



# Ein neues Instrument zur Erfassung der Einstellungen zu Portfolioarbeit in der Lehrerinnen- und Lehrerbildung (EPP-L)

Tim Fütterer · Lina Feder · Colin Cramer

Eingegangen: 6. November 2021 / Überarbeitet: 9. November 2022 / Angenommen: 16. November 2022 /  
Online publiziert: 16. Dezember 2022  
© Der/die Autor(en) 2022

**Zusammenfassung** Portfolioarbeit ist aufgrund von Potenzialen für Professionalisierungsprozesse (z. B. Entwicklung von Reflexionsfähigkeit) ein weit verbreitetes und teilweise sogar verbindlich vorgeschriebenes Konzept in der Lehrerinnen- und Lehrerbildung. Für die Wirksamkeit einer Portfolioarbeit (d. h. für die Entfaltung der Potenziale) erweisen sich positive Einstellungen der Anwenderinnen und Anwender zu Portfolioarbeit als bedeutsam. Bislang liegen allerdings keine Befunde zu Einstellungen zu Portfolioarbeit vor, die über Spezifika vereinzelter, standortspezifischer Portfoliokonzeptionen hinausreichen. Ein Grund hierfür ist, dass Instrumente zur Erfassung von Einstellungen bislang nicht konzeptionsübergreifend nutzbar sind. Dieser Beitrag stellt ein neues Instrument zur Erfassung von Einstellungen von Lehramtsstudierenden zu Portfolioarbeit (EPP-L) vor, das unabhängig von spezifischen Konzeptionen eingesetzt werden kann. Ausgehend von einem Pool aus 45 Items, die auf Basis theoretisch abgeleiteter und systematisch identifizierter Potenziale von Portfolioarbeit entwickelt wurden, konnten in einem iterativen Prozess (z. B. kognitive Pretests, Expertenratings) 18 finale Items entwickelt werden. Die psychometrische Überprüfung des EPP-L erfolgte anhand von  $N_1=424$  und  $N_2=374$  Lehramtsstudierenden. Der EPP-L erweist sich als reliabel (interne Konsis-

---

Die Daten sowie die vollständigen reproduzierbaren Codes, die für die Analysen in dieser Studie verwendet wurden, sind auf begründete Anfrage beim korrespondierenden Autor erhältlich. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin und den Autoren.

✉ Tim Fütterer  
Hector-Institut für Empirische Bildungsforschung, Universität Tübingen,  
Europastraße 6, 72072 Tübingen, Deutschland  
E-Mail: [tim.fuetterer@uni-tuebingen.de](mailto:tim.fuetterer@uni-tuebingen.de)

Tim Fütterer · Lina Feder · Colin Cramer  
Tübingen School of Education, Universität Tübingen, Tübingen, Deutschland

Lina Feder · Colin Cramer  
Institut für Erziehungswissenschaft, Universität Tübingen, Tübingen, Deutschland

tenz) und valide (Konstrukt- und Kriteriumsvalidität). Theoriekonform zeigen sich starke Zusammenhänge der Einstellungen mit dem wahrgenommenen Nutzen von Portfolioarbeit und keine bis schwache Zusammenhänge mit konstruktivistischen und transmissiven lerntheoretischen Überzeugungen. Differential Item Functioning Analysen zeigen die Messinvarianz des EPP-L über die Variable Geschlecht. Konfirmatorische Faktorenanalysen bestätigen einen Generalfaktor. Der EPP-L bietet die Möglichkeit, Evidenz zur Portfolioarbeit in der Lehrerinnen- und Lehrerbildung über vereinzelte Programmevaluationen hinaus zu generieren.

**Schlüsselwörter** Portfolioarbeit · Einstellungen · Lehrerinnen- und Lehrerbildung · Instrument · Validierung

### **A new questionnaire to assess the attitudes towards portfolio work in teacher education (EPP-L)**

**Abstract** Portfolio work is a widespread and sometimes even mandatory concept in teacher education due to its potentials for professionalization processes (e.g., development of reflective skills). For the effectiveness of portfolio work (i.e., for the development of its potentials), positive attitudes of users toward portfolio work were shown to be important. However, so far, no findings exist on attitudes toward portfolio work that go beyond specifics of isolated, location-specific portfolio concepts. One reason for the lack of findings is that instruments to capture attitudes cannot be used across concepts. This paper presents a new instrument to capture pre-service teachers' attitudes toward portfolio work (EPP-L) that can be used independently of specific portfolio conceptions. Based on a pool of 45 items developed along theoretically derived and systematically identified potentials of portfolio work, 18 items were developed in an iterative process (e.g., cognitive pretests, expert ratings). Psychometric testing of the EPP-L was conducted using  $N_1=424$  and  $N_2=374$  pre-service teachers. The EPP-L proved to be reliable (internal consistency) and valid (construct and criterion validity). Consistent with theory, for example, attitudes showed strong correlations with perceived usefulness of portfolio work and non to weak correlations with constructivist and transmissive learning theory beliefs. Differential item functioning analyses showed that the EPP-L was measurement invariant across the variable gender. Confirmatory factor analyses confirmed a general factor. The EPP-L provides the first opportunity to generate evidence on portfolio work in teacher education beyond isolated program evaluations.

**Keywords** Portfolio work · Attitudes · Teacher education · Instrument · Validation

Portfolioarbeit ist in der Ausbildung von Lehrpersonen weit verbreitet und gilt als relevantes Konzept in der Lehrerinnen- und Lehrerbildung. So wird der Portfolioarbeit national wie international ein hohes Innovationspotenzial zugeschrieben (Barton und Collins 1993; Feder und Cramer 2019; Fütterer 2019; Gläser-Zikuda und Hascher 2007; Hascher und Sonntagbauer 2013; Strudler und Wetzel 2008). Zum Beispiel wird konstatiert, dass Portfolioarbeit die Reflexivität angehender Lehrpersonen för-

dert, indem berufsrelevante Handlungssituationen auf der Grundlage von Professionswissen bearbeitet werden (Borko et al. 1997; Gläser-Zikuda 2009; Häcker 2012). Allerdings wird argumentiert, dass sich die Potenziale nur bei der Sicherstellung von Gelingensbedingungen entfalten (Fütterer 2019), etwa wenn die Anwenderinnen und Anwender positive Einstellungen zu Portfolioarbeit haben (Imhof und Picard 2009; Ziegelbauer 2016). In dieser Arbeit werden Einstellungen als evaluative Reaktionen auf Einstellungsobjekte (hier: Portfolioarbeit) definiert (Ajzen 2001; Albarracín et al. 2018) und durch das Ausmaß an Zustimmung oder Ablehnung ausgedrückt (Kopietz und Echterhoff 2016). Nicht berücksichtigt werden in dieser Studie affektive Aspekte von Einstellungen (d.h. Gefühle oder Emotionen).

Einstellungen sind mit Ausprägungen menschlichen Handelns verknüpft (Ajzen 1987; Glasman und Albarracín 2006): Für die Portfolioarbeit wird daher angenommen, dass Einstellungen zu Portfolioarbeit mit der Arbeitsweise von angehenden Lehrpersonen bei der Bearbeitung (z. B. Kontinuität) von Portfolios zusammenhängen. Weil Einstellungen mittels Selbsteinschätzungen niedrigschwellig zu erfassen sind und deren Erhebung zugleich den ersten Schritt einer Evaluationsmaßnahme kennzeichnet (Kirkpatrick und Kirkpatrick 2006), existieren in diesem Bereich die meisten empirischen Befunde in der Portfolioforschung (Feder und Cramer 2019; Hofmann et al. 2016). In deren Zusammenschau zeigen angehende Lehrpersonen häufig negative Einstellungen zur Portfolioarbeit, wobei die Befundlage heterogen ist. Je nach Standort wurden positive (z. B. Christen und Hofmann 2013) oder negative Einstellungen (z. B. Streblov et al. 2013) festgestellt. Derartige Befunde sind allerdings kaum vergleichbar, da die Einstellungen unterschiedlich und meist unter Bezugnahme auf standortspezifische Merkmale der Portfoliokonzeptionen operationalisiert wurden (Gläser-Zikuda et al. 2020). Validierte Instrumente wurden bislang kaum eingesetzt. Angesichts des Mangels an geteilten Kerncharakteristika von Portfolioarbeit überrascht die Heterogenität der Operationalisierungen nicht. Für die Lehrerinnen- und Lehrerbildung ergibt sich aus dem Fehlen von Instrumenten für eine standardisierte Erfassung von Einstellungen zur Portfolioarbeit, dass sich bisher kaum abschätzen lässt, ob die der Portfolioarbeit zugesprochenen Potenziale beispielsweise hinsichtlich der Reflexionsfähigkeit von (angehenden) Lehrpersonen tatsächlich Wirkung entfalten. In diesem Zusammenhang bietet die systematische Identifikation grundlegender Potenziale von Portfolioarbeit (Feder und Cramer 2018) erstmalig die Möglichkeit, ein Instrument zur Erfassung von Einstellungen zu Portfolioarbeit standortunabhängig zu entwickeln. Im Beitrag wird deshalb in einem ersten Schritt die Entwicklung eines solchen Instrumentes dokumentiert. Das Instrument wurde basierend auf einer Vorstudie (Feder et al. 2021) iterativ über Pretests entwickelt. In einem zweiten Schritt werden zwei Validierungsstudien mit  $N_1=424$  und  $N_2=374$  Lehramtsstudierenden zur Prüfung der Reliabilität, Konstruktvalidität, Kriteriumsvalidität und geschlechtsspezifischen Testfairness des finalen Instruments vorgestellt.

# 1 Theoretischer Hintergrund

## 1.1 Portfolioarbeit und deren Potenziale

Portfolioarbeit ist in der deutschsprachigen Ausbildung von Lehrpersonen weit verbreitet und in einigen Bundesländern der Bundesrepublik Deutschland sogar verbindlich vorgeschrieben (Monitor-Lehrerbildung 2019). Obwohl eine spezifische Definition von Portfolioarbeit weder sinnvoll noch möglich ist, gelten zum einen das Sammeln von Dokumenten eigener Leistungen und zum anderen die reflexive Bezugnahme auf diese Dokumente als zentrale Merkmale von Portfolioarbeit (Fütterer 2019). Während Portfolioarbeit in Deutschland überwiegend in der ersten (Studium) und zweiten Phase (Referendariat) etabliert ist, kommt sie in der dritten Phase (berufsbegleitende Fort- und Weiterbildung) seltener zum Einsatz (Gläser-Zikuda et al. 2020). Der Begriff Portfolioarbeit vereint eine Vielzahl höchst unterschiedlicher Konzepte (Gläser-Zikuda et al. 2020; Häcker 2012). Portfoliokonzepte unterscheiden sich etwa im Hinblick auf den zeitlichen Umfang der Portfolioarbeit oder den Grad der institutionellen Verbindlichkeit (Häcker und Lissmann 2007). Die verschiedenen Konzepte adressieren unterschiedliche Potenziale von Portfolioarbeit, die je nach Anwendungskontext stark divergieren (Koch-Priewe 2013); teilweise werden sogar gegenläufige Potenziale adressiert (z. B. Reflexivitätsförderung versus Leistungsmessung). In einer systematischen Inhaltsanalyse deutschsprachiger Literatur zur Portfolioarbeit in der Ausbildung von Lehrpersonen haben Feder und Cramer (2018) neun empirisch voneinander abgrenzbare Potenziale herausgearbeitet.

Die Potenziale beziehen sich auf die Unterstützung der individuellen professionellen Entwicklung (angehender) Lehrpersonen (Aspekt *Professionalisierungsinstrument*) oder auf die Optimierung von Strukturen lehrerbildender Studiengänge (Aspekt *Strukturinstrument*). Diese Potenziale werden je nach Standort und Portfoliokonzeption unterschiedlich stark adressiert und können Ausgangspunkt für eine konzeptionsübergreifende Analyse von Einstellungen zu Portfolioarbeit sein.

## 1.2 Einstellungen zu Portfolioarbeit

Für die Entfaltung von Potenzialen einer Portfolioarbeit gelten positive Einstellungen der Anwenderinnen und Anwender zur Portfolioarbeit als zentrale Voraussetzung (Brouër 2005, 2007; Fütterer 2019; Hascher und Sonntagbauer 2013; Imhof und Picard 2009; Klampfer 2013; Ziegelbauer 2016; für Einstellungen allgemein siehe: Ajzen 1987; Ajzen und Fishbein 1980). Entsprechend wird im Rahmen von Evaluationsmaßnahmen vorgeschlagen, zunächst Einstellungen der Anwenderinnen und Anwender zu erfassen (Kirkpatrick und Kirkpatrick 2006). Zur Portfolioarbeit liegen bereits Befunde zu den Einstellungen von Anwenderinnen und Anwendern vor (vgl. Bryant und Chittum 2013; Feder und Cramer 2019; Feder et al. 2021; Hofmann et al. 2016), die gemischt ausfallen. Einerseits zeigen sich vereinzelt positive Einstellungen (z. B. Artmann und Herzmann 2016; Christen und Hofmann 2013), andererseits legen Befunde aus Reviews nahe, dass angehende Lehrpersonen eher negative Einstellungen zu Konzepten der Portfolioarbeit haben (Cakir und Balcikanli 2012; Feder und Cramer 2019; Feder et al. 2021; Hofmann et al. 2016).

Zum Beispiel zeigt sich bei Evaluationen standortbezogener Portfoliokonzepte teilweise deutliche Ablehnung (z. B. Pineker und Störtländer 2013; Streblov et al. 2013). Allerdings resultieren die Befunde überwiegend aus Konzeptevaluationen einzelner Hochschulen, weshalb der Vergleich bisheriger Befunde kaum möglich ist (Feder et al. 2019). Unter anderem wegen der konzeptionspezifischen Forschung zu Portfolioarbeit und der daraus resultierenden Schwierigkeit, Befunde miteinander vergleichen zu können, liegen erwartungskonform kaum umfassende Überblicksarbeiten zur empirischen Befundlage zu Portfolioarbeit vor und insgesamt wird ein „großes Forschungsdefizit“ (Hofmann et al. 2016, S. 35) attestiert. Das in dieser Studie vorgestellte Instrument nimmt dieses Desiderat und bietet die Möglichkeit, über einzelne Programmevaluationen hinaus Erkenntnisse zur Portfolioarbeit und dessen Steuerung und Entwicklung zu gewinnen. Dadurch eröffnen sich erste Ansatzpunkte, um Chancen und Grenzen von Portfolioarbeit in der Lehrerinnen- und Lehrerbildung kritisch-konstruktiv zu prüfen.

### 1.3 Operationalisierung der Einstellungen zu Portfolioarbeit

Die Messung von Einstellungen zur Portfolioarbeit erfolgte bislang überwiegend mittels (Gruppen)Interviews und -diskussionen (Artmann und Herzmann 2016; Ziegelbauer und Gläser-Zikuda 2016) sowie Fragebogenerhebungen (z. B. Brouër 2007; Fütterer 2019; Pineker und Störtländer 2013; Schlag 2015; Streblov et al. 2013; Ziegelbauer 2016). Wenn eine standardisierte Erfassung mittels Fragebogen erfolgte, wurden Einstellungen unterschiedlich definiert und unter Verwendung verwandter Konstrukte operationalisiert. So bewerteten etwa in der Studie von Streblov et al. (2013) Lehramtsstudierende anhand von 20 Items den Nutzen der Portfolioarbeit in der ersten Phase u. a. bezogen auf die dortige, standortspezifische Handreichung. Fütterer (2019) untersuchte Einstellungen zur Portfolioarbeit ebenfalls über die Bewertung ihres Nutzens in der zweiten Phase der Lehrerinnen- und Lehrerbildung, u. a. bezüglich Phasenspezifika. Unter Diplomhandelslehramtsstudierenden wurde ein Fragebogen zur Akzeptanz der Portfolioarbeit eingesetzt, der auf den spezifischen Anwendungskontext abzielt (Brouër 2007; Brouër und Gläser-Zikuda 2010). Schlag (2015) untersuchte mit einer Skala aus sechs Items die Akzeptanz der portfoliobasierten Dokumentation eines Praktikums u. a. unter Rekurs auf das dort bereitgestellte Material. Diese und weitere Instrumente sind mit Blick auf ihre Entwicklung wenig dokumentiert, stark konzept- bzw. standortspezifisch und nicht konsequent validiert (z. B. ohne Rekurs auf Gütekriterien). Zusammenfassend existiert national und international kein theoretisch fundiertes, validiertes Instrument.

### 1.4 Zielsetzungen und Annahmen

Das Ziel dieses Beitrages ist, ein reliables, valides und bezüglich Geschlechtsgruppen faires Instrument zur Erfassung von Einstellungen von Lehramtsstudierenden zu den systematisch identifizierten Potenzialen von Portfolioarbeit zu entwickeln. Ein reliables und valides Instrument zur Erfassung von Einstellungen zu Portfolioarbeit unabhängig von standortspezifischen Portfoliokonzeptionen erscheint besonders hinsichtlich der Vergleichbarkeit von Befunden aus unterschiedlichen Studien zur

Portfolioarbeit notwendig, um einen ersten Schritt für belastbarere Aussagen zur Tragfähigkeit des Einsatzes von Portfolios in der Lehrerinnen- und Lehrerbildung über Hochschulen hinweg zu ermöglichen. Ein bezüglich Geschlechtergruppen faires Instrument ist besonders deshalb wichtig, weil sich in aktuellen Studien gezeigt hat, dass sich angehende Lehrpersonen hinsichtlich verschiedener Personenmerkmale (z. B. Selbstwirksamkeit, Selbstreflexion, Enthusiasmus) unterscheiden (Holzberger et al. 2021). Zudem zeigen Frauen und Männer konkret bei der Portfolioarbeit unterschiedliche Arbeitsweisen (z. B. bezüglich des Umfangs ihrer Reflexionen oder ihrer Gewissenhaftigkeit bei der Bearbeitung; Fütterer 2019). Auch zeigte sich, dass Nutzerinnen der Portfolioarbeit ein höheres Potenzial zuschreiben (Feder et al. 2021), die Portfolios besser organisieren (Dreyer 2015), mehr Artefakte einbinden (Taylor et al. 2012) oder längere Texte zu Reflexionsaufgaben schreiben (Fütterer 2019). Das potenzialbezogene Instrument zu Einstellungen zur Portfolioarbeit sollte demzufolge sensitiv für potenzielle Geschlechtsunterschiede sein. Zugleich sind Portfolios im Studium Werkzeuge, um Lernprozesse zu organisieren. Auch diesbezüglich unterscheiden sich Frauen und Männer, wenn sich etwa Frauen in Lernprozessen mehr mit anderen Lernenden austauschen (z. B. Caspi et al. 2008) und Austauschprozesse lernwirksamer nutzen, indem sie sich stärker argumentativ in Diskussionen einbringen und eher Feedback von anderen annehmen (z. B. Noroozi et al. 2022). Auch unterscheiden sich Frauen und Männer bezüglich der intellektuellen Risikobereitschaft (d. h. das Ausmaß, in welchem im akademischen Kontext Situationen [wie beispielsweise Prüfungssituationen] aufgesucht werden, in welchen man mangelnde Fähigkeiten offenbaren könnte; Byrnes et al. 1999) oder bezüglich ihrer strategischen Ansätze im Umgang mit Stress im akademischen Kontext (z. B. Bonneville-Roussy et al. 2017). Für die Erfassung der Einstellung angehender Lehrpersonen zu Portfolioarbeit bedeutet dies, dass Frauen und Männer mit derselben Einstellung zu Portfolioarbeit beispielsweise auf ein Item zum Potenzial von Portfolios für die Gestaltung von Austausch gleich antworten sollten, doch Frauen machen generell positivere Erfahrungen mit dem Austausch beim Lernen. Es muss also sichergestellt werden, dass sich im Antwortverhalten der Geschlechtergruppen tatsächlich die Einstellung zu Portfolioarbeit ausdrückt. Die Items des Instruments müssen daher so gestaltet sein, dass kein Geschlecht diskriminiert wird (Testfairness; Chen et al. 2017; De Ayala 2009; Jin und Chen 2020).

Wir gehen von einer umso positiveren Einstellung zur Portfolioarbeit aus, je mehr Potenzial dieser insgesamt zugeschrieben wird. Entsprechend der unterschiedlichen Potenziale von Portfolioarbeit (Feder und Cramer 2019) und den empirischen Einsichten aus korrelativen Vorstudien zum Zusammenhang dieser Facetten (Feder et al. 2021) kann nicht von einem eindimensionalen Konstrukt *Einstellung zu Portfolioarbeit* ausgegangen werden. Die Berücksichtigung von Einstellungen zu unterschiedlichen Aspekten einer Portfolioarbeit erscheint zugleich mit Blick auf Erkenntnisse aus der Einstellungsforschung plausibel, wonach Personen gleichzeitig mehrere und sogar ambivalente Einstellungen zum selben Einstellungsobjekt (hier: Portfolioarbeit) haben können (Ajzen 2001), welche die Vorhersagekraft von Einstellungen auf tatsächlich realisiertes Verhalten moderieren können (Glasman und Albarracín 2006). Um in dieser Studie eine grundlegende Einstellung zu Portfolioarbeit abbilden zu können und zugleich unterschiedlich ausgeprägte Einstellungen zu

verschiedenen Aspekten (d. h. Potenzialen) von Portfolioarbeit zuzulassen, nehmen wir an, dass eine Bi-Faktor-Struktur oder ein Modell zweiter Ordnung geeignet ist, um die Beziehung zwischen den Potenzialen und das übergeordnete Konstrukt (g-Faktor) abzubilden. Folglich sollten einerseits die expliziten Einstellungen zu den Potenzialen von Portfolioarbeit (Einstellungsobjekte) jeweils als eindimensionale Konstrukte sowie andererseits das übergeordnete Konstrukt (g-Faktor) *Einstellung zu Portfolioarbeit* erfasst werden (Tab. 1).

Reliabel ist das Instrument, wenn es eine hohe interne Konsistenz zeigt. Valide ist das Instrument, wenn sich starke Zusammenhänge des g-Faktors mit verwandten (konvergente Validität als Aspekt der Konstruktvalidität) bzw. nicht-verwandten Konstrukten (divergente Validität) sowie externen Kriterien (Kriteriumsvalidität) zeigen. Auf Grundlage des Forschungsstandes wurden für die Validierungsstudie erwartete Zusammenhänge zwischen dem g-Faktor (Einstellungen zu Portfolioarbeit) und anderen Konstrukten formuliert (Axiome: Hartig et al. 2012).

Da Einstellungen zu Portfolioarbeit bisher häufig im Sinne deren Akzeptanz operationalisiert wurden (Klampfer 2013; Schlag 2015; Ziegelbauer 2016), erwarten wir positive Zusammenhänge zwischen den operationalisierten, potenzialbezogenen Einstellungen und den bereits vorliegenden Akzeptanzskalen zu Portfolioarbeit (konvergente Validität). Portfolioarbeit in der Lehrerinnen- und Lehrerbildung zielt u. a. darauf ab, dass angehende Lehrpersonen ihren eigenen Lernprozess reflektieren.

**Tab. 1** Kategorien der Potenziale von Portfolioarbeit

Hauptkategorie	Kategorie	Beschreibung
Professionalisierungsinstrument	(1) Reflexivität	Portfolioarbeit führt zu einer verstärkten Reflexivität angehender Lehrpersonen
	(2) Berufsbezogene Fähigkeiten	Portfolioarbeit unterstützt die Entwicklung individueller berufsbezogener Fähigkeiten angehender Lehrpersonen, etwa in Bezug auf die berufsbezogene Kompetenzentwicklung
	(3) Selbstständigkeit	Portfolioarbeit trägt zu einer stärkeren Selbstbestimmung, Selbststeuerung und Verantwortung angehender Lehrpersonen bei und fördert so die Selbstständigkeit
	(4) Methode Portfolio	Portfolioarbeit trägt zur Entwicklung der Methodenkompetenz angehender Lehrpersonen zum „Portfolio“ bei
Strukturinstrument	(5) Innovierung der Lehrerinnen- und Lehrerbildung	Portfolioarbeit trägt zu einer Veränderung der lehrerbildenden Studiengänge bei, indem es z. B. (verbindlich) als Reforminstrument eingesetzt wird oder zur Generierung von Steuerungswissen dienen kann
	(6) Sammlung und Dokumentation	Portfolioarbeit bietet einen Ort, an dem systematisch über die Ausbildung hinweg Dokumente und Artefakte gesammelt werden können
	(7) Begleitung und Austausch	Portfolioarbeit trägt zu einem verstärkten Austausch zwischen Dozentinnen und Dozenten und Studentinnen und Studenten als auch zwischen Studentinnen und Studenten bei
	(8) Leistungsmessung	Portfolioarbeit wird zur Leistungsmessung eingesetzt
	(9) Brückenfunktion	Portfolioarbeit bietet die Möglichkeit, unterschiedliche Ausbildungsbestandteile (z. B. auf institutioneller Ebene) aufeinander zu beziehen

Die Tabelle ist übernommen von Feder und Cramer (2018).

Nehmen Lehrpersonen jedoch nicht ihren eigenen Lernprozess in den Blick, sondern den Lernprozess ihrer Schülerinnen und Schüler, liegt ein Perspektivwechsel vor, bei welchem sich Einstellungen entwickeln können. Einstellungen zur Portfolioarbeit, in welchen der eigene Lernprozess reflektiert wird, und Einstellungen zu Lernprozessen von Schülerinnen und Schülern, stellen unterschiedliche Konstrukte dar. Aus diesem Grund nehmen wir an, dass sich keine bis geringe Zusammenhänge zwischen potenzialbezogenen Einstellungen und konstruktivistischen und transmissiven lerntheoretischen Überzeugungen (divergente Validität) zeigen. Da Einstellungen eng mit motivationalen Konstrukten zusammenhängen (Fütterer 2019; Klampfer 2013; Ziegelbauer et al. 2013), erwarten wir zudem positive Zusammenhänge mit dem Interesse, dem wahrgenommenen Nutzen sowie der intrinsischen Motivation bezüglich Portfolioarbeit (Kriteriumsvalidität).

## 2 Methode

Das Instrument (EPP-L) wurde in einem iterativen Verfahren zunächst sowohl theoretisch fundiert als auch empirisch konstruiert (rationale und induktive Testkonstruktion: Bühner 2011) und anschließend empirisch validiert.

### 2.1 Konstruktion des Instrumentes

Zu Beginn der Konstruktion des Instrumentes stand die Inhaltsvalidität im Fokus. Zu jedem der theoretisch und systematisch identifizierten Potenziale (Feder et al. 2019) wurden fünf Items entwickelt. Die Iteminhalte orientierten sich dabei an den Kategorienbezeichnungen, Beschreibungen und Ankerbeispielen der theoretischen Vorarbeiten (Feder und Cramer 2018, 2019). Der anfängliche Itempool bestand aus 45 Items, die sich auf 9 Inhaltsaspekte verteilen (Beispielitem: „Ich denke, dass Portfolios zur Reflexion anregen“; Onlinematerial 1). Alle Items wurden mittels einer fünfstufigen Antwortskala erhoben (1 = *trifft überhaupt nicht zu* bis 5 = *trifft voll und ganz zu*), wobei eine hohe Ausprägung eine hohe Potenzialzuschreibung und damit eine positive Einstellung ausdrückte. Hinsichtlich der Zumutbarkeit (z. B. zeitliche Belastung von Testpersonen) und Testökonomie (z. B. Kosten und Dauer zukünftiger Erhebungen), wurde eine Reduktion der Itemanzahl so angestrebt, dass jedes Potenzial durch drei Items jeweils eindimensional erfasst wird.

Um sicher zu stellen, dass die Items tatsächlich die neun Inhaltsaspekte repräsentieren, wurden *erstens* die 45 Items fünf Expertinnen und Experten für Portfolioarbeit zur kritischen Beurteilung vorgelegt (Onlinematerial 2). Um *zweitens* die Verständlichkeit der Items und Probleme bei der Bearbeitung aus Anwendungsperspektive zu gewährleisten, wurden mit sieben Lehramtsstudierenden kognitive Tests (Lenzner et al. 2015) mittels der Methode des lauten Denkens durchgeführt (vgl. Bühner 2011; Moosbrugger und Kelava 2012). *Drittens* wurden die 45 Items einem Pretest unter  $N_{P1} = 183$  Lehramtsstudierenden unterzogen (Alter in Jahren:  $M = 20.84$ ,  $SD = 2.59$ ; 68 % weiblich, 31 % männlich; 72 % im 1. Fachsemester, 7 % im 2. Fachsemester, 13 % im 3. Fachsemester), um anhand explorativer Faktorenanalysen (EFA; Rotation = Geomin [Robustheitsanalyse: Rotation = Varimax], Schätzer = Maximum Likeli-

hood) Informationen zu empirisch zugrundeliegenden Faktoren und entsprechenden Ladungsmustern der Items zu erhalten (faktorielle Validität; Tab. 9 und 10 im Anhang). Entscheidungen zum Ein- oder Ausschluss von Items wurden primär auf der Grundlage der Ladungsmuster getroffen. Zudem sollten Items unterschiedlicher Schwierigkeit ( $0,20 \leq p_i \leq 0,80$ ) beibehalten werden, wobei mittelschwierige Items ( $0,45 \leq p_i \leq 0,55$ ) überrepräsentiert sein sollen (Bühner 2011; Kelava und Moosbrugger 2012). Zudem wurden Items aufgrund der Einschätzungen der Expertinnen und Experten aus inhaltlichen Gründen nicht weiter einbezogen. Im Zweifelsfall wurden inhaltliche Argumente bei der Itemselektion stärker gewichtet als Statistiken. Nach-

**Tab. 2** Überblick über die Items des Instrumentes EPP-L

PiF	Item EPP-L	Itemformulierung
		Inwieweit stimmen Sie den folgenden Aussagen zu? Ich traue Portfolioarbeit zu, dass sie ...
<i>Entwicklung von Reflexivität</i>		
1	9	Das Nachdenken über Erfahrungen in der Praxis aus unterschiedlichen Blickwinkeln vertieft
7	5	Das Nachdenken über den persönlichen Entwicklungsprozess vertieft
13	17	Die Reflexionsfähigkeit fördert
<i>Entwicklung von Selbstständigkeit</i>		
2	11	Die Eigenständigkeit fördert
8	7	Die Fähigkeit zum selbstregulierten Lernen fördert
14	19	Die Fähigkeit fördert, eigenverantwortlich zu agieren
<i>Möglichkeit der Innovierung der Ausbildung</i>		
3	26	Die Lehrerbildung reformiert
9	30 <sup>a</sup>	Die Lehrerbildung reformiert, indem sie z. B. unterschiedliche Ausbildungsbestandteile (z. B. Bildungswissenschaften, Fachwissenschaften und Fachdidaktiken) verbindet
15	35 <sup>a</sup>	Die Lehrerbildung reformiert, indem sie z. B. die Phasen der Lehrerbildung (Studium und Referendariat) verbindet
<i>Möglichkeit des Sammelns und der Dokumentation</i>		
4	42 <sup>a</sup>	Die Grundlage zur Präsentation individueller Kompetenzen in einer Bewerbung bildet
10	37	Die Grundlage für eine Bewerbung bildet
16	32 <sup>a</sup>	Die Grundlage zur Präsentation erarbeiteter Inhalte (z. B. in einer Bewerbung) bildet
<i>Möglichkeit der Begleitung und des Austauschs</i>		
5	23 <sup>a</sup>	Den Austausch zwischen den Lernenden fördert
11	28	Die Kooperation zwischen den Lernenden fördert
17	33	Die Kommunikation zwischen Lernenden fördert
<i>Möglichkeit der Leistungsmessung</i>		
6	24 <sup>a</sup>	für Prozesse der Leistungsbeurteilung geeignet ist
12	29	Eine alternative Form der Leistungsbeurteilung (z. B. zu Klausuren) ist
18	44 <sup>a</sup>	für die Kontrolle von Lernprozessen geeignet ist

PiF Position im Fragebogen

<sup>a</sup>Angepasstes Item

dem *viertens* ein zweiter Pretest mit zwei Messzeitpunkten in derselben Stichprobe ( $N_{P2} = 305$  Lehramtsstudierende, Alter in Jahren:  $M = 21,34$ ,  $SD = 2,79$ ; 64 % weiblich, 36 % männlich; 71 % im 2. Fachsemester, 20 % im 4. Fachsemester, 7 % im 6. Fachsemester) durchgeführt wurde, bei welchem auf ein vierstufiges Antwortformat (1 = *trifft überhaupt nicht zu*, 2 = *trifft eher nicht zu*, 3 = *trifft eher zu*, 4 = *trifft voll und ganz zu*) umgestellt wurde, um Antworttendenzen zur Mitte hin vorzubeugen, liegt im Ergebnis das Instrument (EPP-L) bestehend aus 18 Items (6 Skalen mit je 3 Items) vor (Tab. 2).

## 2.2 Validierung des EPP-L

### 2.2.1 Stichproben

Zur Validierung des Instrumentes wurden Stichproben aus zwei Studien herangezogen. Die erste Stichprobe setzt sich aus  $N_1 = 424$  Lehramtsstudierenden zusammen ( $M_{\text{Alter}} = 20,53$ ,  $SD_{\text{Alter}} = 2,16$ , weiblich [ $w$ ] = 58 %, männlich [ $m$ ] = 40 %): 330 Bachelorstudierende ( $M_{\text{Alter}} = 19,82$ ,  $SD_{\text{Alter}} = 1,70$ ,  $w = 65$  %,  $m = 32$  %) sowie 94 Masterstudierende ( $M_{\text{Alter}} = 23,13$ ,  $SD_{\text{Alter}} = 1,71$ ,  $w = 32$  %,  $m = 68$  %). Es handelt sich um Gelegenheitsstichproben, die an zwei deutschen Universitäten über verpflichtende Einführungsveranstaltungen akquiriert wurden. Weil für alle Skalen des EPP-L Messinvarianz zwischen beiden Lehramtssubgruppen vorlag (Messinvarianztestung nach Kleinke et al. 2017; Beurteilung von Root Mean Square Error Of Approximation [RMSEA], Comparative Fit Index [CFI], Standardized Root Mean Square Residual [SRMR]), wurden diese zusammengefasst.  $N_1$  als größere der beiden Stichproben diene als Grundlage für die meisten Validierungsanalysen. Weil Einstellungen jedoch erfahrungsabhängig sind (Glasman und Albarracín 2006), wurde zur Absicherung der Befunde zusätzlich auf eine zweite Stichprobe zurückgegriffen, in der die Befragten über mehr Erfahrung (d. h. mehr Lerngelegenheiten) mit Portfolioarbeit verfügten:  $N_2 = 374$  Lehramtsstudierende im Master oder auf Staatsexamen ( $M_{\text{Alter}} = 23,37$ ,  $SD_{\text{Alter}} = 4,49$ ,  $w = 78$  %,  $m = 21$  %). Diese zweite Gelegenheitsstichprobe wurde an zwölf deutschen Hochschulen über Dozierende im Lehramt akquiriert.

### 2.2.2 Instrumente zur Validierung

In den beiden Stichproben wurden unterschiedliche Validierungsinstrumente eingesetzt, was im Folgenden jeweils in Klammern ausgewiesen wird. Die Akzeptanz des Einsatzes von Portfolios wurde hinsichtlich der *konvergenten Validität* über vier Instrumente erfasst: eine Adaption der Skala von Schlag (2015; AZ1) mit 6 Items (z. B. „Ich würde das Portfolio anderen Studierenden empfehlen“) und guter interner Konsistenz ( $\alpha = 0,84$  [ $N_1$ ]), eine Adaption der Skala von Klampfer (2013; AZ2) mit drei Items (z. B. „Die Arbeit mit dem Portfolio gefällt mir gut“) und einer guten internen Konsistenz ( $\alpha = 0,80$  [ $N_1$ ]) sowie einer Skala von Ziegelbauer (2016; AZ3) mit sechs Items (z. B. „Ich nehme an, dass das Portfolio ein gutes Instrument zur Unterstützung meiner Professionalisierung als Lehrer/in ist“) und ebenfalls guter interner Konsistenz ( $\alpha = 0,76$  [ $N_1$ ]).

Zur Prüfung der *divergenten Validität* wurden transmissive (TRA) und konstruktivistische (KON) lerntheoretische Überzeugungen anhand zweier Skalen von Kunter et al. (2013, siehe auch Kunter et al. 2017) herangezogen. Die Skala zu konstruktivistischen lerntheoretischen Überzeugungen besteht aus sechs Items (z. B., „Man sollte Schüler(inne)n erlauben, sich eigene Wege zur Bearbeitung von Aufgaben auszudenken, bevor die Lehrkraft vorführt, wie diese zu beantworten sind“) und zeigt in den vorliegenden Daten eine gute interne Konsistenz ( $\alpha=0,84 [N_1]$ ). Die Skala zu transmissiven lerntheoretischen Überzeugungen besteht aus fünf Items (z. B., „Schüler(innen) lernen durch die Demonstration von Beispielaufgaben am besten“) und zeigt vorliegend ebenfalls gute interne Konsistenz ( $\alpha=0,74 [N_1]$ ).

Zur Prüfung der Kriteriumsvalidität wurde eine Skala zum Interesse an der Portfolioarbeit von Klampfer (2013) bestehend aus zwei Items (z. B., „Ich denke, dass die Arbeit mit E-Portfolios nützlich für die Weiterentwicklung in der Schulpraxis ist“) genutzt (INT). Zudem wurde der wahrgenommene Nutzen der Portfolioarbeit (NUZ) mit zwei Skalen erfasst: zum einen mit den sechs Items (z. B., „Ich nehme an, dass Portfolioarbeit für mein Studium hilfreich ist“) der Skala von Ziegelbauer (2016; NUZ1), die in den vorliegenden Daten eine sehr gute interne Konsistenz ( $\alpha=0,92 [N_1]$  bis  $0,93 [N_2]$ ) aufweist; zum anderen mit den fünf Items (z. B., „Was ich beim Arbeiten an und mit dem Portfolio lerne, ist später im Beruf sehr nützlich für mich“) der Skala von Wolfsgruber (2020; NUZ2), die vorliegend eine gute interne Konsistenz ( $\alpha=0,78 [N_2]$ ) zeigt. Darüber hinaus wurde die Motivation zur Portfolioarbeit (MOT) mit den vier Items der Skala von Ziegelbauer (2016) zur intrinsischen Motivation für Portfolioarbeit im Lehramtsstudium erfasst (z. B., „Ich arbeite mit dem Portfolio, weil es mir Spaß macht“), die in den vorliegenden Daten ebenfalls eine gute interne Konsistenz aufweist ( $\alpha=0,85 [N_1]$  bis  $0,88 [N_2]$ ).

Um diese Instrumente einsetzen zu können, wurden einzelne Items, in denen auf konzeptspezifische Aspekte abgezielt wurde, verallgemeinert adaptiert (z. B. wurde bei einem Item das Wort *Arbeitsheft* gestrichen). Alle Indikatoren wurden mittels vierstufiger Antwortskala erhoben (1 = *trifft überhaupt nicht zu*, 2 = *trifft eher nicht zu*, 3 = *trifft eher zu*, 4 = *trifft voll und ganz zu*).

### 2.2.3 Statistische Analysen

Für die Datenaufbereitung wurden die Software SPSS (v24.0.0.2) und R 4.0.2 (R Core Team 2021), für alle statistischen Analysen das Statistikprogramm Mplus 8 (Muthén und Muthén 1998–2017) verwendet.

Um eine Vergleichbarkeit zu anderen Studien zu ermöglichen, wurde Cronbach's Alpha ( $\alpha$ ) als Maß der internen Konsistenz berechnet. Zudem wurde McDonald's Omega ( $\omega$ ), welches die interne Konsistenz robuster schätzt (Hayes und Coutts 2020), für jede eindimensionale Skala als auch Omega-Subskala-spezifisch ( $\omega_{\text{Skala-S}}$ ) berechnet (Schermelleh-Engel und Gädé 2020). Zudem wurden die Itemschwierigkeiten  $p_i$  und Itemtrennschärfen  $r_{it}$  berechnet (Döring und Bortz 2016).

Die *Konstruktvalidität* des Instrumentes wurde zunächst mittels der Dimensionalität und Faktorstruktur der Items geprüft. Anhand explorativer Faktorenanalysen (EFA) mit einer obliquen Rotation (Geomin; Browne 2001) und einer Maximum-Likelihood (ML) Schätzung wurde geprüft, ob die 18 Items des EPP-L tatsächlich,

wie theoretisch angenommen, am besten durch sechs Inhaltsaspekte (d. h. sechs Faktoren) abgebildet werden.  $\chi^2$ -Differenztests ermöglichen den Vergleich der Passung der Modelle mit zwei bis sechs Faktoren (Satorra und Bentler 2001). Die latente Faktorenstruktur in konfirmatorischen Faktorenanalysen (CFA) wurde mit gegenüber nicht-normalverteilten Daten robustem Maximum-Likelihood-Schätzverfahren (MLR) geprüft. Dafür wurde das Mehrfaktorenmodell (mit sechs Faktoren), ein Modell zweiter Ordnung und ein Bi-Faktor Modell aufgestellt. Um zu prüfen, ob ein Generalfaktor (Einstellung zur Portfolioarbeit) angenommen werden kann, wurde das Mehrfaktorenmodell gegen ein Modell zweiter Ordnung und ein Bi-Faktor Modell mit  $\chi^2$ -Differenztests geprüft. Zur Beurteilung der Dominanz eines Generalfaktors wurde Omega-hierarchisch ( $\omega_H$ ) berechnet (Schermelleh-Engel und Gäde 2020).

Die *inhaltliche Bedeutung* der Faktoren des Instrumentes EPP-L wurde geprüft, indem die konvergente und divergente Validität untersucht wurde. Dafür wurde ein korrelativer Ansatz in Strukturgleichungsmodellen herangezogen (Geiser 2010; Moosbrugger und Kelava 2012): Die jeweiligen Instrumente wurden separat latent modelliert und mit der Einstellung zur Portfolioarbeit korreliert und so Messfehler berücksichtigt.

Die *praktische Relevanz* des Instrumentes wurde durch Analyse der Kriteriumsvalidität geprüft. Dabei wurden die Konstrukte auf Basis der Instrumente latent modelliert und ebenfalls mit der Einstellung zur Portfolioarbeit korreliert.

Um ein mit Blick auf die *Testfairness* unterschiedliches Antwortverhalten zwischen den Geschlechtern auf dieselben Items der Einstellungen zu Portfolioarbeit zu identifizieren, wurden für jedes Item Mehrere-Indikatoren-mehrere-Ursachen-Modelle (multiple-indicators multiple causes; MIMIC: Jöreskog und Goldberger 1975; DIF-Analyse) berechnet, um akkurate Schätzungen auch bei kleinen Stichproben zu ermöglichen (Jin und Chen 2020). Wie von Chun et al. (2016) empfohlen, wurde zunächst ein Baseline-Modell aufgestellt, bei welchem das gemessene latente Konstrukt (in diesem Fall ein Generalfaktor *Einstellung zu Portfolioarbeit*) auf die dichotome Geschlechtsvariable (0 = männlich, 1 = weiblich) regressiert wurde. Gegen dieses Baseline-Modell wurden dann 18 MIMIC-Modelle, die sich vom Baseline-Modell dadurch unterschieden, dass die jeweiligen Regressionen der 18 Items auf die Geschlechtsvariable einbezogen wurden, getestet ( $\chi^2$ -Differenztests). Es wurde also geprüft, ob ein Modell, bei welchem das Geschlecht die Ausprägung (d. h. das Antwortverhalten) auf einem der 18 Items des EPP-L vorhersagt, besser zu den Daten passt. Sollte diese Passung besser sein als im Baseline-Modell (bei welchem die Vorhersage der Ausprägung auf einem Item durch das Geschlecht nicht berücksichtigt wird), dann läge eine geschlechterspezifische Diskriminierung (DIF) durch dieses Item vor, weil in den Modellen zugleich die Ausprägung der generellen Einstellung zu Portfolioarbeit kontrolliert wird. Zudem wurde geprüft, ob die Regressionen der Items auf die Geschlechtsvariable statistisch signifikant sind. Für Robustheitsanalysen wurden die Skalenvarianzen als Schätzer extremen Antwortverhaltens als zusätzliche Prädiktoren in die MIMIC Modelle aufgenommen (Chen et al. 2017).

Zur Bewertung der Größe der Korrelationen (Effektstärken) orientieren wir uns an Cohen (1988), bezogen auf die interne Konsistenz an Taber (2018), bezüglich der

Itemschwierigkeiten und -trennschärfen an Döring und Bortz (2016) und hinsichtlich der Modellfitwerte an Schermelleh-Engel et al. (2003).

### 2.2.4 Umgang mit fehlenden Werten

Für die 4% fehlenden Werte in Stichprobe  $N_1$  und die 0,4% fehlenden Werte in Stichprobe  $N_2$  zeigt sich bezogen auf die 18 Items des Instrumentes, die Items der Validierungsinstrumente und die verfügbaren demografischen Variablen (Alter, Geschlecht, Fachsemester) mit dem Test nach Little (1988), dass die Annahme, wonach die fehlenden Werte völlig zufällig fehlen (MCAR Bedingung), nicht verworfen werden muss ( $p > 0,05$ ). Die modellbasierte Schätzung *Full Information Maximum Likelihood* (FIML) wurde als adäquate Methode zur Behandlung der fehlenden Werte eingesetzt (Graham 2012; Lüdtke et al. 2007; Schafer und Graham 2002). Dabei wurde das Alter der Teilnehmenden (mit Ausnahme der EFA, bei denen keine Hilfsvariablen aufgenommen werden konnten) als kontinuierliche Hilfsvariable verwendet, indem der Befehl *auxiliary* (m) in Mplus 8 verwendet wurde.

## 3 Ergebnisse

### 3.1 Mittelwerte und Reliabilitäten der Skalen sowie Schwierigkeiten der Items

Für die Berechnung der Mittelwerte und der Reliabilitäten der Skalen sowie der Itemschwierigkeiten wurde auf die größere Stichprobe ( $N_1 = 424$  Lehramtsstudierende) zurückgegriffen. Die in Tab. 3 abgebildeten und geschlechterdifferenzierten Mittelwerte und Standardabweichungen sind nahe dem theoretischen Skalenmittelwert und die Mittelwerte liegen unter Frauen durchweg höher. Nur bezüglich der Skalen *Entwicklung von Reflexivität* und *Möglichkeit der Innovierung der Ausbildung* sind die geschlechtsspezifischen Unterschiede nicht statistisch signifikant. Gemessen an Cohen's  $d$  sind die Mittelwertunterschiede für die Gesamtskala am bedeutsamsten.

Es zeigen sich gute interne Konsistenzwerte der sechs Skalen und der Gesamtskala, auch bei separaten Analysen für Frauen und Männer (Tab. 4).

Die Schwierigkeitsparameter liegen zwischen 53% und 81%, wobei das Item EPP-L17 ( $p_i = 0,81$ ) das einfachste (stärkste Zustimmung) und EPP-L37 ( $p_i = 0,53$ ) das schwierigste Item (geringste Zustimmung) ist. Bis auf das Item EPP-L9 ( $r_{ii} = 0,41$ ) weisen alle Items hohe Trennschärfen auf.

### 3.2 Die Validität des EPP-L

#### 3.2.1 Konstruktvalidität

Der Vergleich der Fit-Statistiken der Modelle von einem Faktor bis sechs Faktoren im Rahmen explorativer Faktorenanalysen zeigt: Das Modell mit sechs Faktoren passt anhand der Fit-Indizes (Schermelleh-Engel et al. 2003) gut zu den Daten (Tab. 11 im Anhang).

**Tab. 3** Manifeste Mittelwerte und Standardabweichungen und Mittelwertunterschiede zwischen Geschlechtern der Skalen sowie der Gesamtskala

	Item EPP-L	Stichproben						Mittelwertunterschiede ( $M_w$ und $M_m$ )	
		Gesamt		Pro Geschlechtsgruppe				$p$	$d$
		$M$	$SD$	$M_w$	$SD_w$	$M_m$	$SD_m$		
Entwicklung von Reflexivität	9	3,10	0,56	3,13	0,54	3,07	0,59	0,293	0,107
	5								
	17								
Entwicklung von Selbstständigkeit	11	2,94	0,68	3,02	0,66	2,82	0,70	0,005	0,296
	7								
Möglichkeit der Innovierung der Ausbildung	26	2,37	0,67	2,46	0,65	2,23	0,66	<0,001	0,352
	30								
	35								
Möglichkeit des Sammelns und der Dokumentation	42	2,38	0,69	2,48	0,69	2,23	0,67	<0,001	0,367
	37								
	32								
Möglichkeit der Begleitung und des Austauschs	23	2,43	0,76	2,48	0,74	2,38	0,80	0,191	0,131
	28								
	33								
Möglichkeit der Leistungsmessung	24	2,69	0,74	2,87	0,69	2,45	0,75	<0,001	0,587
	29								
	44								
Gesamtskala	–	2,65	0,50	2,74	0,48	2,53	0,50	<0,001	0,430

$N_1 = 424$ .  $M_w$  Mittelwert Frauen,  $SD_w$  Standardabweichung Frauen,  $M_m$  Mittelwert Männer,  $SD_m$  Standardabweichung Männer

$d = \text{Cohen's } d$

Erwartungskonform laden diejenigen Items, die jeweils theoretisch den sechs Skalen des EPP-L zugeordnet wurden, auch empirisch auf einem gemeinsamen Faktor, während die übrigen Items geringe Ladungen ( $\lambda < 0,300$ ) auf dem jeweiligen Faktor aufweisen (Tab. 12 im Anhang). Lediglich Item 9 weist eine geringe Ladung ( $\lambda = 0,272$ ) auf dem zugeordneten Faktor (F1: *Entwicklung von Reflexivität*) auf, die Ladungen des Items auf den anderen Faktoren reicht von  $\lambda = -0,020$  (F4: *Möglichkeit des Sammelns und der Dokumentation*) bis  $\lambda = 0,237$  (F3: *Möglichkeit der Innovierung der Ausbildung*). Allerdings lädt Item 9 auf Faktor F1 höher als auf allen anderen Faktoren. Zudem sind die Ladungen der drei Items 9, 5 und 17 von allen 18 Items die einzigen statistisch signifikanten. Die Beträge der Ladungen aller anderen Items sind niedriger als 0,165 (bei 13 Items sogar  $\lambda < 0,100$ ). Insgesamt kann auf Basis der Ladungsmuster von einer zuverlässigen Zuordnung der Items zu den sechs Faktoren ausgegangen werden, wodurch eine gute inhaltliche Interpretation der Faktoren möglich ist. Die Existenz der sechs Faktoren bedeutet, dass sich die Einstellungen zu unterschiedlichen Aspekten der Portfolioarbeit voneinander differenzieren lassen. Die inhaltliche Bedeutsamkeit dieser sechs Faktoren

**Tab. 4** Itemstatistiken und Reliabilitäten der Skalen

Item EPP-L	Itemstatistiken				Reliabilitäten der Skalen		
	<i>M</i>	<i>SD</i>	$R_{it}$	$p_i$	$\alpha$ ( $\alpha_w/\alpha_m$ )	$\omega$ ( $\omega_w/\omega_m$ )	$\omega_{Skala-S}$ ( $\omega_{Skala-Sw}/\omega_{Skala-Sm}$ )
9	2,98	0,64	0,41	0,74	0,71	0,72	0,43
5	3,11	0,75	0,61	0,78	(0,68/0,73)	(0,74/0,76)	(0,34/0,51)
17	3,22	0,72	0,58	0,81			
11	2,97	0,82	0,58	0,74	0,79	0,80	0,22
7	2,84	0,83	0,64	0,71	(0,78/0,79)	(0,77/0,74)	(0,20/0,25)
19	3,01	0,80	0,66	0,75			
26	2,19	0,76	0,62	0,55	0,77	0,77	0,22
30	2,38	0,79	0,63	0,60	(0,76/0,76)	(0,78/0,81)	(0,25/0,37)
35	2,52	0,84	0,57	0,63			
42	2,56	0,79	0,63	0,64	0,79	0,80	0,33
37	2,13	0,84	0,65	0,53	(0,80/0,78)	(0,81/0,79)	(0,29/0,40)
32	2,45	0,81	0,64	0,61			
23	2,51	0,85	0,74	0,63	0,89	0,89	0,63
28	2,42	0,81	0,81	0,60	(0,88/0,90)	(0,88/0,91)	(0,63/0,62)
33	2,37	0,86	0,79	0,59			
24	2,65	0,88	0,59	0,66	0,73	0,74	0,27
29	2,70	1,04	0,56	0,68	(0,73/0,73)	(0,67/0,75)	(0,29/0,25)
44	2,73	0,81	0,54	0,68			

( $\alpha_w/\alpha_m$ ) der Gesamtskala = 0,90 (0,90/0,90),  $\omega_T$  ( $\omega_{Tw}/\omega_{Tm}$ ) Omega-total der Gesamtskala = 0,87 (0,87/0,87),  $\omega_H$  ( $\omega_{Hw}/\omega_{Hm}$ ) Omega-hierarchisch der Gesamtskala = 0,65 (0,65/0,57)

\*\*\*  $p < 0,001$

ergibt sich beim Blick auf die jeweils zugehörigen Items (Tab. 2). So bezieht sich beispielsweise die Einstellung einer Person gegenüber *Entwicklung von Reflexivität* auf die Eignung von Portfolioarbeit, Reflexionsfähigkeit zu fördern. Davon zu unterscheiden ist z. B. die Einstellung gegenüber *Entwicklung von Selbstständigkeit*, welche die Relevanz von Portfolioarbeit für eigenverantwortliches Handeln betont.

Bis auf eine Ausnahme korrelieren alle sechs Skalen zu diesen Faktoren gering bis moderat positiv (Cohen 1988) und statistisch signifikant. Eine Ausnahme bildet die Beziehung zwischen der Skala *Möglichkeit der Innovierung der Ausbildung* und der Skala *Möglichkeit der Leistungsmessung*, die nicht statistisch signifikant korrelieren (Tab. 5).

Die Prüfung der latenten Struktur der sechs Skalen auf Grundlage der Stichproben  $N_1$  und  $N_2$  erfolgte, indem das Mehrfaktorenmodell (M7) sowohl gegen das Modell zweiter Ordnung (M8) als auch gegen das Bi-Faktor Modell (M9) getestet wurde (Tab. 6 und 13 im Anhang). Die  $\chi^2$ -Differenztests ( $N_1$ : M7–M9:  $\Delta\chi^2 = 29.548$ ,  $\Delta df = 3$ ,  $p < 0,001$ ;  $N_1$ : M8–M9:  $\Delta\chi^2 = 22.586$ ,  $\Delta df = 9$ ,  $p = 0,007$ ;  $N_1$ : M7–M9:  $\Delta\chi^2 = 18.784$ ,  $\Delta df = 3$ ,  $p < 0,001$ ;  $N_1$ : M8–M9:  $\Delta\chi^2 = 35.538$ ,  $\Delta df = 9$ ,  $p < 0,001$ ) und die Modell-Fit-Indizes legen die Existenz eines Generalfaktors *Einstellung zu Portfolioarbeit* nahe.

Vergleiche zwischen dem Modell zweiter Ordnung und dem Bi-Faktor Modell können kaum sinnvoll realisiert werden (Mansolf und Reise 2017). Vielmehr pas-

**Tab. 5** Korrelation der sechs Faktoren der explorativen Faktorenanalyse

Faktor	Korrelationen					
	F1	F2	F3	F4	F5	F6
Entwicklung von Reflexivität (F1)	1	–	–	–	–	–
Entwicklung von Selbstständigkeit (F2)	0,424*	1	–	–	–	–
Möglichkeit der Innovierung der Ausbildung (F3)	0,587*	0,376*	1	–	–	–
Möglichkeit des Sammelns und der Dokumentation (F4)	0,504*	0,324*	0,473*	1	–	–
Möglichkeit der Begleitung und des Austauschs (F5)	0,534*	0,352*	0,449*	0,569*	1	–
Möglichkeit der Leistungsmessung (F6)	0,272*	0,246*	0,148	0,325*	0,432*	1

$N_1 = 423$ . Rotation = Geomin, Schätzer = Maximum Likelihood.

\* $p < 0,05$

**Tab. 6** Schätzungen der konfirmatorischen Faktorenanalysen

Getestetes Modell	$\chi^2$	<i>Df</i>	RMSEA [90%-KI]	CFI	SRMR	AIC	aBIC
M7: Mehrfaktorenmodell	257,353	120	0,052 [0,43, 0,61]	0,953	0,046	17.826,016	17.904,014
M8: Modell zweiter Ordnung	279,947	129	0,053 [0,44, 0,61]	0,948	0,050	17.832,618	17.902,728
M9: Bi-Faktor Modell	227,550	117	0,047 [0,38, 0,56]	0,962	0,038	17.799,560	17.880,187

$N_1 = 424$ . Robustes Maximum-Likelihood-Schätzverfahren (Schätzer = MLR). Alle Itemparameter wurden frei geschätzt und die Varianzen der latenten Faktoren wurden auf 1,0 fixiert.  $\chi^2$  Statistik der Anpassungsgüte, *AIC* Akaike Information Criterion, *aBIC* Adjusted Bayesian Information Criteria (angepasst an die Stichprobengröße), *CFI* Comparative Fit Index, *df* Freiheitsgrade, *KI* Konfidenzintervall, *RMSEA* Root Mean Square Error of Approximation, *SRMR* Standardized Root Mean Square Residual

sen aufgrund der Modell-Fit-Indizes beide Modelle gut zu den Daten. Weil jedoch nur das Bi-Faktormodell durchweg gute Modell-Fit-Statistiken aufweist und zudem sparsamer ist, wurde für die folgenden Validierungsanalysen das Bi-Faktor Modell herangezogen. Alle Statistiken (z. B. standardisierte Ladungen) der drei Modelle basierend auf  $N_1$  sind in Abb. 1 dargestellt.

Mit Blick auf AIC und aBIC passen die Modelle mit Generalfaktor im Vergleich zum Mehrfaktormodell nur unwesentlich anders zu den Daten: das Modell zweiter Ordnung passt etwas schlechter, das Bi-Faktor Modell passt etwas besser. Das Omega-hierarchisch von  $\omega_H = 0,65$  bezüglich des Bi-Faktor Modells verdeutlicht allerdings, dass durch den Generalfaktor ein substanzieller Anteil der wahren Varianz an der Gesamtvarianz der Items erklärt wird.





**Tab. 8** Korrelationen zwischen den latenten Konstrukten zur Überprüfung der Validität

	Kriteriumsvalidität					
	NUZ1		NUZ2		MOT	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Einstellung zu Portfolioarbeit	0,827	<0,001	0,849	<0,001	0,668	<0,001

$N_2=374$ . *MOT* Die Motivation zur Portfolioarbeit, *NUZ1* Der wahrgenommene Nutzen der Portfolioarbeit von Ziegelbauer (2016), *NUZ2* Der wahrgenommene Nutzen der Portfolioarbeit von Wolfsgruber (2020)

### 3.2.2 Konvergente und divergente Validität sowie Kriteriumsvalidität

Zur Prüfung der konvergenten und divergenten Validität sowie der Kriteriumsvalidität wird der Generalfaktor *Einstellung zu Portfolioarbeit* aus dem Bi-Faktor Modell mit den entsprechenden latenten Konstrukten korreliert ( $N_1$ : Tab. 7;  $N_2$ : Tab. 8).

Erwartungskonform zeigen sich starke positive Zusammenhänge mit den Konstrukten, die ebenfalls Einstellungen zu (d.h. meist Akzeptanz von) Portfolioarbeit messen ( $r > 0,81$ ,  $p < 0,001$ ). Zudem sind die Zusammenhänge mit dem Interesse an der Portfolioarbeit (INT), der Motivation zur Portfolioarbeit (MOT) und dem wahrgenommenen Nutzen der Portfolioarbeit (NUZ) wie erwartet statistisch signifikant und mittel bis stark positiv. Des Weiteren zeigt sich erwartungsgemäß ein schwacher, statistisch signifikanter Zusammenhang mit dem Konstrukt transmissive Überzeugungen (TRA) sowie kein Zusammenhang mit konstruktivistischen lerntheoretischen Überzeugungen (KON).

### 3.3 Testfairness

Im Kontext dieser Studie läge *Uniform Differential Item Functioning (DIF)* vor, wenn Individuen mit gleichen Einstellungen zu Portfolioarbeit unterschiedliche Wahrscheinlichkeiten besitzen, den Antwortkategorien eines Items des Generalfaktors *Einstellung zu Portfolioarbeit* deshalb zuzustimmen, weil sie einem bestimmten Geschlecht angehören (0 = männlich, 1 = weiblich). Ein einheitlich gruppenabhängiges Item (Englisch *uniform DIF*, im Folgenden kurz DIF) läge vor, wenn das Item konsistent über alle Ausprägungen der Einstellungen zu Portfolioarbeit hinweg von unterschiedlichen Geschlechtsgruppen unterschiedlich beantwortet wird. Das Baseline Modell, gegen welches alle MIMIC-Modelle pro Item getestet wurden, zeigt einen guten Fit (AIC = 17.917.436, BIC = 18.303.454, aBIC = 17.998.826, RMSEA = 0,047, 90 %-KI [0,039; 0,056], CFI = 9,58, TLI = 0,947, SRMR = 0,039). Der Vergleich der erweiterten MIMIC-Modelle unter Verwendung der Stichprobe  $N_1$ , die einen direkten Effekt des Geschlechts auf einzelne Items enthalten, zeigt, dass zwei der 18 Items (EPP-L24 und 29) der Skala *Möglichkeit der Leistungsmessung* einheitlich gruppenabhängige Items sind (siehe grau hinterlegte Zeilen in Tab. 14 im Anhang). Alle anderen Items diskriminierten keine Geschlechtsgruppe, werden also nicht konsistent von einem Geschlecht anders beantwortet als von anderen. Um die Robustheit der Ergebnisse zu testen, wurden die gleichen MIMIC-Modelle pro Item erneut geprüft und dabei der direkte Effekt der individuellen Ska-

lenvarianz auf die jeweiligen Items berücksichtigt, um extremes Antwortverhalten zu kontrollieren. Die DIF Effekte bestätigen die Ergebnisse.

## 4 Diskussion

Portfolioarbeit ist ein zentraler, teilweise verpflichtender Baustein der Professionalisierung von angehenden Lehrpersonen. Bislang lag kein validiertes Instrument vor, das losgelöst von spezifischen Portfoliokonzeptionen und Anwendungskontexten Einstellungen von Lehramtsstudierenden zur Portfolioarbeit erfassen konnte. In diesem Beitrag wurde die Entwicklung und Validierung eines Instrumentes (EPP-L) vorgestellt, welches die Einstellungen eindimensional und sechs ihnen zugrundeliegende Inhaltsaspekten von Portfolioarbeit mit insgesamt 18 Items erfasst. Mit dem EPP-L liegt erstmals ein theoretisch fundiertes, reliables und valides Instrument vor, welches konzeptionsübergreifend Einstellungen zu Portfolioarbeit standardisiert erfasst. Es kann einen Beitrag zur Qualitätssicherung bei der Gestaltung von Professionalisierungsprozessen in der Lehrerinnen- und Lehrerbildung leisten.

Ein solches Instrument ist insbesondere deshalb erforderlich, weil sich in der Literatur und Praxis der Lehrerinnen- und Lehrerbildung zwar vielfältige Potenzialzuschreibungen an die Portfolioarbeit finden (Feder und Cramer 2018), es bislang aber kaum Evidenz für Effekte dieser postulierten Potenziale von Portfolioarbeit gibt (Feder und Cramer 2019). Das hier entwickelte Instrument zu Einstellungen zur Portfolioarbeit ist angesichts der großen Verbreitung von Portfolioarbeit in der Lehrerinnen- und Lehrerbildung und den damit verbundenen Ressourcen (bei zugleich wenigen belastbaren Befunden im deutschsprachigen Raum) ein wichtiger Ausgangspunkt, um ein genaueres und belastbares Bild von deren tatsächlichen Erträge zu erhalten. Perspektivisch legt sich eine Adaption auch mit Blick auf die zweite Phase (Referendariat) und dritte Phase (Fort- und Weiterbildung) der Lehrerinnen- und Lehrerbildung nahe.

Zugleich bietet das Instrument EPP-L die Möglichkeit, Chancen und Grenzen von Portfolioarbeit in der Ausbildung von Lehrpersonen insgesamt kritisch-konstruktiv zu prüfen. Sollte die sich konform zum Forschungsstand zeigende Skepsis angehender Lehrpersonen gegenüber Portfolioarbeit erhärten, sich also eine potenzielle „Prozessillusion“ (Hascher und Sonntagbauer 2013) der Portfolioarbeit mit Blick auf die Professionalisierung bestätigen, so muss auch infrage gestellt werden, in welchem Umfang Portfolioarbeit diejenige Aufmerksamkeit erfährt, die ihr derzeit etwa bildungspolitisch oder im Rahmen der *Qualitätsoffensive Lehrerbildung* zuteilwird. In diesem Zusammenhang muss einschränkend darauf hingewiesen werden, dass Schlussfolgerungen zu konkretem Verhalten von angehenden Lehrpersonen bezüglich der Portfolioarbeit basierend auf Einstellungen nur eingeschränkt möglich sind. Zwar haben sich Zusammenhänge zwischen Einstellungen und Verhalten in unterschiedlichen Bereichen als substantiell erwiesen (Glasman und Albarracín 2006), gleichzeitig zeigten sich jedoch Moderationen durch unterschiedliche Variablen, wie beispielsweise das Ausmaß direkter Erfahrungen mit dem Einstellungsobjekt (hier: Portfolioarbeit). Zudem ist limitierend die Erfassung von Einstellungen nicht mit tatsächlich realisierten Potenzialen beziehungsweise Wirkungen von Portfolioarbeit

gleichzusetzen. Um beispielsweise die Effekte von Portfolioarbeit auf die Professionalisierung von (angehenden) Lehrpersonen zu untersuchen, müssten zukünftig Interventionsstudien (z. B. randomisierte kontrollierte Studien) in der Lehrerinnen- und Lehrerbildung durchgeführt werden.

Über die Frage der Einstellungen zu Portfolioarbeit und damit perspektivisch zum Nutzungsverhalten hinaus ist weiterhin zu klären, ob Portfolioarbeit die institutionell als bedeutsam erachteten Anforderungen (z. B. an eine Steigerung von Kohärenz zwischen Phasen, Komponenten und Akteuren der Lehrerinnen- und Lehrerbildung) erfüllt. Kohärenz im Sinne einer sinnhaften Verknüpfung der Elemente in der Lehrerinnen- und Lehrerbildung könnte sich letztlich auch eher über eine informell-individuelle Relationierungsleistung (Cramer 2020) und damit weniger durch Portfolioarbeit als vielmehr in der Interaktion der Akteure (Lehramtsstudierende und Dozierende) einstellen. Dabei kann ein Portfolio einen unterstützenden Beitrag leisten, z. B. indem es derjenige Ort ist, an dem verschiedene Komponenten des Lehramtsstudiums (Fachwissenschaften, Fachdidaktiken, Bildungswissenschaften, Schulpraxis) bewusst aufeinander bezogen werden. Dadurch werden verschiedene Perspektiven auf denselben Gegenstand eingenommen. Erst dann schöpft Portfolioarbeit ihr besonderes Potenzial aus, zeit- und ortsunabhängig in einem meta-reflexiven Sinne zu einer Relationierung der disziplinär, paradigmatisch und theoretisch sehr heterogenen Elemente der Lehrerinnen- und Lehrerbildung beizutragen (Cramer et al. 2019). Dieser Beitrag zur Professionalisierung kann von der Portfolioarbeit aber nur dann ausgehen, wenn die Einstellungen Studierender (und Dozierender) ihr gegenüber entsprechend positiv sind und die Akteure in der Lehrerinnen- und Lehrerbildung sich mehrheitlich bestimmten Zielen der Portfolioarbeit verpflichten. Genau hier leistet das Instrument über seine wissenschaftliche Relevanz hinaus auch einen Beitrag zur Evaluation und Weiterentwicklung der Lehrerinnen- und Lehrerbildung.

Dabei verdeutlichen die Modellfitindizes, dass in der Forschungspraxis sowohl die Modellierung über ein Bi-Faktormodell als auch über ein Mehrfaktormodell vertretbar ist, wenngleich der Generalfaktor einen substanziellen Anteil der wahren Varianz der Items aufklärt. Das wiederum bedeutet, dass die einzelnen Skalen des Instrumentes EPP-L in zukünftiger Forschung parallel eingesetzt und ausgewertet werden können und eine aufwendigere Modellierung eines Bi-Faktormodells nicht zwingend notwendig erscheint. Es liegt damit ein gut handhabbares Instrument vor, das effizient eingesetzt werden kann.

#### 4.1 Limitationen

Zur adäquaten Einschätzung der Güte des EPP-L sollten zwei Limitationen bedacht werden: Erstens ist das Instrument zwar weitestgehend messinvariant bezüglich des Geschlechts, allerdings sind bezüglich zwei der Items unter Frauen höhere Antwortkategorien wahrscheinlicher als unter Männern, obwohl dieselbe Einstellung zu Potenzialen von Portfolioarbeit vorliegt. Dies könnte beispielsweise daran liegen, dass Frauen und Männer Leistungsbeurteilungssituationen unterschiedlich gegenüberstehen und die Worte in den Items daher unterschiedliche Assoziationen auslösen und/oder die Items unterschiedlich verstanden werden. Zwar lässt sich annehmen, dass der Itempool des EPP-L insgesamt zur Erfassung des Generalfaktors

*Einstellung zu Portfolioarbeit* geeignet ist, doch sollte die unterschiedliche Funktionsweise des EPP-L bezüglich des Geschlechts insbesondere bei der Interpretation der Skala *Möglichkeit der Leistungsmessung* berücksichtigt werden. Hier ist nicht auszuschließen, dass bezogen auf Frauen und Männer unterschiedliche Konstrukte erfasst werden. Sollte sich die Verletzung der Testfairness bei den beiden Items in künftiger Forschung als robust erweisen, so ist es ratsam, auf den Einsatz dieser Skala zu verzichten. Künftig kann zudem nach der Relevanz alternativer erklärender Variablen (personenbezogene Variablen wie z. B. Selbstwirksamkeit) gefragt werden. Neben der dichotomen Unterscheidung Frau/Mann wurde keine weitere Gendergruppe (divers) einbezogen, da empirisch abgesicherte Aussagen angesichts der geringen Inzidenz dieser Gruppe in der Stichprobe kaum möglich sind. Schlussfolgerungen zum Verhalten des Instrumentes für diese Gendergruppe sind aus den Analysen in dieser Studie nicht möglich.

Zweitens erscheint es lohnend, die Items des EPP-L weiter zu optimieren. So wäre zu überlegen, ob einzelne Items so ausgetauscht werden können, dass eine höhere Streuung der Zustimmungswerte (d. h. der Schwierigkeitsindizes) erzielt wird (vgl. Döring und Bortz 2016). Zudem weist das Item EPP-L9 („Ich traue Portfolioarbeit zu, dass sie das Nachdenken über Erfahrungen in der Praxis aus unterschiedlichen Blickwinkeln vertieft“) eine niedrige Ladung bezüglich des Faktors *Entwicklung von Reflexivität* auf. Dieses Item könnte zukünftig umformuliert oder gegebenenfalls ersetzt werden.

## 4.2 Ausblick

Das Instrument EPP-L beansprucht aufgrund seiner theoretischen Fundierung auf den zuvor identifizierten Potenzialen von Portfolioarbeit, unabhängig von Standorten und spezifischen Konzeptionen eingesetzt werden zu können. Damit verbindet sich großes Potenzial zur Bearbeitung des Desiderats, über einzelne Konzeptevaluationen hinaus Erkenntnisse zur Portfolioarbeit und Einsatz in der Lehrerinnen- und Lehrerbildung zu gewinnen. Es ergeben sich Anhaltspunkte für eine Optimierung der Portfolioarbeit, indem deutlich wird, welche Potenziale seitens der Nutzenden als besonders tragfähig angesehen werden und welche erwarteten Potenziale keine Entsprechung finden. Alternativ könnten sehr kritische Einstellungen bestimmte Portfoliokonzeptionen generell in Frage stellen und so Reformbedarf markieren. Wie sich die Einstellungen insgesamt quantifizieren, muss durch standortübergreifende (internationale) Forschung in größeren Stichproben geklärt werden, um das Potenzial von Portfolioarbeit insgesamt zu evaluieren. Erste Überlegungen dazu werden im Rahmen des Forschungsnetzwerks *Portfolioarbeit in der Bildung* (<https://uniteubingen.de/de/168549>) angestellt. Solche Forschung kann dazu beitragen, Synergien zwischen den bislang stark unterschiedlichen Portfoliokonzeptionen in der Ausbildung von Lehrpersonen konzeptionell zu stärken und diese so evidenzbasierten Forschungs- und Entwicklungsprozessen zugänglich zu machen.

## 5 Anhang

**Tab. 9** Itemschwierigkeiten ( $p_i$ ), Trennschärfen ( $r_{it}$ ), Cronbach's Alpha ( $\alpha$ ) und McDonald's Omega ( $\omega$ ) beim ersten Pretest

Item EPP-L	$P_i$	$R_{it}$	Faktor	$\alpha$	$\omega$
1	0,79	0,61	Entwicklung von Reflexivität	0,79	0,83
5	0,60	0,66			
9	0,64	0,63			
13	0,66	0,46			
17	0,82	0,76			
2	0,70	0,49	Entwicklung berufsbezogener Fähigkeiten	0,70	0,73
6	0,59	0,52			
10	0,67	0,53			
14	0,82	0,50			
18	0,65	0,46	Entwicklung von Selbstständigkeit	0,78	0,81
3	0,64	0,54			
7	0,73	0,59			
11	0,73	0,69			
15	0,62	0,49			
19	0,67	0,62	Entwicklung der Fähigkeit, Portfolios im Unterricht einzusetzen	0,57	0,59
4	0,62	0,42			
8	0,79	0,36			
12	0,53	0,28			
16	0,65	0,35			
20	0,74	0,38	Möglichkeit der Innovierung der Ausbildung	0,43	0,51
21	0,65	0,04			
26	0,76	0,30			
31	0,59	0,43			
36	0,69	0,18			
41	0,72	0,29	Möglichkeit des Sammelns und der Dokumentation	0,30	0,31
22	0,50	0,11			
27	0,68	0,16			
32	0,55	0,31			
37	0,47	0,13			
42	0,66	0,14	Möglichkeit der Begleitung und des Austauschs	0,52	0,53
23	0,61	0,36			
28	0,64	0,38			
33	0,60	0,40			
38	0,42	0,18			
43	0,73	0,35	Möglichkeit der Leistungsmessung	0,60	0,61
24	0,64	0,24			
29	0,45	0,45			
34	0,80	0,48			
39	0,51	0,39			
44	0,60	0,36			

**Tab. 9** (Fortsetzung)

Item EPP-L	$P_i$	$R_{it}$	Faktor	$\alpha$	$\omega$
25	0,54	0,47	Möglichkeit der „Brückenfunktion“	0,65	0,70
30	0,56	0,60			
35	0,67	0,48			
40	0,63	0,46			
45	0,61	0,25			

$N=183$ . Nummerierung der Items entsprechend der Position der Items im Fragebogen

**Tab. 10** Faktorladungen der explorativen Faktorenanalyse beim ersten Pretest

Item EPP-L	1 Faktor Modell		2 Faktoren Modell			
	$\beta$	$p$	1. Faktor		2. Faktor	
	$\beta$	$p$	$\beta$	$p$	$\beta$	$p$
1 <sup>c</sup>	0,689	<0,05	Modell			
5 <sup>b</sup>	0,758	<0,05	konver-			
9 <sup>b</sup>	0,686	<0,05	gierte			
13 <sup>nb</sup>	0,496	<0,05	nicht			
17 <sup>b</sup>	0,862	<0,05				
2 <sup>a,b</sup>	0,580	<0,05	Modell			
6 <sup>c</sup>	0,683	<0,05	konver-			
1a <sup>a,b</sup>	0,657	<0,05	gierte			
14 <sup>c</sup>	0,568	<0,05	nicht			
18 <sup>c</sup>	0,532	<0,05				
3 <sup>c</sup>	0,620	<0,05	0,748	<0,05	-0,131	n. s.
7 <sup>b</sup>	0,678	<0,05	0,559	n. s.	0,130	n. s.
11 <sup>b</sup>	0,812	<0,05	0,827	<0,05	0,006	n. s.
15 <sup>c</sup>	0,551	<0,05	0,308	n. s.	0,274	n. s.
19 <sup>b</sup>	0,708	<0,05	0,002	n. s.	0,901	<0,05
4 <sup>c</sup>	0,612	<0,05	0,672	<0,05	0,003	n. s.
8 <sup>c</sup>	0,539	<0,05	0,583	<0,05	-0,009	n. s.
12 <sup>a,b</sup>	0,374	<0,05	0,206	n. s.	0,240	n. s.
16 <sup>c</sup>	0,483	<0,05	0,335	n. s.	0,205	n. s.
20 <sup>c</sup>	0,447	<0,05	-0,001	n. s.	0,845	n. s.
21 <sup>c</sup>	0,131	n. s.	Modell			
26 <sup>b</sup>	0,522	<0,05	konver-			
31 <sup>b</sup>	0,643	<0,05	gierte			
36 <sup>c</sup>	0,323	<0,05	nicht			
41 <sup>b</sup>	0,412	<0,05				
22 <sup>c</sup>	0,152	n. s.	Modell			
27 <sup>c</sup>	0,184	n. s.	konver-			
32 <sup>b</sup>	0,905	<0,05	gierte			
37 <sup>b</sup>	0,233	n. s.	nicht			
42 <sup>b</sup>	0,165	n. s.				

**Tab. 10** (Fortsetzung)

Item EPP-L	1 Faktor Modell		2 Faktoren Modell			
	$\beta$	$p$	1. Faktor		2. Faktor	
	$\beta$	$p$	$\beta$	$p$	$\beta$	$p$
23 <sup>b</sup>	0,480	<0,05	0,417	<0,05	0,133	n. s.
28 <sup>b</sup>	0,679	<0,05	0,719	<0,05	-0,004	n. s.
33 <sup>b</sup>	0,658	<0,05	0,646	<0,05	0,002	n. s.
38 <sup>c</sup>	0,092	n. s.	-0,120	n. s.	0,481	n. s.
43 <sup>c</sup>	0,285	<0,05	0,004	n. s.	0,837	n. s.
24 <sup>b</sup>	0,167	n. s.	0,632	<0,05	-0,010	n. s.
29 <sup>b</sup>	0,754	<0,05	-0,005	n. s.	0,766	<0,05
34 <sup>c</sup>	0,768	<0,05	0,003	n. s.	0,776	<0,05
39 <sup>c</sup>	0,376	<0,05	0,366	<0,05	0,276	<0,05
44 <sup>b</sup>	0,318	<0,05	0,433	<0,05	0,200	n. s.
25 <sup>b</sup>	0,625	<0,05	0,695	<0,05	-0,051	n. s.
30 <sup>b</sup>	0,815	<0,05	0,745	<0,05	0,042	n. s.
35 <sup>b</sup>	0,559	<0,05	0,582	<0,05	-0,003	n. s.
40 <sup>c</sup>	0,524	<0,05	0,000	n. s.	1,875	n. s.
45 <sup>c</sup>	0,263	<0,05	0,226	n. s.	0,068	n. s.

$N=183$ . Nummerierung der Items entsprechend der Position der Items im Fragebogen. Rotation= Geomin, Schätzer= Maximum Likelihood

<sup>a</sup>Items wurden zu einem Faktor zusammengeführt

<sup>b</sup>Items, die beibehalten wurden

<sup>c</sup>Items, die nicht beibehalten wurden

**Tab. 11** Fit-Statistiken der Lösungen der explorativen Faktorenanalyse

Modell	M1	M2	M3	M4	M5	M6
Anzahl der Faktoren	1	2	3	4	5	6
Chi-Quadrat Statistik	$\chi^2(135, 423) = 1221,15, p < 0,001$	$\chi^2(118, 423) = 688,35, p < 0,001$	$\chi^2(102, 423) = 452,98, p < 0,001$	$\chi^2(87, 423) = 321,57, p < 0,001$	$\chi^2(73, 423) = 221,73, p < 0,001$	$\chi^2(60, 423) = 123,65, p < 0,001$
Eigenwerte	6,716	1,736	1,509	1,060	0,954	0,901
<i>Deskriptive Maße der Gesamtmodellanpassung</i>						
RMSEA	0,138	0,107	0,090	0,080	0,069	0,050
[90 %-KI]	[0,131, 0,145]	[0,099, 0,115]	[0,082, 0,099]	[0,071, 0,089]	[0,059, 0,080]	[0,037, 0,063]
SRMR	0,089	0,064	0,045	0,038	0,031	0,017
<i>Deskriptive Maße auf Basis von Modellvergleichen</i>						
CFI	0,663	0,823	0,891	0,927	0,954	0,980
TLI	0,618	0,771	0,837	0,872	0,903	0,950

**Tab. 11** (Fortsetzung)

Modell	M1	M2	M3	M4	M5	M6
<i>Deskriptive Maße der Modellsparsamkeit</i>						
AIC	16.103,458	15.604,663	15.401,296	15.299,883	15.228,040	15.155,963
BIC	16.322,016	15.892,027	15.753,417	15.712,715	15.697,535	15.678,074
aBIC	16.150,656	15.666,720	15.477,337	15.389,034	15.329,428	15.268,714

$N_1 = 423$ . Rotation = Geomin, Schätzer = Maximum Likelihood, 18 Items. *AIC* Akaike Information Criterion, *BIC* Bayesian Information Criterion, *aBIC* Adjusted Bayesian Information Criteria (angepasst an die Stichprobengröße), *CFI* Comparative Fit Index, *KI* Konfidenzintervall, *RMSEA* Root Mean Square Error of Approximation, *SRMR* Standardized Root Mean Square Residual, *TLI* Tucker-Lewis Index

**Tab. 12** Faktorladungen der Lösung mit sechs Faktoren der explorativen Faktorenanalyse

Faktor	Item EPP-L	Ladungen ( $\lambda$ ) der Items auf den Faktor					
		F1	F2	F3	F4	F5	F6
Entwicklung von Reflexivität (F1)	9	0,272	–	–	–	–	–
	5	0,734	–	–	–	–	–
	17	0,729	–	–	–	–	–
Entwicklung von Selbstständigkeit (F2)	11	–	0,517	–	–	–	–
	7	–	0,737	–	–	–	–
	19	–	0,733	–	–	–	–
Möglichkeit der Innovierung der Ausbildung (F3)	26	–	–	0,797	–	–	–
	30	–	–	0,643	–	–	–
	35	–	–	0,477	–	–	–
Möglichkeit des Sammelns und der Dokumentation (F4)	42	–	–	–	0,618	–	–
	37	–	–	–	0,807	–	–
	32	–	–	–	0,613	–	–
Möglichkeit der Begleitung und des Austauschs (F5)	23	–	–	–	–	0,791	–
	28	–	–	–	–	0,871	–
	33	–	–	–	–	0,846	–
Möglichkeit der Leistungsmessung (F6)	24	–	–	–	–	–	0,644
	29	–	–	–	–	–	0,691
	44	–	–	–	–	–	0,518

$N_1 = 423$ . Rotation = Geomin, Schätzer = Maximum Likelihood. Es werden nur Faktorladungen  $>0,3$  dargestellt, außer die Faktorladung eines Items zum zugehörigen Faktor ist  $<0,3$

**Tab. 13** Schätzungen der konfirmatorischen Faktorenanalysen für die drei Modelle

Getestetes Modell	$\chi^2$	<i>Df</i>	RMSEA [90 %-KI]	CFI	SRMR	AIC	aBIC
M7: Mehrfaktorenmodell	206,552	120	0,044 [0,34, 0,54]	0,978	0,032	15.292,961	15.359,848
M8: Modell zweiter Ordnung	243,006	129	0,049 [0,39, 0,58]	0,972	0,039	15.315,062	15.375,185
M9: Bi-Faktor Modell	188,021	117	0,040 [0,29, 0,51]	0,982	0,032	15.279,109	15.348,250

$N_2 = 374$ . Robustes Maximum-Likelihood-Schätzverfahren (Schätzer = MLR). Alle Itemparameter wurden frei geschätzt und die Varianzen der latenten Faktoren wurden auf 1,0 fixiert.  $\chi^2$  Statistik der Anpassungsgüte, *AIC* Akaike Information Criterion, *aBIC* Adjusted Bayesian Information Criteria (angepasst an die Stichprobengröße), *CFI* Comparative Fit Index, *df* Freiheitsgrade, *KI* Konfidenzintervall, *RMSEA* Root Mean Square Error of Approximation, *SRMR* Standardized Root Mean Square Residual

**Tab. 14** Informationskriterien des einheitlichen MIMIC-DIF-Testverfahrens für jedes der 18 Items des EPP-L, Modellvergleiche und DIF-Effekte

	LL	Npar	SCF	AIC	BIC	aBIC	Modellvergleich			Uniform DIF Effekte		
							TRd	p	β	SE	p	
Baseline	-8862,7	72	1,471	17.917,4	18.303,5	17.998,8	-	-	-	-	-	-
9	-8862,4	74	1,466	17.918,7	18.308,7	18.001,0	0,698	0,404	-0,038	0,046	0,409	
5	-8862,7	74	1,466	17.919,4	18.309,5	18.001,7	0,031	0,861	0,006	0,040	0,890	
17	-8861,1	74	1,466	17.916,1	18.306,2	17.998,4	3,318	0,069	-0,073	0,041	0,071	
11	-8862,7	74	1,467	17.919,4	18.309,5	18.001,7	0,001	0,977	0,001	0,040	0,985	
7	-8862,7	74	1,466	17.919,4	18.309,4	18.001,7	0,009	0,922	-0,001	0,037	0,977	
19	-8862,0	74	1,467	17.917,9	18.308,0	18.000,2	1,430	0,232	-0,046	0,038	0,231	
26	-8862,7	74	1,467	17.919,4	18.309,5	18.001,7	0,015	0,902	-0,004	0,040	0,919	
30	-8862,7	74	1,466	17.919,4	18.309,5	18.001,7	0,008	0,931	-0,001	0,038	0,989	
35	-8862,7	74	1,466	17.919,4	18.309,5	18.001,7	0,007	0,932	0,001	0,040	0,974	
42	-8862,6	74	1,465	17.919,2	18.309,3	18.001,5	0,245	0,621	-0,018	0,037	0,623	
37	-8862,7	74	1,466	17.919,4	18.309,4	18.001,6	0,091	0,763	0,011	0,038	0,776	
32	-8862,3	74	1,466	17.918,6	18.308,7	18.000,9	0,812	0,368	0,032	0,036	0,373	
23	-8861,8	74	1,466	17.917,7	18.307,7	17.999,9	1,855	0,173	0,046	0,033	0,171	
28	-8862,4	74	1,466	17.918,8	18.308,8	18.001,0	0,669	0,413	0,024	0,029	0,412	
33	-8861,8	74	1,466	17.917,7	18.307,7	17.999,9	1,834	0,176	-0,040	0,030	0,181	
24	-8859,9	74	1,465	17.913,7	18.303,7	17.995,9	6,428	0,011	0,094	0,039	0,015	
29	-8856,4	74	1,465	17.906,7	18.296,8	17.989,0	13,619	<0,001	0,149	0,042	<0,001	
44	-8861,0	74	1,466	17.915,9	18.306,0	17.998,2	3,715	0,054	-0,077	0,040	0,054	

N<sub>1</sub> = 474. Das Geschlecht der Teilnehmerinnen und Teilnehmer wurde mit 0 (= Männer) und 1 (= Frauen) kodiert. Es wurde ein robustes Maximum-Likelihood-Schätzverfahren (MLR) angewendet. Pro Item wurde ein Modell aufgestellt, welches gegen das Baseline-Modell getestet wurde. AIC Akaike's Information Criterion, BIC Bayesian Information Criterion, aBIC sample size-adjusted BIC, Npar Anzahl der freien Parameter, LL Log-Likelihood Wert, SCF Skalierungskorrekturfaktor, TRd korrigierte Differenz zwischen  $\chi^2$ -Testwerten nach Satorra & Bentler

**Zusatzmaterial online** Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s42010-022-00161-y>) enthalten.

**Förderung** Dieses Forschungsprojekt wurde unterstützt von der Tübingen Postdoctoral Academy for Research on Education (PACE) des Hector-Instituts für Empirische Bildungsforschung an der Universität Tübingen; PACE wird vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert. Das diesem Artikel zugrundeliegende Vorhaben wird im Rahmen der gemeinsamen „Qualitäts-offensive Lehrerbildung“ von Bund und Ländern mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1911 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Interessenkonflikt** T. Fütterer, L. Feder und C. Cramer geben an, dass kein Interessenkonflikt besteht.

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

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

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

## Literatur

- Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. *Advances in Experimental Social Psychology*, 20, 1–63. [https://doi.org/10.1016/S0065-2601\(08\)60411-6](https://doi.org/10.1016/S0065-2601(08)60411-6).
- Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, 52(1), 27–58. <https://doi.org/10.1146/annurev.psych.52.1.27>.
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, N.J.: Prentice-Hall.
- Albarracín, D., Sunderrajan, A., Lohmann, S., Chan, S. & Jiang, D. (2018). The psychology of attitudes, motivation, and persuasion. In D. Albarracín & B. T. Johnson (Hrsg.), *Basic Principles*. The handbook of attitudes. London: Routledge.
- Artmann, M. & Herzmann, P. (2016). Portfolioarbeit im Urteil von Studierenden. Ergebnisse einer Interviewstudie zur LehrerInnenbildung im Kölner Modellkolle. In S. Ziegelbauer & M. Gläser-Zikuda (Hrsg.), *Portfolio als Innovation in Schule, Hochschule und LehrerInnenbildung. Perspektiven aus Praxis, Forschung und Lehre* (S. 131–146). Bad Heilbrunn: Klinkhardt.
- Barton, J. & Collins, A. (1993). Portfolios in teacher education. *Journal of Teacher Education*, 44(3), 200–210. <https://doi.org/10.1177/002248719304400307>.
- Bonneville-Roussy, A., Evans, P., Verner-Filion, J., Vallerand, R.J. & Bouffard, T. (2017). Motivation and coping with the stress of assessment: gender differences in outcomes for university students. *Contemporary Educational Psychology*, 48, 28–42. <https://doi.org/10.1016/j.cedpsych.2016.08.003>.
- Borko, H., Michalec, P., Timmons, M. & Siddle, J. (1997). Student teaching portfolios: a tool for promoting reflective practice. *Journal of Teacher Education*, 48(5), 345–357. <https://doi.org/10.1177/0022487197048005004>.
- Brouër, B. (2005). Portfolio – Ein Beitrag zur schulpraktischen Kompetenzentwicklung. *karlsruher pädagogische beiträge*, 61, 7–23.

- Brouër, B. (2007). Pädagogische Portfolios in der Diplomhandelslehrer-Ausbildung. In E. M. Hertle & P. F. E. Sloane (Hrsg.), *Portfolio – Kompetenzen – Standards. Neue Wege in der Lehrerbildung für berufsbildende Schulen* (S. 31–45). Paderborn: Eusl-Verlagsgesellschaft.
- Brouër, B. & Gläser-Zikuda, M. (2010). Förderung selbstregulativer Fähigkeiten im Kontext selbstorganisierten Lernens. In J. Seifried, E. Wuttke, R. Nickolaus & P. F. E. Sloane (Hrsg.), *Lehr-Lern Forschung in der kaufmännischen Berufsbildung. Ergebnisse und Gestaltungsaufgaben* (Bd. 23, S. 123–136). Stuttgart: Franz Steiner.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36(1), 111–150. [https://doi.org/10.1207/S15327906MBR3601\\_05](https://doi.org/10.1207/S15327906MBR3601_05).
- Bryant, L. H. & Chittum, J. R. (2013). ePortfolio effectiveness: A(n ill-fated) search for empirical support. *International Journal of EPortfolio*, 3(2), 189–198.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson.
- Byrnes, J. P., Miller, D. C. & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125(3), 367–383. <https://doi.org/10.1037/0033-2909.125.3.367>.
- Cakir, A. & Balcikanli, C. (2012). The use of the EPOSTL to foster teacher autonomy: ELT student teachers' and teacher trainers' views. *Australian Journal of Teacher Education*. <https://doi.org/10.14221/ajte.2012v37n3.7>.
- Caspi, A., Chajut, E. & Saporta, K. (2008). Participation in class and in online discussions: gender differences. *Computers & Education*, 50(3), 718–724. <https://doi.org/10.1016/j.compedu.2006.08.003>.
- Chen, H.-F., Jin, K.-Y. & Wang, W.-C. (2017). Modified logistic regression approaches to eliminating the impact of response styles on DIF detection in Likert-Type scales. *Frontiers in Psychology*, 8, 1143. <https://doi.org/10.3389/fpsyg.2017.01143>.
- Christen, A. & Hofmann, M. (2013). E-Reflexionsportfolio an der Pädagogischen Hochschule St. Gallen (PHSG, Schweiz). Entwicklungen im Studienjahr 2009/2010. In D. Miller & B. Volk (Hrsg.), *E-Portfolio an der Schnittstelle von Studium und Beruf* (S. 133–146). Münster: Waxmann. [http://www.pedocs.de/frontdoor.php?source\\_opus=9811&la=de](http://www.pedocs.de/frontdoor.php?source_opus=9811&la=de).
- Chun, S., Stark, S., Kim, E. S. & Chernyshenko, O. S. (2016). MIMIC methods for detecting DIF among multiple groups: exploring a new sequential-free baseline procedure. *Applied Psychological Measurement*, 40(7), 486–499. <https://doi.org/10.1177/0146621616659738>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale, NJ: Erlbaum.
- Cramer, C. (2020). Kohärenz und Relationierung in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König, M. Rothland, & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 269–279). Bad Heilbrunn: Klinkhardt utb. <https://doi.org/10.35468/hblb2020-031>.
- Cramer, C., Harant, M., Merk, S., Drahmman, M., & Emmerich, M. (2019). Meta-Reflexivität und Professionalität im Lehrerinnen- und Lehrerberuf. *Zeitschrift für Pädagogik*, 65(3), 401–423. <https://doi.org/10.25656/01:23949>.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (5. Aufl.). Heidelberg: Springer.
- Dreyer, J. M. (2015). Reliability of recognition of prior learning (RPL) assessments for teaching practice courses in open distance electronic learning (ODEL). *International Journal of Educational Sciences*, 8(1), 111–117. <https://doi.org/10.1080/09751122.2015.11917597>.
- Feder, L. & Cramer, C. (2018). Potenziale von Portfolioarbeit in der Lehrerbildung. Eine Analyse der Portfolioliteratur. *DDS – Die Deutsche Schule*, 110(4), 354–368. <https://doi.org/10.31244/dds.2018.04.06>.
- Feder, L. & Cramer, C. (2019). Portfolioarbeit in der Lehrerbildung. Ein systematischer Forschungsüberblick. Portfolio in teacher education. A research synthesis. *Zeitschrift für Erziehungswissenschaft*, 22(5), 1225–1245. <https://doi.org/10.1007/s11618-019-00903-2>.
- Feder, L. & Cramer, C. (in Begutachtung). *Research on portfolios in teacher education: A systematic review*. Manuskript zur Veröffentlichung eingereicht.
- Feder, L., Cramer, C., Bohl, T. & Wenz, K. (2019). Portfolioarbeit in der Lehrerbildung. Potenziale – empirische Forschungslage – konzeptuelle Kontextualisierung. In BMBF (Hrsg.), *Verzahnung von Theorie und Praxis im Lehramtsstudium* (S. 40–47).
- Feder, L., Fütterer, T. & Cramer, C. (2021). Einstellungen Studierender zur Portfolioarbeit. Theoriebasierte Erfassung und erste deskriptive Befunde. In N. Beck, T. Bohl, & S. Meissner (Hrsg.), *Vielfältig herausgefordert. Forschungs- und Entwicklungsfelder der Lehrerbildung auf dem Prüfstand. Diskurse und Ergebnisse der ersten Förderphase der Qualitätsoffensive Lehrerbildung an der Tübingen School of Education* (S. 209–221). Tübingen: Tübingen University Press. <https://doi.org/10.15496/publikation-52627>.

- Fütterer, T. (2019). *Professional Development Portfolios im Vorbereitungsdienst. Die Wirksamkeit von Lernumgebungen auf die Qualität der Portfolioarbeit*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-24064-6>.
- Geiser, C. (2010). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung* (1. Aufl.). Wiesbaden: Springer VS.
- Gläser-Zikuda, M. (2009). Potenziale und Grenzen von Lerntagebuch und Portfolio im Bildungsbereich. In M. Gläser-Zikuda (Hrsg.), *Lerntagebuch und Portfolio auf dem Prüfstand* (2. Aufl., S. 95–100). Landau: Empirische Pädagogik.
- Gläser-Zikuda, M. & Hascher, T. (Hrsg.). (2007). *Lernprozesse dokumentieren, reflektieren und beurteilen. Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis*. Bad Heilbrunn: Klinkhardt.
- Gläser-Zikuda, M., Feder, L. & Hofmann, F. (2020). Portfolioarbeit in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 706–712). Bad Heilbrunn: Klinkhardt utb. <https://www.handbuch-lehrerbildung.net/download/85-portfolioarbeit-in-der-lehrerinnen-und-lehrerbildung/>.
- Glasman, L.R. & Albarracín, D. (2006). Forming attitudes that predict future behavior: a meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, 132(5), 778–822. <https://doi.org/10.1037/0033-2909.132.5.778>.
- Graham, J.W. (2012). *Missing data. Analysis and design*. New York, NY: Springer. <https://doi.org/10.1007/978-1-4614-4018-5>.
- Häcker, T. (2012). Portfolioarbeit im Kontext einer reflektierenden Lehrer/innenbildung. In R. Egger & M. Merkt (Hrsg.), *Lernwelt Universität* (S. 263–289). Wiesbaden: VS. [https://doi.org/10.1007/978-3-531-18941-3\\_15](https://doi.org/10.1007/978-3-531-18941-3_15).
- Häcker, T. & Lissmann, U. (2007). Möglichkeiten und Spannungsfelder der Portfolioarbeit – Perspektiven für Forschung und Praxis. *Empirische Pädagogik*, 21(2), 209–239.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 143–171). Berlin, Heidelberg: Springer.
- Hascher, T. & Sonntagbauer, C. (2013). Portfolio in der Lehrer/innenbildung – Bilanz, Rahmung und Ausblick. In B. Koch-Priewe, T. Leonhard, A. Pineker & J.C. Störtländer (Hrsg.), *Portfolio in der LehrerInnenbildung. Konzepte und empirische Befunde* (S. 287–298). Bad Heilbrunn: Klinkhardt.
- Hayes, A.F. & Coutts, J.J. (2020). Use omega rather than Cronbach's Alpha for estimating reliability. But ... *Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>.
- Hofmann, F., Wolf, N., Klaß, S., Grassmé, I. & Gläser-Zikuda, M. (2016). Portfolios in der LehrerInnenbildung. Ein aktueller Überblick zur empirischen Befundlage. In M. Boos, A. Krämer & M. Kricke (Hrsg.), *Portfolioarbeit phasenübergreifend gestalten: Konzepte, Ideen und Anregungen aus der LehrerInnenbildung* (S. 23–39). Münster: Waxmann.
- Holzberger, D., Maurer, C., Kunina-Habenicht, O. & Kunter, M. (2021). Ready to teach? A profile analysis of cognitive and motivational-affective teacher characteristics at the end of pre-service teacher education and the long-term effects on occupational well-being. *Teaching and Teacher Education*, 100, 103285. <https://doi.org/10.1016/j.tate.2021.103285>.
- Imhof, M. & Picard, C. (2009). Views on using portfolio in teacher education. *Teaching and Teacher Education*, 25(1), 149–154. <https://doi.org/10.1016/j.tate.2008.08.001>.
- Jin, K.-Y. & Chen, H.-F. (2020). MIMIC approach to assessing differential item functioning with control of extreme response style. *Behavior Research Methods*, 52(1), 23–35. <https://doi.org/10.3758/s13428-019-01198-1>.
- Jöreskog, K.G. & Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351), 631–639. <https://doi.org/10.2307/2285946>.
- Kelava, A. & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl. S. 75–102). Berlin, Heidelberg: Springer.
- Kirkpatrick, D.L. & Kirkpatrick, J.D. (2006). *Evaluating training programs: the four levels* (3. Aufl.). San Francisco: Berrett-Koehler.
- Klampfer, A. (2013). *E-Portfolios als Instrument zur Professionalisierung in der Lehrer- und Lehrerinnenausbildung*. Glückstadt: Hülsbusch.
- Kleinke, K., Schlüter, E., & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus: Eine praktische Einführung* (2. Aufl.). Berlin: de Gruyter.
- Koch-Priewe, B. (2013). Das Portfolio in der LehrerInnenbildung – Verbreitung, Zielsetzungen, Empirie, theoretische Fundierungen. In B. Koch-Priewe, T. Leonhard, A. Pineker & J.C. Störtländer (Hrsg.),

- Portfolio in der LehrerInnenbildung. Konzepte und empirische Befunde* (S. 41–73). Bad Heilbrunn: Klinkhardt.
- Kopietz, R. & Echterhoff, G. (2016). Einstellungen (Attitudes). In D. Bierhof & H.-W. Frey (Hrsg.), *Sozialpsychologie* (S. 581–616). Göttingen: Hogrefe.
- Kunter, M., Linninger, C., Schulze-Stocker, F., Kunina-Habenicht, O. & Lohse-Bossenz, H. (2013). *Evaluation des reformierten Vorbereitungsdienstes in Nordrhein-Westfalen*. Frankfurt am Main: Goethe-Universität.
- Kunter, M., Baumert, J., Leutner, D., Terhart, E., Seidel, T., Dicke, T., Holzberger, D., Kunina-Habenicht, O., Linninger, C., Lohse-Bossenz, H., Schulze-Stocker, F. & Stürmer, K. (2017). Dokumentation der Erhebungsinstrumente der Projektphasen des BilWiss-Forschungsprogramms von 2009 bis 2016. <http://nbn-resolving.de/urn:nbn:de:hebis:30:3-428005>. Zugegriffen: 15.11.2022.
- Lenzner, T., Neuert, C. & Otto, W. (2015). *Kognitives Pretesting*. GESIS Survey Guidelines. Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften. [https://doi.org/10.15465/GESIS-SG\\_010](https://doi.org/10.15465/GESIS-SG_010).
- Little, R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>.
- Lüdtko, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>.
- Mansolf, M. & Reise, S.P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, 61, 120–129. <https://doi.org/10.1016/j.intell.2017.01.012>.
- Monitor Lehrerbildung (2019). Vorgaben zur Führung eines Portfolios. [https://www.monitor-lehrerbildung.de/web/diagramm/blfrage26\\_1](https://www.monitor-lehrerbildung.de/web/diagramm/blfrage26_1).
- Moosbrugger, H. & Kelava, A. (Hrsg.). (2012). *Testtheorie und Fragebogenkonstruktion* (2. Aufl.). Heidelberg: Springer.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide*. Los Angeles: Muthén & Muthén.
- Noroozi, O., Banihashem, S. K., Taghizadeh Kerman, N., Parvaneh Akhteh Khaneh, M., Babayi, M., Ashrafi, H. & Biemans, H. J. A. (2022). Gender differences in students' argumentative essay writing, peer review performance and uptake in online learning environments. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2022.2034887>.
- Pineker, A. & Störtländer, J. C. (2013). Gestaltung von praktikumsbezogenen Reflexionsanlässen im Rahmen des „Bielefelder Portfolio Praxisstudien“ – Zwei hochschuldidaktische Varianten und ihre Evaluation. In B. Koch-Priewe, T. Leonhard, A. Pineker & J. C. Störtländer (Hrsg.), *Portfolio in der LehrerInnenbildung. Konzepte und empirische Befunde* (S. 193–211). Bad Heilbrunn: Klinkhardt.
- R Core Team (2021). R: A language and environment for statistical computing. <http://www.R-project.org>. Zugegriffen: 01.11.2021.
- Satorra, A. & Bentler, P.M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514. <https://doi.org/10.1007/BF02296192>.
- Schafer, J.L. & Graham, J.W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>.
- Schermelleh-Engel, K. & Gade, J.C. (2020). Modellbasierte Methoden der Reliabilitätsschätzung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 335–368). Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-662-61532-4\\_15](https://doi.org/10.1007/978-3-662-61532-4_15).
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Schlag, M. (2015). *Das strukturierte Portfolio zur Dokumentation des Orientierenden Praktikums für Lehramtsstudierende*. Mainz: Johannes Gutenberg-Universität Mainz. <https://doi.org/10.25358/openscience-1814>.
- Streblo, L., Rumpold, V. & Valdorf, N. (2013). Einschätzung der Portfolioarbeit durch Lehramtsstudierende – Empirische Ergebnisse einer studienverlaufsbegleitenden Befragung in Bielefeld. In B. Koch-Priewe, T. Leonhard, A. Pineker & J. C. Störtländer (Hrsg.), *Portfolio in der LehrerInnenbildung. Konzepte und empirische Befunde* (S. 122–135). Bad Heilbrunn: Klinkhardt.
- Strudler, N. & Wetzel, K. (2008). Costs and benefits of electronic portfolios in teacher education. *Journal of Computing in Teacher Education*, 24(4), 135–142. <https://doi.org/10.1080/10402454.2008.10784599>.
- Taber, K.S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>.

- Taylor J., Dunbar-Hall, P. & Rowley, J. L. (2012) The e-portfolio continuum: Discovering variables for e-portfolio adoption within music education. *Australasian Journal of Educational Technology*, 28(8), 1362–1381. <https://doi.org/10.14742/ajet.776>.
- Wolfsgruber, M. (2020). *Einstellung und Motivation bei der Portfolioarbeit. Eine empirische Untersuchung der Arbeit mit dem Entwicklungsportfolio am Standort Linz*. Linz: Linz School of Education.
- Ziegelbauer, S. (2016). Akzeptanz als Bedingung der erfolgreichen Implementation und Anwendung der pädagogisch-didaktischen Innovation Portfolio. In S. Ziegelbauer & M. Gläser-Zikuda (Hrsg.), *Portfolio als Innovation in Schule, Hochschule und LehrerInnenbildung. Perspektiven aus Praxis, Forschung und Lehre* (S. 73–85). Bad Heilbrunn: Klinkhardt.
- Ziegelbauer, S. & Gläser-Zikuda, M. (Hrsg.). (2016). *Portfolio als Innovation in Schule, Hochschule und LehrerInnenbildung. Perspektiven aus Praxis, Forschung und Lehre*. Bad Heilbrunn: Klinkhardt.
- Ziegelbauer, S., Ziegelbauer, C., Limprecht, S. & Gläser-Zikuda, M. (2013). Bedingungen für gelingende Portfolioarbeit in der Lehrerinnen- und Lehrerbildung – Empiriebasierte Entwicklung eines adaptiven Portfoliokonzepts. In B. Koch-Priewe, T. Leonhard, A. Pineker & J.C. Störtländer (Hrsg.), *Portfolio in der LehrerInnenbildung. Konzepte und empirische Befunde* (S. 112–121). Bad Heilbrunn: Klinkhardt.