsmerkmalen (H3.3) bei der Definition von Unterrichtsqualität weist eine signifikante Korrelation mit der Häufigkeit der Thematisierung dieser drei Bereiche beim Geben von Feedback auf.

H4: Mit zunehmender Theorieverdrossenheit sinkt die Häufigkeit der Thematisierung von Tiefenstrukturen (H4.1) und steigen die Häufigkeiten der Thematisierung von Oberflächenmerkmalen (H4.2) und Persönlichkeitsmerkmalen (H4.3) beim Geben von Feedback.

H5.1: Personenbezogene Variablen (H5.1.1 Geschlecht, H5.1.2 Alter, H5.1.3 Berufserfahrung) haben einen Einfluss auf die Theorieverdrossenheit.

H5.2: Personenbezogene Variablen (H5.2.1 Geschlecht, H5.2.2 Alter, H5.2.3 Berufserfahrung) haben einen Einfluss auf die Häufigkeit der Verwendung von Feedback-Kriterien.

2 Methode

2.1 Design

Um die Forschungsfragen adäquat zu beantworten, wurde eine Mixed-Methods-Studie realisiert (Cresswell & Plano, 2018). Mittels eines teilstandardisierten Fragebogens wurde in einer querschnittlich angelegten Onlinebefragung in einem 60-tägigen Zeitraum von Mitte Februar bis Mitte April 2020 sowohl qualitatives als auch quantitatives Forschungsmaterial erhoben. Um eine heterogene Stichprobe zu erhalten, erfolgte die Rekrutierung über den zentralen E-Mail-Verteiler des Bundesarbeitskreises der Seminar- und Fachleiter/innen e.V. (BAK). Dieser bundesweite Berufsverband hat nach eigenen Angaben etwa 2000 Mitglieder aus allen Bundesländern, Lehrämtern und Schularten (BAK, 2016). Der elektronische Fragebogen wurde den Teilnehmenden über die Enterprise Feedback Suite (EFS) zur Bearbeitung

zugänglich gemacht und ermöglichte eine breit angelegte und gleichzeitig ökonomische Datenerhebung. Die Studie wurde unter Einhaltung der Prinzipien des Ethik-Kodex der Deutschen Gesellschaft für Erziehungswissenschaft DGfE (2016) durchgeführt.

2.2 Stichprobencharakterisierung

Insgesamt nahmen 127 Lehrerbildner*innen der zweiten Phase der Lehrerbildung an der Onlinebefragung teil. 66 der Befragten (52 %) sind weiblich. Das Durchschnittsalter der Befragten lag bei 52.08 Jahren ($SD = 8.11$; Range: 31–65). Es nahmen Lehrerbildner*innen aus insgesamt 10 der 16 Bundesländer der Bundesrepublik Deutschland teil. 94.4 % der Teilnehmenden gaben an, über eine Berufserfahrung als Lehrperson von mehr als 10 Jahren zu verfügen. Nur 2.4 % hatten weniger als 5 Jahre Berufserfahrung. Die befragten Lehrerbildner*innen sind in sämtlichen Schularten tätig (24 in Grundschulen, 15 in Mittelschulen, 11 in Realschulen, 67 in Gymnasien, 32 in sonstigen Schulen, wie bspw. Berufsschulen, Gemeinschaftsschulen oder Sonderpädagogischen Förderzentren).

2.3 Datenerhebung

Bei der Erhebung der Daten wurde, insofern möglich, auf bestehende reliable und valide Skalen zurückgegriffen, oder diese wurden adaptiert. Falls dies nicht möglich war, wurden neue Instrumente sorgfältig entwickelt und einem quantitativen Prätest ($N = 12$) unterzogen, um die Verständlichkeit der Items und einen reibungslosen Ablauf der Befragung sicherzustellen (vgl. Döring & Bortz 2016).

2.3.1 Unterrichtsqualität

Die Teilnehmenden wurden gebeten, die aus ihrer Sicht wichtigsten Merkmale von Unterrichtsqualität (mindestens fünf bis maximal zehn) in freie Textfelder einzutragen.

2.3.2 Feedback-Kriterien

Den Befragten wurden darauffolgend Items zu möglichen Feedback-Kriterien präsentiert, für die sie jeweils angeben sollten, wie oft sie die enthaltenen Merkmale thematisieren, wenn sie Lehramtsanwärter*innen Feedback zu deren Unterricht geben. Dazu wurde eine Likert-artige Skala mit den Polen 1 (*thematisiere ich nie*) bis 5 (*thematisiere ich immer*) als diskretes Antwortformat benutzt. Die Items enthielten Inhalte aus den Bereichen Tiefenstrukturen (Wertschätzung, Klassenführung, Lernerfolgssicherung, Motivierung, Herausforderung, Zusammenarbeit und Rückmeldung sowie Klarheit), Oberflächenmerkmale (u. a. Einsatz verschiedener Sozialformen, Einsatz verschiedener Methoden, Medieneinsatz) und personenbezogene Merkmale (u. a. ausgeprägte Lehrerpersönlichkeit, Charisma, Humor, Autorität). Die

Items zu den Tiefenstrukturen wurden aus dem Fragebogen teaCh (Wisniewski & Zierer, 2020) übernommen und hinsichtlich der Zielpopulationen Lehrerbildner*innen und Lehramtsanwärter*innen adaptiert.

2.3.3 Theorieverdrossenheit

Um die Theorieverdrossenheit der Teilnehmenden zu erfassen, wurde eine Skala von Bleck et al. (2020) verwendet. Auf einer 5-stufigen Likert-artigen Skala mit den Polen 1 (*stimme überhaupt nicht zu*) bis 5 (*stimme voll und ganz zu*) gaben die Befragten ihre Zustimmung zur Bedeutung von wissenschaftlichen Erkenntnissen für ihre pädagogische Praxis (u. a. "Theoretische Überlegungen sind für den Unterrichtsalltag nicht relevant").

2.3.4 Personenmerkmale

Es wurde erfasst, wie groß die bisherige Berufserfahrung der Teilnehmer*innen als Lehrperson ist. Darüber hinaus wurden Geschlecht und Alter der Teilnehmenden erfasst.

2.4 Datenaufbereitung

Die Datenaufbereitung und -analyse des qualitativen Forschungsmaterials erfolgte mit der Software MAXQDA 2018, die der quantitativen Daten mit der Statistiksoftware R (R Core Team, 2019, Version 3.6.2) und den Paketen lavaan (Rosseel, 2012, Version 0.6-5) und mctest (Imdad & Aslam M., 2018, Version 1.3). Zeitbasiert wurden die Antworten der Teilnehmenden auf alle Items, die in weniger als einer Sekunde pro Item beantwortet wurden, als fehlend gesetzt, da davon ausgegangen werden konnte, dass diese Angaben nicht ernsthaft bearbeitet wurden.

2.5 Datenanalyse

Die offenen Antworten der Befragten zu den aus ihrer Sicht wichtigsten Merkmalen von Unterrichtsqualität wurden in Anlehnung an die qualitative Inhaltsanalyse nach Kuckartz (2018) kodiert. Während die Hauptkategorien (Tiefenstrukturen, Oberflächenmerkmale und personenbezogene Merkmale) aufgrund des spezifischen Erkenntnisinteresses dieser Studie deduktiv gesetzt wurden, erfolgte die Entwicklung und Ausdifferenzierung der Subkategorien induktiv am Forschungsmaterial (siehe Tabelle 1). Die Antworten der Befragten wurden jeweils von den Autoren des vorliegenden Beitrags den Kategorien Tiefenstrukturen, Oberflächenmerkmale, und personenbezogene Merkmale zugeordnet. Dabei wurde das gesamte Material doppelt kodiert. Die Intercoder-Reliabilität war mit Cohens $\kappa = .85$ und Krippendorfs $\alpha = .82$ (Krippendorff, 2004) sehr gut.

Hinsichtlich des quantitativen Forschungsmaterials wurden zunächst die deskriptiven Statistiken berechnet. Aufgrund der nicht vorliegenden Normalverteilung der

Daten wurden Korrelationen als Spearman's rho berechnet. Zur Prüfung der internen Konsistenzen der verwendeten Skalen wurde zunächst McDonald's omega herangezogen. Darüber hinaus wurden konfirmatorische Faktorenanalysen durchgeführt und auf dieser Basis ein latentes Strukturgleichungsmodell konzipiert. Als Schätzverfahren wurde die Robust Maximum Likelihood Estimation (MLR) verwendet, welche robust gegenüber Nicht-Normalität der Daten ist. Zur Prüfung der Anpassungsgüte des Modells an die Daten wurden konservative Fit-Statistiken verwendet. Während von Hu & Bentler (1999) Cut-off-Werte von $\chi^2/df \leq 3$, $CFI \geq .95$, $TLI \geq .95$, $RMSEA \leq .06$ und $SRMR \leq .08$ für einen guten Modellfit vorgeschlagen werden, werden gerade für komplexere Modelle ein $CFI > .90$ und ein $TLI > .90$ als Hinweise auf einen akzeptablen Modellfit aufgefasst (Browne & Cudeck, 1993; Kline, 2005, Marsh, Hau, & Wen, 2004).

Abweichend von der streng konfirmatorischen Faktorenanalyse wurde im Sinne eines modellgenerierenden Ansatzes ein tentatives Modell an den Daten geprüft und bei mangelnder Passung verändert (Floyd & Widaman, 1995; Jöreskog, 1993). Konkret bedeutete dies für unser Modell, dass auf der Basis von Modification Indices Korrelationen zwischen Residualvarianzen ähnlicher Items zugelassen wurden (Landis, Edwards, & Cortina, 2009).

Die konfirmatorischen Faktorenanalysen lieferten für die Häufigkeit der Thematisierung drei Dimensionen (Tiefenstrukturen, Oberflächenmerkmale und Persönlichkeitsmerkmale) als Feedback-Kriterien ($CFI = .96$; $TLI = .95$; $RMSEA = .038$; $SRMR = .08$) und das Konstrukt der Theorieverdrossenheit ($CFI = .97$; $TLI = .93$; $RMSEA = .04$; $SRMR = .09$) jeweils einen akzeptablen bis guten Modellfit. Die beiden Messmodelle sind als Teil des Strukturgleichungsmodells in Abbildung 3 dargestellt.

Bei der Regressionsanalyse der Personenvariablen deuteten hohe Interkorrelationen zweier unabhängiger Variablen ($r = .86$ zwischen Alter und Berufserfahrung) zusammen mit unterschiedlichen Vorzeichen der standardisierten Regressionskoeffizienten auf Multikollinearität hin, was durch verschiedene Detektionsverfahren (Farrar & Glauber, 1967; Klein, 1962; Theil, 1971) bestätigt wurde. Daher wurde die Analyse stufenweise durchgeführt, d.h. es wurde zunächst das Maximalmodell mit allen Prädiktoren gerechnet und dann das Alter ausgeschlossen.

3 Ergebnisse

3.1 Qualitative Ergebnisse

Zur Exploration der Vorstellungen von Lehrerbildner*innen zur Unterrichtsqualität wurden deren Freitext-Antworten inhaltsanalytisch ausgewertet. Tabelle 1 zeigt die Kategorien und die Anzahl der Kodierungen.

Tabelle 1: Kategoriensystem und Anzahl der Kodierungen zu den von den befragten Lehrerbildner*innen genannten Merkmalen von Unterrichtsqualität

Kategorie	Anzahl der Kodierungen
Tiefenstrukturen	507
Verständlichkeit/inhaltliche Klarheit	67
Kognitive Aktivierung	91
Konstruktive Unterstützung	260
Classroom Management	86
Oberflächenmerkmale	124
Angemessener/r Sozialformwahl und -wechsel	19
Methodenwahl, -einsatz und -wechsel	30
Medienwahl, -einsatz und -wechsel	18
Schüleraktivierung	37
Arbeitsraum- und Arbeitsmaterialgestaltung	5
Lehrplanbezug	12
Personenbezogene Merkmale	120
Fachliche Kompetenzen	52
Persönlichkeitsmerkmale	57
Nicht eindeutig zuordenbar	168

Die Antworten der befragten Lehrerbildner*innen zu den aus ihrer Sicht wichtigsten Merkmalen von Unterrichtsqualität weisen eine hohe Heterogenität auf, was sich u. a. an der Anzahl und Unterschiedlichkeit der entwickelten Kategorien zeigt. Die Lehrerbildner*innen nannten durchschnittlich 7.24 Merkmale ($SD = 2.38$). Neben 357 fehlenden Angaben und 168 nicht eindeutig zuordenbaren Antworten (z. B. "Wissenschaftspropädeutik", "gute Schulleitung", "Teamfähigkeit") konnten 751 den Kategorien Tiefenstrukturen (507), Oberflächenmerkmale (124) und Lehrpersonenmerkmale (120) zugeordnet werden.

Diesbezüglich zeigt sich, dass die Mehrzahl der Nennungen auf Tiefenstrukturen (u. a. "Kognitive Aktivierung") entfällt, wobei der Kategorie konstruktive Unterstützung 260 Kodierungen zugeordnet wurden. Als Oberflächenmerkmale wurden primär Antworten odiiert, welche den Kategorien Schüleraktivierung (37; z. B. „Vielfältige Schüleraktivierung“), Methodenwahl, -einsatz, -wechsel (25; z. B. „Methodenwechsel“), angemessene/r Sozialformwahl und -wechsel (19; u. a. „Offene

Lernsettings“) und Medienwahl, -einsatz, -wechsel (18; „Gute Medien“ etc.) zugeordnet wurden. Hinsichtlich der Lehrpersonenmerkmale wurde zwischen Persönlichkeitsmerkmalen (57; z. B. „Authentizität“) und Kompetenzen (52; u. a. „Medienkompetenz“) unterschieden.

3.2 Quantitative Ergebnisse

3.2.1 Deskriptive Statistiken

Die Skalenmittelwerte für die Einschätzungen, welche Kriterien beim Geben von Feedback berücksichtigt werden (siehe Abbildung 2), waren am höchsten für Tiefenstrukturen (M = 3.57; SD = 0.51), gefolgt von Oberflächenmerkmalen (M = 2.82; SD = 0.65) und personenbezogenen Merkmalen (M = 2.55; SD = 0.81).

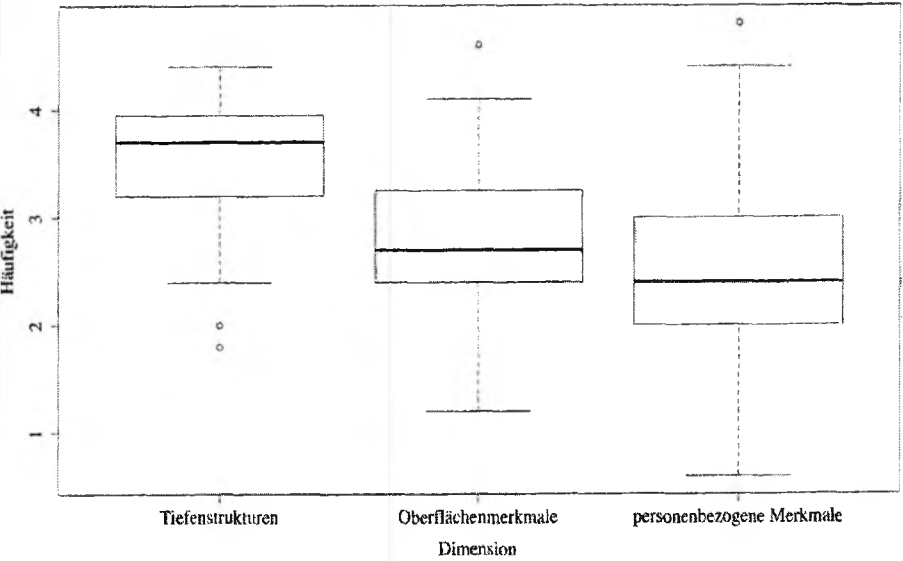


Abbildung 2: Skalenmittelwerte der Häufigkeitseinschätzungen

Die internen Konsistenzen lagen bei $\omega = .90$ für Tiefenstrukturen, $\omega = .89$ für Oberflächenmerkmale, und $\omega = .82$ für personenbezogene Merkmale. Die vier Items zur Einschätzung der Theorieverdrossenheit wiesen ebenfalls eine akzeptable innere Konsistenz auf ($\omega = .85$) und der aggregierte Gesamtwert für diese Skala lag insgesamt unterhalb des theoretischen Mittelwerts (M = 2.28; SD = 0.78).

3.2.2 Zusammenhänge zwischen subjektiven Vorstellungen von Unterrichtsqualität und der Häufigkeit der Verwendung der Feedback-Kriterien

Die Zusammenhänge zwischen den Anteilen von frei genannten Tiefenstrukturen, Oberflächenmerkmalen und Persönlichkeitsmerkmalen als Teile der subjektiven Vorstellungen der Unterrichtsqualität und der Einschätzung, wie häufig diese Merkmale als Bestandteile des gegebenen Feedbacks verwendet werden, sind in Tabelle 2 zusammengefasst. Die subjektiven Vorstellungen korrelieren untereinander signifikant, sind jedoch weitgehend unabhängig von der Häufigkeit der Feedback-Kriterien-Verwendung, mit Ausnahme der Verwendung von Tiefenstrukturen. Die Häufigkeit der Verwendung der Tiefenstrukturen im Feedback hängt positiv mit den Anteilen von frei genannten Tiefenstrukturmerkmalen und negativ mit den Anteilen von frei genannten Oberflächenmerkmalen zusammen.

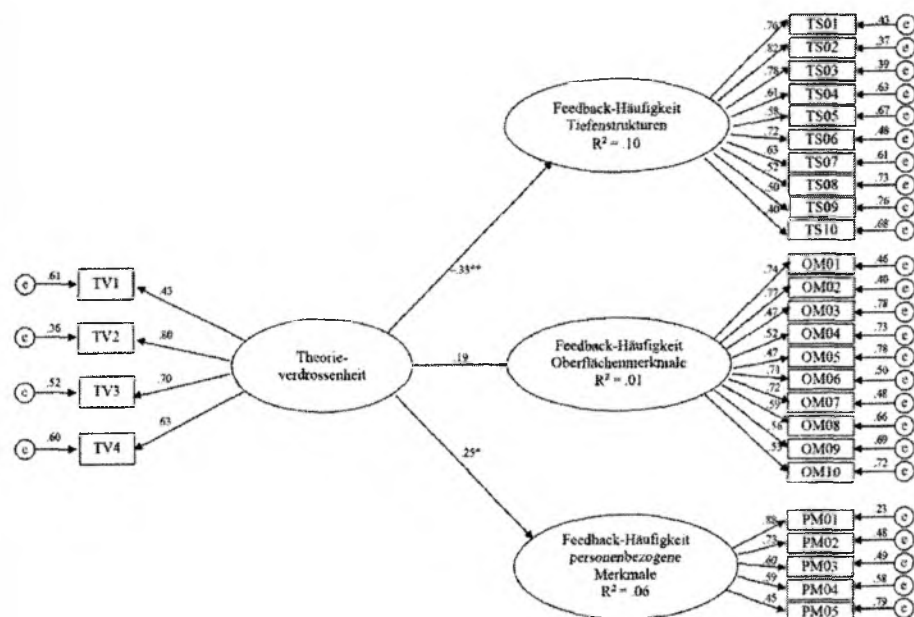
Tabelle 2: Korrelationsmatrix: Gewichtungen und Häufigkeiten im Feedback von Dimensionen der Unterrichtsqualität

		Gewichtung OM	Gewichtung PM	Häufigkeit TS	Häufigkeit OM	Häufigkeit PM
Gewichtung TS	rho	-.53	-.63	.22	-.04	-.06
	p	<.001	<.001	.01	.68	.49
Gewichtung OM	rho		-.24	-.19	.05	-.08
	p		.01	.03	.55	.37
Gewichtung PM	rho			-.03	.01	.10
	p			.76	.91	.25
Häufigkeit TS	rho				.33	.13
	p				<.001	.14
Häufigkeit OM	rho					.53
	p					<.001

Anmerkungen: TS, Tiefenstrukturen; OM, Oberflächenmerkmale; PM, Personenbezogene Merkmale

3.2.3 Regression der Häufigkeit der Feedback-Kriterien auf die Theorieverdrossenheit

Abbildung 3 zeigt die Regression der latenten Häufigkeits-Variablen auf die latente Variable Theorieverdrossenheit als Teil des Strukturgleichungsmodells. Während die Häufigkeit von Oberflächenmerkmalen im Feedback von der Theorieverdrossenheit unabhängig ist, werden mit zunehmender Theorieverdrossenheit mehr personenbezogene Merkmale und weniger Tiefenstrukturen thematisiert.



Anmerkungen: Modellfit für das Gesamtmodell: Comparative Fit Index (CFI) = .92; Tucker-Lewis Index (TLI) = .91; Root Mean Square Error of Approximation (RMSEA) = .048; Standardized Root Mean Square Residual (SRMR) = .08

Abbildung 3: Strukturgleichungsmodell

Regression der Theorieverdrossenheit und der Häufigkeit der Feedback-Kriterien auf Personenvariablen

Die standardisierten Regressionskoeffizienten der Variablen Geschlecht, Alter und Berufserfahrung sind in Tabelle 3a dargestellt.

Tabelle 3a: Regression der Häufigkeit der Feedback-Kriterien auf Geschlecht, Alter und Berufserfahrung

		Geschlecht	Alter	Berufserfahrung
Häufigkeit Tiefenstrukturen	β	0.11	-0.31	0.60
	SE	0.12	0.02	0.13
	p	0.20	0.17	< 0.01
Häufigkeit Oberflächenmerkmale	β	0.12	-0.09	0.30
	SE	0.13	0.02	0.10
	p	0.22	0.62	0.09
Häufigkeit Personenbezogene Merkmale	β	-0.06	0.31	-0.20
	SE	0.15	0.02	0.12
	p	0.56	0.12	0.29
Theorieverdrossenheit	β	-0.20	0.09	-0.06
	SE	0.21	0.02	0.13
	p	0.11	0.62	0.73

Anmerkungen: n = 128. β , standardisierter Regressionskoeffizient; SE, Standardfehler; Geschlecht: 1: männlich; 2: weiblich

Zur Vermeidung der Multikollinearität wurde das Alter in Tabelle 3b aus der Regressionsanalyse ausgeschlossen

Tabelle 3b: Regression der Häufigkeit der Feedback-Kriterien auf Geschlecht und Berufserfahrung

		Geschlecht	Berufserfahrung
Häufigkeit Tiefenstrukturen	β	0.07	0.34
	SE	0.13	0.07
	p	0.44	< 0.01
Häufigkeit Oberflächenmerkmale	β	0.10	0.20
	SE	0.13	0.06
	p	0.33	0.06
Häufigkeit Personenbezogene Merkmale	β	-0.08	0.04
	SE	0.15	0.07
	p	0.41	0.68
Theorieverdrossenheit	β	-0.19	-0.01
	SE	0.21	0.06
	p	0.13	0.97

Anmerkungen: n = 128. β , standardisierter Regressionskoeffizient; SE, Standardfehler; Geschlecht: 1: männlich; 2: weiblich

4 Diskussion

Ziel der vorliegenden Studie war es, zu explorieren, auf welche Inhalte sich das Feedback von Lehrerbildner*innen an Lehramtsanwärter*innen bezieht und wie die Feedback-Kriterien mit subjektiven Vorstellungen von Unterrichtsqualität zusammenhängen. Zusätzlich wurde die Variable der Theorieverdrossenheit verwendet, um zu prüfen, wie negative Einstellungen gegenüber wissenschaftlichen Erkenntnissen sowohl mit den Vorstellungen von Unterrichtsqualität als auch mit den Inhalten des Feedbacks zusammenhängen.

Zur Beantwortung der ersten Forschungsfrage, welche Vorstellungen von Unterrichtsqualität Lehrerbildner*innen haben (F1), konnte anhand einer qualitativen Inhaltsanalyse gezeigt werden, dass die Befragten auf mannigfaltige Kriterien zur Beurteilung von Unterrichtsqualität zurückgreifen, wobei sich diese in ihren Antworten überproportional häufig auf Tiefenstrukturen bezogen. Insgesamt deuten die qualitativen Ergebnisse darauf hin, dass sich die subjektiven Vorstellungen der Lehrerbildner*innen zur Unterrichtsqualität *interpersonal* deutlich unterscheiden, dass es aber einen Kern von hauptsächlich tiefenstrukturellen Merkmalen gibt, zu dem ein weitgehender Konsens zu bestehen scheint.

Die Einschätzung der Lehrerbildner*innen darüber, wie häufig sie bestimmte Aspekte thematisieren, wenn sie Lehramtsanwärter*innen Feedback geben (F2) zeigte, dass Tiefenstrukturen am häufigsten als Kriterien herangezogen werden, während die personenbezogenen Merkmale am seltensten thematisiert werden. Da Tiefenstrukturen den Lernerfolg von Schüler*innen am besten vorhersagen (Praetorius et al., 2014; Seidel & Shavelson, 2007), kann es als günstig angesehen werden, dass diese auch am häufigsten als Feedback-Kriterien genannt werden. Bei den personenbezogenen Merkmalen wird das Merkmal der „ausgeprägten Lehrerpersönlichkeit“ abweichend von den restlichen Items innerhalb dieser Kategorie häufig thematisiert ($M = 3.94$; $SD = 1.02$), interessanterweise auch deutlich häufiger als verwandte Merkmale wie „Charisma“ und „natürliche Autorität“. Die „ausgeprägte Lehrerpersönlichkeit“ bezieht sich auf einen Bereich, der für Lehramtsanwärter*innen als Lernende wenig Informationen zur Veränderung bietet (Hattie & Timperley, 2007) und durch ein hohes Potential der Selbstwertbedrohung zu heftigen affektiven Reaktionen führen kann, die eine kognitive Reflexion verhindern (DeNisi & Kluger, 2000; Ilgen & Davis, 2000). Die hohe Häufigkeit der Thematisierung bestätigt bisherige Befunde zur Überschätzung dieses Merkmals für die erfolgreiche Berufsausübung durch Lehrpersonen (Bromme & Haag, 2004; Helmke, 2007; Herrmann & Hertrampf, 1997; Lipowsky, 2003), wobei die Lehrerpersönlichkeit weder klar operationalisierbar ist (Mayr & Neuweg, 2006), noch sich ein Effekt von Persönlichkeitsmerkmalen der Lehrperson auf den Lernerfolg von Schüler*innen nachweisen lässt (Klassen & Tze, 2014; Weinert & Helmke, 1996). Es muss

davon ausgegangen werden, dass die Betonung der Persönlichkeit als Feedback-Kriterium eine internal-stabile Attribution durch die Lehramtsanwärter*innen begünstigt und in Folge dessen der Motivation zur Verbesserung der Unterrichtsqualität eher im Wege steht.

Die theoriegeleitete Einteilung der Feedbackinhalte anhand der drei Bereiche ließ sich faktorenanalytisch in den Einschätzungen der Lehrerbildner*innen abbilden. Auf der Basis des bestätigten Messmodells wurde der Frage nachgegangen, wie die Feedbackinhalte mit den Vorstellungen von Unterrichtsqualität von Lehrerbildner*innen zusammenhängen (F3). Dabei zeigte sich, dass die Gewichtung der drei Bereiche Tiefenstrukturen, Oberflächenmerkmale und personenbezogene Merkmale, wenn Unterrichtsqualität frei definiert werden soll, und die Häufigkeit, mit welcher die Dimensionen im Feedback berücksichtigt werden, relativ unabhängig voneinander sind. Es besteht lediglich ein niedriger signifikant positiver Zusammenhang zwischen der Gewichtung und Häufigkeit von Tiefenstrukturen und ein signifikant negativer Zusammenhang zwischen der Gewichtung von Oberflächenmerkmalen und der Thematisierung von Tiefenstrukturen im Feedback (H3.1). Wie häufig Lehrerbildner*innen beim Geben von Feedback auf Oberflächenmerkmale und personenbezogene Merkmale eingehen, ist dagegen unabhängig davon, wie stark sie die verschiedenen Bereiche gewichten, wenn sie Unterrichtsqualität definieren sollen (H3.2 und H3.3).

Ein Zusammenhang zwischen Feedback-Inhalten und der Theorieverdrossenheit (F4) konnte für die Bereiche der Tiefenstrukturen und personenbezogenen Merkmale nachgewiesen werden. Je stärker Lehrerbildner*innen wissenschaftlichen Erkenntnissen gegenüber negativ eingestellt sind, desto häufiger thematisieren sie personenbezogene Merkmale (H4.3) und desto seltener thematisieren sie Tiefenstrukturen (H4.1) im Feedback. Dies ist insofern plausibel, da wissenschaftliche Erkenntnisse auf hohe Effekte der Tiefenstrukturen auf den Lernerfolg von Schüler*innen hinweisen, während kaum Hinweise auf Effekte durch personenbezogene Merkmale vorliegen. Eine signifikante Regression der Thematisierung von Oberflächenmerkmalen auf die Theorieverdrossenheit konnte nicht nachgewiesen werden (H4.2).

Die letzte Forschungsfrage bezog sich darauf, inwieweit die untersuchten Merkmale von Personenvariablen wie Geschlecht, Alter und Berufserfahrung abhängen (F5).

Es ließ sich keine signifikante Regression der Theorieverdrossenheit auf die Variablen Geschlecht (H5.1.1), Alter (H5.1.2) oder Berufserfahrung (H5.1.3) nachweisen. Die bisherige Forschung konnte zeigen, dass das Wissen von Lehrpersonen bezüglich wissenschaftlicher Inhalte relativ unabhängig von der Berufserfahrung ist (Brunner

et al., 2006). Unsere Befunde deuten darauf hin, dass dies auch für die Einstellungen gegenüber wissenschaftlichen Inhalten gilt.

Weiterhin konnte nachgewiesen werden, dass die Berufserfahrung zum Teil mit den Feedback-Inhalten zusammenhängt, während sich für Geschlecht (H5.2.1) und Alter (H5.2.2) keine solchen Zusammenhänge ergaben. Mit steigender Berufserfahrung werden Tiefenstrukturen häufiger thematisiert (H5.2.3).

Bei der Interpretation der Ergebnisse gilt es einige Einschränkungen der in diesem Beitrag beschriebenen Studie zu beachten: Das realisierte Querschnittsdesign erlaubt zwar Aussagen über Zusammenhänge, nicht aber kausale Rückschlüsse (Döring & Bortz, 2016). Weiterführende Untersuchungen zur Analyse von Ursache-Wirkungs-Zusammenhängen erscheinen sinnvoll, nicht zuletzt, um die Ergebnisse dieser Untersuchung zu replizieren und zu validieren. Obgleich als Stichprobe Personen mit einem großen Erfahrungsspektrum und aus mehreren Bundesländern rekrutiert werden konnten, ist zu konstatieren, dass aufgrund ihres geringen Umfangs und einer etwaigen positiven Selektion der Teilnehmenden Verzerrungen nicht ausgeschlossen werden können.

Insbesondere da es sich bei Lehrerbildner*innen um eine wenig erforschte Population handelt, sollte in Folgestudien der Frage nachgegangen werden, wie diese verstärkt zur Teilnahme motiviert werden können (z. B. durch angemessene Incentivierung). Auch wenn sowohl qualitative als quantitative Methoden eingesetzt wurden, um sich den Vorstellungen der Befragten zu nähern, kamen dabei ausschließlich Selbstberichtsverfahren zum Einsatz. Diese setzen voraus, dass die befragten Personen Fragen richtig verstehen, adäquate Introspektions- und Kommunikationsfähigkeiten besitzen und darauf antworten möchten (Rosen, Porter & Rogers, 2017). Um die Feedbackkultur in der zweiten Phase der Lehrerbildung vertiefend und multiperspektivisch zu erforschen, erscheint es aufbauend auf den Ergebnissen dieser Studie sinnvoll, diese mit den Einschätzungen von Lehramtsanwärter*innen abzugleichen.

Trotz der genannten Einschränkungen lassen sich aus der Studie erste Implikationen zum Feedback in der zweiten Phase der Lehrerbildung ableiten: Die Selbsteinschätzungen unserer Stichprobe bestätigen die Entkopplung der zweiten Phase der Lehrerbildung von wissenschaftlichen Erkenntnissen (vgl. Wenzl et al., 2017) nicht. Das Feedback von Lehrerbildner*innen orientiert sich zu einem großen Teil an Kriterien, die durch Forschung generiert wurden. Eine Einschränkung stellt das Merkmal der „ausgeprägten Lehrerpersönlichkeit“ dar, welches aus wissenschaftlicher Sicht nicht nur als unbedeutendes, sondern sogar als ungünstiges und professionalisierungshemmendes Feedback-Kriterium aufgefasst werden muss und trotzdem immer noch

häufig herangezogen wird. 5.3.2). Obwohl anzunehmen ist, dass die „Lehrerpersönlichkeit“ von vielen Lehrerbildner*innen nicht in einem rein psychologischen Sinne als Konfiguration stabiler zentraler Charaktereigenschaften (Allport, 1949), sondern eher als globalerer Begriff aufgefasst wird, der verschiedene Bereiche wie Fertigkeiten, Ethos und Haltung umschließt (Haag & Streber, 2020), birgt die Überbetonung dieses Konstruktes als Inhalt von Feedback die Gefahr, bei Lehramtsanwärter*innen den Eindruck zu erzeugen, die Qualität von Unterricht hänge weniger vom Verhalten in der individuellen Situation als vielmehr von dispositionellen Merkmalen ab.

Literaturverzeichnis

- Alger, C. L. (2009). Secondary teachers' conceptual metaphors of teaching and learning. *Teaching and Teacher Education*, 25 (5), 743-751.
- Allport, G. (1949). *Personality*. New York: Henry Holt Co.
- BAK (Bundesarbeitskreis der Seminar- und Fachleiter/innen e.V.) (2016). BAK Lehrerbildung. Verfügbar unter: http://bak-lehrerbildung.de/downloads2/2016_01_18_bak_flyer.pdf [21.06.2020]
- Bastian, J., Combe, A. & Reh, S. (2002). Professionalisierung und Schulentwicklung. *Zeitschrift für Erziehungswissenschaft*, 5 (3), 417-435.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469-520.
- Baumert, J. & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (S. 25-48). Boston, MA: Springer US.
- Behnke, K. (2016). *Umgang mit Feedback im Kontext Schule*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Bleck, V. & Lipowsky, F. (2020). Dröge, nutzlos, praxisfern? Wie verändert sich die Bewertung wissenschaftlicher Studieninhalte in Praxisphasen? In I. Ulrich & A. Gröschner (Hrsg.), *Praxissemester im Lehramtsstudium in Deutschland: Wirkungen auf Studierende* (S. 97-127). Wiesbaden: VS.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2008). *Kompetenzmessung bei angehenden Lehrerinnen und Lehrern*. Münster: Waxmann.
- Bromme, R. & Haag, L. (2008). Forschung zur Lehrerpersönlichkeit. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 803-819). Wiesbaden: VS.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit In: Bollen KA, Long JS, eds. *Testing Structural Equation Models*. Beverly Hills, CA: Sage, 136-162.
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T., Jordan, A., Klusmann, U., Tsai, Y.-M. & Neubrand, M. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, 9 (4), 521-544.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Creswell, J. & Plano Clark, V. (2018). *Designing and conducting mixed methods research*. Los Angeles: Sage.
- Deutsche Gesellschaft für Erziehungswissenschaft (DGfE) (2016). *Ethik-Kodex der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)*. Verfügbar unter: https://www.dgfe.de/fileadmin/OrdnerRedakteure/Satzung_etc/Ethikkodex_2016.pdf [28.03.2020].
- DeNisi, A. & Kluger, A. (2000). Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management Perspectives*, 14 (1), 129-139.
- Döbrich, P. & Abs, H. (2008). Evaluation der zweiten Phase der Lehrerbildung. *Schulverwaltung. Hessen, Rheinland-Pfalz*, 13 (3), 70-73.

- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin: Springer Berlin.
- Farrar, D. & Glauber, R. R. (1967). Multicollinearity in regression analysis. *The Review of Economics and Statistics* 49, (1), 92-107.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). *Student ratings of teaching quality in primary school*. 29, 1-9.
- Ferguson, R. F. & Danielson, C. (2014). How framework for teaching and tripod 7Cs evidence distinguish key components of effective teaching. In T. Kane, K. Kerr & R. Pianta (Hrsg.), *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (S. 98-143). San Francisco, CA.: Jossey-Bass, A Wiley Brand.
- Floyd, F. & Widaman, K. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7 (3), 286-299.
- Gärtner, H. (2013). *Wirksamkeit von Schülerfeedback als Instrument der Selbstevaluation von Unterricht*. In: J. Hense, S. Rädiker, W. Böttcher & T. Widmer (Hrsg.), *Forschung über Evaluation. Bedingungen, Prozesse und Wirkungen* (S. 107-124). Münster: Waxmann.
- Gärtner, H. & Vogt, A. (2013). Wie Lehrkräfte Ergebnisse eines Schülerfeedbacks verarbeiten und nutzen. *Unterrichtswissenschaft*, 41 (3), 252-267.
- Gillece, L. (2012). Teachers' pedagogical beliefs. In J. König (Hrsg.), *Teachers' pedagogical beliefs: definition and operationalisation - connections to knowledge and performance - development and change* (pp. 109-129). Münster: Waxmann.
- Haag, L., & Streber, D. (2020). *Lehrerpersönlichkeit*. Bad Heilbrunn: Klinkhardt.
- Hartinger, A., Kleckmann, T. & Hawelka, B. (2006). Der Einfluss von Lehrervorstellungen zum Lernen und Lehren auf die Gestaltung des Unterrichts und auf motivationale Schülervariablen. *Zeitschrift für Erziehungswissenschaft*, 9 (1), 110-126.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77 (1), 81-112.
- Helmke, A. (2017). Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts. Seelze-Velber: Klett/Kallmeyer.
- Herrmann, U. & Hertrampf, H. (1997). Reflektierte Berufserfahrung und subjektiver Qualifikationsbedarf. In S. Buchen, U. Carle & P. Döbrich (Hrsg.), *Jahrbuch für Lehrerforschung. Band 1*. (S. 139-163). Weinheim: Juventa.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Ilgen, D. & Davis, C. (2000). Bearing Bad News: Reactions to Negative Performance Feedback. *Applied Psychology*, 49 (3), 550-565.
- Imdad, M. & Aslam, M. (2018). mctest: Multicollinearity Diagnostic Measures. <https://CRAN.R-project.org/package=mctest>, R package version 1.3 [21.06.2020]
- Jöreskog, K. (1993). Testing structural equation models. In K. A. Bollen (Hrsg.), *Testing structural equation models* (pp. 294-316). Newbury Park: Sage.
- Klassen, R. & Tze, V. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12 (1), 59-76.
- Klein, L. (1962). *An Introduction to Econometrics*. New Jersey: Prentice-Hall.
- Klieme, E., Lipowsky, F. & Rakoczy, K. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. In M. Prenzel, L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127-146). Münster: Waxmann.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford.
- König, J. & Rothland, M. (2012). Motivations for choosing teaching as a career: effects on general pedagogical knowledge during initial teacher education. *Asia-Pacific Journal of Teacher Education*, 40 (3), 289-315.
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30 (3), 411-433.
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz.

- Künsting, J., Neuber, V. & Lipowsky, F. (2016). Teacher self-efficacy as a long-term predictor of instructional quality in the classroom. *European Journal of Psychology of Education*, 31 (3), 299-322.
- Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (S. 97-124). New York, NY: Springer.
- Landis, R. S., Edwards, B. D., & Cortina, J. M. (2009). On the practice of allowing correlated residuals among indicators in structural equation models. In C. E. Lance, R. J. Vandenberg (Hrsg.), *Statistical and methodological myths and urban legends* (S. 195-214). London: Routledge.
- Leuchter, M., Pauli, C., Reusser, K. & Lipowsky, F. (2006). Unterrichtsbezogene Überzeugungen und handlungsleitende Kognitionen von Lehrpersonen. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 562-579.
- Lipowsky, F. (2003). *Wege von der Hochschule in den Beruf*. Bad Heilbrunn/Obb.: Klinkhardt.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling*, 11(3), 320-341.
- Mayr, J. & Neuweg, G. (2006). Der Persönlichkeitsansatz in der Lehrer/innen/forschung. In M. Heinrich & U. Greiner (Hrsg.), *Schauen, was 'rauskommt. Kompetenzförderung, Evaluation und Systemsteuerung im Bildungswesen* (S. 183-206). Wien: Lit.
- Negrini, L. (2016). *Subjektive Überzeugungen von Berufsbildnern*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2-12.
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). *Generic dimensions of teaching quality*. *ZDM Mathematics Education*, 50 (3), 407-426.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: <https://www.R-project.org/> [21.06.2020]
- Reh, S. (2004). Abschied von der Profession, von Professionalität oder vom Professionellen?. *Zeitschrift für Pädagogik*, 50 (3), 358-372.
- Rosen, J., Porter, S. & Rogers, J. (2017). Understanding student self-reports of academic performance and course-taking behavior. *AERA Open*, 3 (2), 1-14.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48 (2), 1-36.
- Schaefer, C. (2002). *Forschung zur Lehrerbildung in Deutschland*. Schweizerische Zeitschrift für Bildungswissenschaften, 24 (1), 65-90.
- Schratz, M., Paseka, A. & Schrittemser, I. (2011). *Pädagogische Professionalität: Quer denken – umdenken – neu denken*. Wien: Facultas.wuv.
- Seidel, T. & Shavelson, R. (2007). Teaching effectiveness research in the past decade. *Review of Educational Research*, 77 (4), 454-499.
- Sembill, D. & Seifried, J. (2009). Konzeptionen, Funktionen und intentionale Veränderungen von Sichtweisen. In O. Zlatkin-Troitschanskaia (Hrsg.), *Lehrprofessionalität: Bedingungen, Genese, Wirkungen und ihre Messung* (S. 345-354). Weinheim: Beltz.
- Slavin, R. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21 (2), 141-157.
- Staub, F. & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains. *Journal of Educational Psychology*, 94 (2), 344-355.
- Strietholt, R. (2009). Referendare beurteilen. *Zeitschrift für Pädagogik*, 55 (4), 622-645.
- Sykes, G., Bird, T. & Kennedy, M. (2010). Teacher education: Its problems and some prospects. *Journal of Teacher Education*, 61 (5), 464-476.
- Terhart, E. (2000). *Perspektiven der Lehrerbildung in Deutschland: Abschlussbericht der von der Kultusministerkonferenz eingesetzten Kommission*. Weinheim: Beltz.
- Theil, H. (1971). *Principles of Econometrics*. New York: John Wiley & Sons.

- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B. & Trautwein, U. (2016). Student and teacher ratings of instructional quality. *Journal of Educational Psychology*, 108 (5), 705-721.
- Wahl, D. (2013). *Lernumgebungen erfolgreich gestalten*. Bad Heilbrunn: Klinkhardt.
- Wang, J., Lin, E., Spalding, E., Klecka, C. L. & Odell, S. (2011). Quality teaching and teacher education. *Journal of Teacher Education*, 62 (4), 331-338.
- Weinert, F. & Helmke, A. (1996). *Der gute Lehrer: Person, Funktion oder Fiktion?* In A. Leschinsky (Hrsg.), *Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule* (S. 223-233). Weinheim: Beltz.
- Wenzl, T., Wernet, A. & Kollmer, I. (2018). *Praxisparolen*. Wiesbaden: Springer.
- Wilson, S., Rozelle, J. & Mikeska, J. (2011). Cacophony or embarrassment of riches. *Journal of Teacher Education*, 62 (4), 383-394.
- Wisniewski, B. & Zierer, K. (2020). Entwicklung eines Online-Fragebogens zur Erhebung von Unterrichtsqualität durch Lernendenfeedback und erste Validierungsschritte. *Psychologie in Erziehung und Unterricht*, 22 (2), 138-155.
- Wisniewski, B., Zierer, K., Dresel, M., & Daumiller, M. (2020). Obtaining secondary students' perceptions of instructional quality: Two-level structure and measurement invariance. *Learning and Instruction*, 66, 101303.

Autoren

Dr. Benedikt Wisniewski, Universität Augsburg, Lehrstuhl für Schulpädagogik
Stefan T. Siegel, Universität Augsburg, Lehrstuhl für Allgemeine Pädagogik
Korrespondenz an: benedikt.wisniewski@phil.uni-augsburg.de