



# Belastbarkeit von Evaluationsindikatoren in der Hochschulweiterbildung

## Psychometrische Prüfung skalenbasierter Items und ihre Eignung als Indikatoren der Input-, Prozess- und Outputqualität

Janine Marchal 

Eingegangen: 2. Dezember 2025 / Überarbeitet: 13. April 2026 / Angenommen: 15. Mai 2026  
© The Author(s) 2026

**Zusammenfassung** Der Beitrag untersucht die psychometrische Qualität skalenbasierter Evaluationsindikatoren in der Hochschulweiterbildung sowie deren Eignung für eine indikatorengestützte Bildungsberichterstattung. Grundlage der Analyse bilden 7812 standardisierte Teilnehmendenbefragungen aus über 200 formal strukturierten Zertifikatslehrgängen der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz aus dem Zeitraum 2019 bis 2025.

Untersucht werden sechs Skalen, die entlang eines Input-Prozess-Output-Modells zentrale Dimensionen wahrgenommener Qualität aus Teilnehmendensicht erfassen. Geprüft werden interne Konsistenz, Strukturvalidität, empirische Validität und Messinvarianz über Zertifikatslehrgänge und Erhebungszeitpunkte hinweg. Die Ergebnisse zeigen gute bis sehr gute Reliabilitäten, eine bestätigte dreidimensionale Faktorenstruktur sowie metrische und skalare Messinvarianz. Die Skalen eignen sich somit als standardisierte und vergleichsfähige Indikatoren für Qualitätsentwicklung und Bildungsberichterstattung. Zugleich erfassen sie ausschliesslich wahrgenommene Qualität aus Teilnehmendensicht und bilden damit nur eine spezifische Perspektive auf wissenschaftliche Weiterbildung ab.

**Schlüsselwörter** Bildungsberichterstattung · Hochschulweiterbildung · Wissenschaftliche Weiterbildung · Evaluation · Evaluationsindikatoren · Validität · Reliabilität

---

✉ Janine Marchal

Pädagogische Hochschule, Fachhochschule Nordwestschweiz FHNW, Windisch, Schweiz  
E-Mail: [janine.marchal@fhnw.ch](mailto:janine.marchal@fhnw.ch)

## Robustness of evaluation indicators in university continuing education

Psychometric testing of scale-based items and their suitability as indicators of input–process–output quality

**Abstract** This article examines the psychometric quality of scale-based evaluation indicators in university continuing education and their suitability for indicator-based educational reporting. The study is based on 7812 standardized participant surveys from over 200 formally structured certificate programmes at the University of Teacher Education, University of Applied Sciences and Arts Northwestern Switzerland, conducted between 2019 and 2025. Six scales are analysed, capturing key dimensions of perceived quality from the participants' perspective based on an input–process–output model. Internal consistency, structural validity, empirical validity, and measurement invariance across programmes and data collection points are assessed. The results show good to very good reliability, a confirmed three-factor structure, and metric and scalar measurement invariance. The scales can therefore be used as standardized and comparable indicators for quality development and educational reporting. At the same time, they capture only perceived quality from the participants' perspective and thus represent a specific view of university continuing education.

**Keywords** Educational reporting · University continuing education · Academic continuing education · Evaluation · Evaluation indicators · Reliability · Validity

### 1 Einleitung

Bildungsberichterstattung beschreibt und analysiert Bildungssysteme anhand von Indikatoren und stellt Orientierungswissen für Politik, Praxis und Wissenschaft bereit (vgl. Autor:innengruppe Bildungsberichterstattung 2024; SKBF 2023).

Eine zentrale Voraussetzung hierfür ist die Verfügbarkeit geeigneter Indikatoren, die Dimensionen von Bildungssystemen, insbesondere auch Aspekte von Qualität, abbilden. Zur Erfassung solcher Dimensionen gewinnen Evaluationsindikatoren zunehmend an Bedeutung. Sie verdichten Einzeldaten zu Kennzahlen, die Vergleiche zwischen Institutionen sowie Analysen von Entwicklungen über die Zeit hinweg ermöglichen und eine empirische Grundlage für Bildungsberichterstattung und Qualitätsentwicklung bilden. Damit dies möglich ist, müssen die zugrunde liegenden Skalen methodisch geprüft sowie reliabel und valide sein (vgl. Schrader et al. 2016; OECD 2008).

Für die Hochschulweiterbildung liegen zwar umfangreiche Evaluationsdaten zu Lehr-Lern-Prozessen, Organisation, Relevanz, Zufriedenheit und Nutzen vor (vgl. Salland 2018). Zugleich ist sie trotz ihres Bedeutungszuwachses weiterhin durch strukturelle Herausforderungen gekennzeichnet und führt im Hochschulsystem vielfach ein randständiges Dasein (vgl. Jütte 2025). Vor diesem Hintergrund ist bislang nur begrenzt systematisch untersucht, inwieweit die in Evaluationen eingesetzten Skalen reliabel, valide und über unterschiedliche Kontexte hinweg vergleichbar sind. Eine systematische Überführung dieser Daten in ein konsistentes, indikato-

rengestütztes Berichtssystem ist entsprechend bislang nur in Ansätzen ausgearbeitet (Widany et al. 2025).

Der vorliegende Beitrag setzt an der Schnittstelle von Evaluation, Qualitätssicherung, Weiterbildungsforschung und Bildungsberichterstattung an. Dabei nimmt er eine primär governance-orientierte Perspektive im Sinne empirischer Bildungsforschung ein. Im Zentrum steht die Frage, unter welchen psychometrischen Voraussetzungen standardisierte Evaluationsdaten zu Indikatoren verdichtet werden können. Diese Perspektive unterscheidet sich von subjektorientierten erwachsenenbildungswissenschaftlichen Zugängen, die Lernen, Teilnahme und Bildungsprozesse im Kontext biografischer Entwicklungen und individueller Bedeutungszuschreibungen untersuchen (vgl. Jütte und Rohs 2020; Jütte 2025).

Der Beitrag fokussiert auf jene Dimensionen wahrgenommener Qualität, die in standardisierten Teilnehmendenbefragungen erfasst und für vergleichende Analysen sowie für Bildungsberichterstattung genutzt werden können. Damit wird bewusst eine reduzierte indikatorische Perspektive eingenommen, die nicht den Anspruch erhebt, die Komplexität von Weiterbildungsprozessen vollständig abzubilden, sondern einen spezifischen Beitrag zur datenbasierten Beschreibung und Vergleichbarkeit von Weiterbildungsangeboten leistet.

Im Folgenden werden die Begriffe Hochschulweiterbildung, hochschulische Weiterbildung und wissenschaftliche Weiterbildung synonym verwendet. Im Schweizer Kontext existieren weitere Bezeichnungen, etwa tertiäre Weiterbildung oder Weiterbildung auf Hochschulebene, die unterschiedliche Perspektiven auf dieses Handlungsfeld betonen (Fischer 2014). Für die in der vorliegenden Studie untersuchten Weiterbildungsangebote wird einheitlich der Begriff „Zertifikatslehrgänge“ verwendet. Gemeint sind damit formal strukturierte Angebote wie Certificate of Advanced Studies (CAS), Diploma of Advanced Studies (DAS) und Master of Advanced Studies (MAS).

Untersucht wird die psychometrische Qualität der im Evaluationsinstrument der Pädagogischen Hochschule FHNW (PH FHNW) eingesetzten Skalen zur Bewertung dieser Zertifikatslehrgänge. Die Skalen operationalisieren Input-, Prozess- und Output-Dimensionen der Weiterbildungsqualität anhand mehrerer, inhaltlich verwandter Items. Ziel des Beitrags ist es, ihre Reliabilität, Validität und Messinvarianz zu prüfen und zu beurteilen, inwieweit sie als standardisierte und vergleichbare Indikatoren für Bildungsberichterstattung und Qualitätsentwicklung geeignet sind.

## 2 Theoretischer Rahmen und Forschungsstand

### 2.1 Kontext und Rahmenbedingungen der Hochschulweiterbildung

Die Hochschulweiterbildung hat sich zu einem eigenständigen Handlungsfeld von Hochschulen entwickelt. Hüther und Krücken (2018) betonen, dass sie zu einem wichtigen Bestandteil der sogenannten „Third Mission“ geworden ist, über die Hochschulen neben Forschung und Lehre Beiträge zur gesellschaftlichen und ökonomischen Entwicklung leisten. Eingebettet ist dieses Handlungsfeld in die Leitidee des lebenslangen Lernens, die in europäischen und internationalen Referenzrahmen ver-

ankert ist und auf die kontinuierliche Kompetenzentwicklung über den gesamten Lebensverlauf abzielt. Vor diesem Hintergrund sowie angesichts digitaler, demografischer und arbeitsmarktbezogener Transformationen wird der Hochschulweiterbildung eine zunehmende Bedeutung zugeschrieben (vgl. swissuniversities 2021; Autor:innengruppe Bildungsberichterstattung 2024).

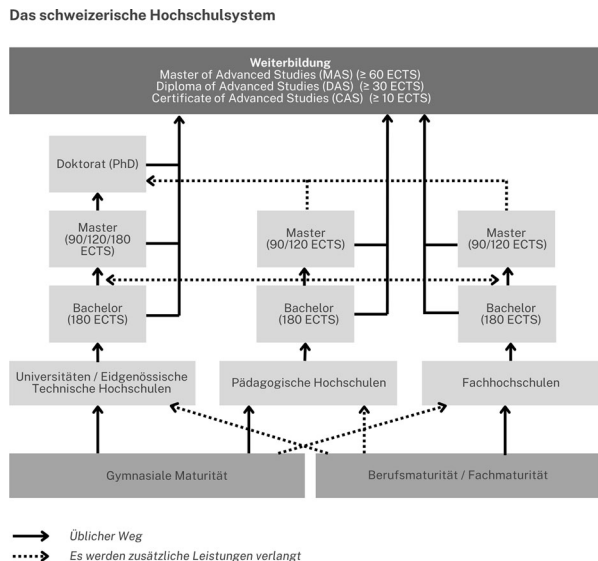
Zertifikatslehrgänge der Hochschulweiterbildung erfüllen eine doppelte Funktion: Einerseits ermöglichen sie berufstätigen Erwachsenen, ihre Kompetenzen zu erweitern, neue berufliche Perspektiven zu erschliessen und ihre Beschäftigungsfähigkeit im Wandel von Arbeits- und Lebenswelten zu sichern. Zugleich unterstützen sie den Wissenstransfer zwischen Hochschule und Praxis. Dieser Transfer kann reziprok gestaltet sein, indem wissenschaftliches Wissen in die Praxis vermittelt wird und zugleich Erfahrungen sowie Problemstellungen aus der Praxis in Hochschulen und wissenschaftliche Diskurse zurückfliessen (vgl. OECD 2019; swissuniversities 2021). Hochschulweiterbildung ist ebenso durch biografische Entwicklungen, individuelle Interessen sowie professionelle Selbstverständnisse geprägt und bewegt sich im Spannungsfeld unterschiedlicher Anforderungen (Jütte und Rohs 2020).

Die Struktur der Hochschulbildung und der Hochschulweiterbildung in der Schweiz lässt sich anhand des folgenden Schemas veranschaulichen (siehe Abb. 1).

Das Schweizer Hochschulsystem unterscheidet zwischen grundständigem Studium und formalisierten Angeboten der Hochschulweiterbildung.

Während Bachelor- und Masterstudiengänge Teil des grundständigen Studiums sind, umfasst die Hochschulweiterbildung eigenständige, meist berufs begleitend organisierte Zertifikatslehrgänge für Personen mit abgeschlossenem Hochschulstudium oder gleichwertiger Qualifikation und Berufserfahrung (vgl. swissuniversities 2021, 2025; Zimmermann 2025). Hierzu zählen insbesondere CAS-, DAS- und MAS-Programme. CAS sind kürzere, in sich abgeschlossene Formate, während DAS und MAS auf umfangreichere Qualifikationen ausgerichtet sind (vgl.

**Abb. 1** Das Schweizer Hochschulsystem. Eigene Darstellung in Anlehnung an swissuniversities (2019). Die Abbildung verdeutlicht die Differenzierung zwischen grundständigem Studium und Hochschulweiterbildung im Schweizer Hochschulsystem. Für die vorliegende Untersuchung sind insbesondere die formal strukturierten Zertifikatslehrgänge der Hochschulweiterbildung relevant, namentlich CAS, DAS und MAS.



Autor:innengruppe Bildungsberichterstattung 2024; Widany et al. 2021; swissuniversities 2021). Die vorliegende Untersuchung konzentriert sich auf diese formal strukturierten Zertifikatslehrgänge, da für sie standardisierte Evaluationsverfahren und damit vergleichbare Daten vorliegen.

Die Zielgruppen der Hochschulweiterbildung unterscheiden sich von jenen des grundständigen Studiums. Die Heterogenität zeigt sich insbesondere in unterschiedlichen beruflichen Hintergründen, Qualifikationsniveaus und institutionellen Kontexten der Teilnehmenden (siehe auch Abschnitt 3.2). Sie bestehen überwiegend aus berufstätigen Erwachsenen, deren Teilnahmeentscheidungen von Rahmenbedingungen wie zeitlicher Flexibilität, Vereinbarkeit mit beruflichen und familiären Verpflichtungen sowie finanziellen Kosten beeinflusst werden. Entsprechend gewinnen Kriterien wie inhaltliche Relevanz, Praxisbezug und institutionelle Anerkennung von Zertifikatslehrgängen an Bedeutung (vgl. Autor:innengruppe Bildungsberichterstattung 2024; OECD 2019; SKBF 2023).

Hochschulen stehen zugleich vor der Herausforderung, ihr Weiterbildungsangebot so auszurichten, dass sowohl externe Bedarfe als auch hochschuleigene organisationale Ziele berücksichtigt werden (vgl. Autor:innengruppe Bildungsberichterstattung 2024; SKBF 2023). Evaluationen können einen wichtigen Beitrag zur Analyse und Weiterentwicklung von Weiterbildungsangeboten leisten, indem sie Rückmeldungen zur wahrgenommenen Qualität, zur Passung von Inhalten, zur didaktischen Gestaltung und zum Nutzen der Angebote bereitstellen (vgl. Widany et al. 2021; Schrader et al. 2016).

Vor diesem Hintergrund stellt sich die Frage, wie Qualität in der Hochschulweiterbildung systematisch erfasst und analysiert werden kann.

## 2.2 Evaluation in der Hochschulweiterbildung

Evaluationen stellen einen zentralen Zugang dar, um Hochschulweiterbildung empirisch zu erfassen und zu bewerten. Erste Ansätze der Lehrevaluation wurden Anfang der 1990er-Jahre im deutschsprachigen Raum etabliert (Rindermann 2003). Ihre Verbreitung und institutionelle Verankerung wurden durch die wachsende gesellschaftliche Relevanz von Hochschulen sowie durch veränderte Anforderungen an Qualitätssicherung und Rechenschaft begünstigt (Salland 2018). Während sich frühe Evaluationsansätze vor allem auf die grundständige Lehre konzentrierten, wurde Evaluation in den letzten Jahrzehnten zunehmend auch auf die Hochschulweiterbildung ausgeweitet. Fragen der Qualität, Evaluation und Bildungsberichterstattung werden dabei im Feld der wissenschaftlichen Weiterbildung seit längerem theoretisch und empirisch diskutiert und systematisiert (vgl. Jütte und Rohs 2020; Widany et al. 2025).

In der Praxis kommen überwiegend standardisierte Teilnehmendenbefragungen zum Einsatz, die in der Regel am Ende von Veranstaltungen oder Modulen durchgeführt werden (Salland 2018). Diese Erhebungen erfassen unterschiedliche Dimensionen von Zertifikatslehrgängen, darunter organisationale Rahmenbedingungen, didaktische Gestaltung, Lehrkompetenz, Praxisbezug sowie die Zufriedenheit und den wahrgenommenen Nutzen aus Sicht der Teilnehmenden (Salland 2018). Die Befragungen basieren häufig auf geschlossenen Fragen mit skalenbasierten Antwortforma-

ten, die eine Standardisierung der Auswertung und die Aggregation zu Kennzahlen ermöglichen. Ergänzend können offene Fragen eingesetzt werden, um qualitative Rückmeldungen zu erfassen (vgl. Döring und Bortz 2016; Rindermann 2003).

Gleichzeitig ist zu berücksichtigen, dass Evaluationen nur einen begrenzten Ausschnitt von Weiterbildungsqualität erfassen. Sie basieren primär auf subjektiven Wahrnehmungen der Teilnehmenden und sind damit kontextgebunden und perspektivisch eingeschränkt (vgl. Salland 2018; Rindermann 2003; Jütte und Rohs 2020). Darüber hinaus erschwert die Heterogenität der Zielgruppen und Angebote eine standardisierte Erfassung, da Teilnehmende unterschiedliche berufliche Hintergründe, Vorerfahrungen, Motivationen und Nutzungskontexte mitbringen (vgl. Autor:innengruppe Bildungsberichterstattung 2024; Jütte und Rohs 2020; swissuniversities 2021).

Auch methodische Grenzen sind zu berücksichtigen. Punktueller Abschlussbefragungen erfassen längerfristige Wirkungen, etwa nachhaltige Kompetenzentwicklung oder berufliche Effekte, nur eingeschränkt (vgl. Rädiker 2013; Döring und Bortz 2016). Zudem können Antworttendenzen, situative Einflüsse oder individuelle Erwartungen die Bewertungen beeinflussen. Evaluationen bilden damit vor allem wahrgenommene Qualität unter konkreten Teilnahmebedingungen ab und lassen weiterführende Dimensionen von Weiterbildung, etwa biografische Einbettungen oder informelle Lernprozesse, nur begrenzt erkennen (Jütte und Rohs 2020).

Vor diesem Hintergrund liefern Evaluationen empirische Informationen, die zur Reflexion und Weiterentwicklung von Weiterbildungsangeboten beitragen können (vgl. Salland 2018; Jütte und Rohs 2020). In aggregierter Form können sie darüber hinaus als Grundlage für Bildungsberichterstattung dienen, indem sie zu Indikatoren verdichtet und für systematische Vergleiche zwischen Angeboten, Institutionen oder Zeitpunkten genutzt werden (vgl. Autor:innengruppe Bildungsberichterstattung 2024; Widany et al. 2021; Schrader et al. 2016; OECD 2008).

### 2.3 Indikatoren als Instrument der Bildungsberichterstattung

Indikatoren sind in der Bildungsforschung ein zentrales Instrument, um komplexe Bildungsprozesse in verdichteter und empirisch überprüfbarer Form abzubilden (vgl. Schrader et al. 2016; OECD 2008). Sie beschreiben auf Basis statistischer Kennzahlen Merkmale von Bildungssystemen (Schrader et al. 2016). Indikatoren können unter anderem aus Daten entstehen, die im Rahmen von Evaluationen erhoben werden. Einzelne Itemantworten werden zu Skalen zusammengeführt und statistisch geprüft, bevor sie als Indikatoren genutzt werden (vgl. Döring und Bortz 2016; Moosbrugger und Kelava 2020).

In der Hochschulweiterbildung können aus Evaluationsdaten abgeleitete Indikatoren eine zentrale Grundlage der externen Bildungsberichterstattung bilden. Sie schaffen Transparenz gegenüber Politik, Praxis und Öffentlichkeit und unterstützen dateninformierte Entscheidungen (vgl. Autor:innengruppe Bildungsberichterstattung 2024; SKBF 2023).

Zur theoretischen Einordnung von Indikatoren kann in der Bildungsforschung das Input–Prozess–Output-Modell (IPO) herangezogen werden. Es beschreibt systematisch Zusammenhänge zwischen Bedingungen, Prozessen und Ergebnissen (Dit-

ton 2002). Aus dieser Perspektive wird die wahrgenommene Gesamtqualität von Weiterbildung als Ergebnis des Zusammenspiels von Input-, Prozess- und Output-Merkmalen verstanden.

Das IPO-Modell unterscheidet zwischen drei Dimensionen: den Inputfaktoren wie Zugangsvoraussetzungen, Vorerfahrungen und Ressourcen der Teilnehmenden, Prozessmerkmalen wie didaktischer Gestaltung, Interaktionsqualität und organisatorischen Abläufen sowie Output-Indikatoren, wie Lernergebnissen, Zufriedenheit oder beispielsweise der Anschlussfähigkeit von Zertifikaten am Arbeitsmarkt (vgl. Schrader et al. 2016; Echarti et al. 2021). Das Modell ermöglicht es, Indikatoren systematisch zu strukturieren und Zusammenhänge zwischen den Dimensionen sichtbar zu machen (Echarti et al. 2021).

Neben dem klassischen IPO-Modell bieten neuere Modellvarianten wie das Input–Prozess–Output–Kontext-Modell (Schrader et al. 2016) oder Mehrebenenmodelle (Köller et al. 2019) erweiterte analytische Perspektiven. Sie ermöglichen differenziertere Analysen auf unterschiedlichen Ebenen (vgl. Schrader et al. 2016; Köller et al. 2019). Für die vorliegende Untersuchung ist diese zusätzliche Komplexität jedoch nicht erforderlich, da sich die Fragestellung auf die Abbildung von Qualitätsdimensionen innerhalb formal strukturierter Zertifikatslehrgänge beschränkt und das IPO-Modell für diesen Zuschnitt anschlussfähig ist.

Damit Indikatoren für Qualitätssicherung, Analyse und Berichterstattung tragfähig sind, müssen sie grundlegende Anforderungen an Datenqualität und Verständlichkeit erfüllen. Dazu zählen insbesondere die Vergleichbarkeit über Weiterbildungsangebote und Zeiträume hinweg sowie die Verständlichkeit für unterschiedliche Adressatengruppen (vgl. Autor:innengruppe Bildungsberichterstattung 2024; OECD 2008).

Für die Hochschulweiterbildung bedeutet dies, dass Indikatoren so entwickelt werden müssen, dass sie nicht nur unmittelbare Outputs wie Zufriedenheit erfassen, sondern auch Aspekte wie die Anschlussfähigkeit im individuellen Bildungsverlauf oder die berufliche Relevanz sichtbar machen. Methodisch erfordert dies die Abstimmung von Erhebungsdesigns und Modelllogiken sowie die Ableitung belastbarer aggregierter Kennzahlen (vgl. Schrader et al. 2016; OECD 2008).

## 2.4 Skalenbasierte Items und Gütekriterien

Die meisten Indikatoren, die aus Evaluationen gewonnen werden, basieren auf skalenbasierten Items. Häufig werden hierfür Likert-Skalen eingesetzt, bei denen Teilnehmende ihre Zustimmung zu Aussagen auf abgestuften Antwortkategorien angeben. Skalenbasierte Items ermöglichen eine differenzierte Erfassung von Einstellungen und Wahrnehmungen und können durch Aggregation zu Skalen verdichtet werden (Döring 2023). Für ihre Nutzung in der Hochschulweiterbildung ist entscheidend, dass sie den psychometrischen Gütekriterien entsprechen (vgl. Widany und Gerhards 2022; Moosbrugger und Kelava 2020).

Die Validität stellt in der Hochschulweiterbildung eine zentrale methodische Herausforderung dar, da die Zielgruppen hinsichtlich Berufsfeldern, Qualifikationsniveaus und Motivationslagen heterogen sind und Lernziele häufig an spezifische Anwendungskontexte gebunden bleiben (vgl. Jütte 2025; Moosbrugger und Kela-

va 2020). Items sollten daher möglichst nicht nur abstrakte Qualitätsdimensionen erfassen, sondern auch kontextsensitive Aspekte wie berufliche Relevanz oder Anwendungsnähe berücksichtigen. Während globale Zufriedenheitsitems nur eingeschränkte Interpretationsmöglichkeiten bieten (vgl. Döring und Bortz 2016; Rindermann 2003), ermöglichen präzise konstruierte und inhaltlich valide Skalen die Entwicklung anschlussfähiger Indikatoren, die über Weiterbildungsangebote, Institutionen und Zeitpunkte vergleichbar sind (OECD 2008). Unabhängig von diesen inhaltlichen Anforderungen sind bei der Konstruktion und Anwendung skalenbasierter Items auch methodische Verzerrungen zu berücksichtigen. Dazu zählen etwa Antworttendenzen wie Akquieszenz oder Präferenz für mittlere Kategorien oder soziale Erwünschtheit. Auch kulturell bedingte Bedeutungsunterschiede können die Messung beeinflussen. Hinzu kommen Stichprobeninstabilitäten, insbesondere bei kleineren Kohorten (Moosbrugger und Kelava 2020). Diese Risiken sind im Kontext der Hochschulweiterbildung besonders zu beachten, da sie die Stabilität und Vergleichbarkeit der Messung zusätzlich einschränken können.

Im Unterschied zur Validität ist die Reliabilität in der Regel weniger problematisch und lässt sich häufig über interne Konsistenzkennwerte prüfen. Gleichwohl handelt es sich auch hierbei um ein wichtiges Gütekriterium für die Beurteilung der Messqualität. Skalen sollten eine ausreichende interne Konsistenz aufweisen und über verschiedene Kohorten und Erhebungszeitpunkte hinweg stabile Ergebnisse liefern (Moosbrugger und Kelava 2020).

Insgesamt verdeutlichen diese Anforderungen, dass die Entwicklung und Prüfung skalenbasierter Instrumente als kontinuierlicher Prozess zu verstehen ist. Dieser umfasst die empirische Überprüfung von Reliabilität und Validität, die Analyse von Messinvarianz sowie gegebenenfalls die Revision der eingesetzten Items. Nur unter diesen Bedingungen können Skalen als Grundlage für belastbare und vergleichsfähige Indikatoren dienen, die sowohl wahrgenommene lehrgangsbezogene Qualität als auch übergreifende Entwicklungen im Weiterbildungssystem abbilden (vgl. Moosbrugger und Kelava 2020; Widany und Gerhards 2022).

## 2.5 Empirische Befundlage und Desiderate

Vor dem Hintergrund der bisherigen Ausführungen stellt sich die Frage, wie belastbar die empirische Evidenz zu skalenbasierten Evaluationsinstrumenten in der Hochschulweiterbildung ist, insbesondere im Hinblick auf ihre Eignung als Grundlage für indikatorengestützte Bildungsberichterstattung (vgl. Schrader et al. 2016; Widany und Gerhards 2022).

Ein Blick auf den bisherigen Forschungsstand zeigt, dass sich im Feld der wissenschaftlichen Weiterbildung ein differenzierter, jedoch fragmentierter Diskurs entwickelt hat. Frühere Arbeiten untersuchen insbesondere internationale Entwicklungen, institutionelle Strukturen und Angebotsformate (Hanft und Knust 2007). Ergänzend liefern Bildungsberichte sowie Sekundäranalysen, etwa auf Basis des Adult Education Survey (AES) – einer international vergleichenden Erhebung zur Weiterbildungsteilnahme – umfangreiche Daten zu Teilnahme, Angebotsformen und strukturellen Merkmalen (vgl. Autor:innengruppe Bildungsberichterstattung 2024; Bilger und Koubek 2024; SKBF 2023).

Neuere Arbeiten bündeln unterschiedliche Perspektiven auf wissenschaftliche Weiterbildung und beleuchten insbesondere organisationale, governance-bezogene und didaktische Aspekte des Feldes (vgl. Jütte und Rohs 2020; Widany et al. 2025). Für den Schweizer Kontext zeigen Studien zudem spezifische institutionelle Strukturen und die Ausgestaltung der Zertifikatsformate (CAS, DAS, MAS) auf (Zimmermann 2025). Darüber hinaus wird der Diskurs zu Bildungsberichterstattung, Qualität und Evaluation in der wissenschaftlichen Weiterbildung in einschlägigen Fachzeitschriften und Publikationsreihen kontinuierlich geführt und weiterentwickelt.

Trotz dieser breiten Wissensbasis wird in der Literatur wiederholt darauf hingewiesen, dass die Qualität der zugrunde liegenden Messinstrumente bislang nur begrenzt systematisch untersucht ist. Darüber hinaus zeigen bestehende Arbeiten, dass sowohl amtliche Statistiken als auch andere Datenquellen bislang keine umfassende und systematische Abbildung wissenschaftlicher Weiterbildung ermöglichen. Insbesondere fehlen empirische Analysen, die die psychometrischen Eigenschaften eingesetzter Skalen, etwa hinsichtlich Reliabilität, Validität und Messinvarianz, systematisch prüfen (vgl. Kuper et al. 2016; Widany et al. 2025).

Im Vergleich zu anderen Bildungsbereichen – insbesondere Schule und grundständigem Studium – liegen für die Hochschulweiterbildung bislang nur wenige Studien vor, die Reliabilität, Validität und Faktorenstrukturen von Evaluationsskalen systematisch untersuchen (vgl. Ditton 2002; Moosbrugger und Kelava 2020; Widany et al. 2025). Stattdessen konzentrieren sich viele Arbeiten auf globale Zufriedenheitsmasse oder einzelne Qualitätsaspekte, ohne deren psychometrische Fundierung vertieft zu analysieren (vgl. Salland 2018; Widany und Gerhards 2022).

Damit bleibt offen, inwieweit die in der Hochschulweiterbildung eingesetzten skalenbasierten Instrumente die notwendigen messtechnischen Voraussetzungen erfüllen, um als Grundlage für belastbare und vergleichsfähige Indikatoren zu dienen. Insbesondere liegen bislang nur begrenzt systematische Analysen zur Messinvarianz sowie zur Stabilität der Skalen über verschiedene Zertifikatslehrgänge und Erhebungszeitpunkte hinweg vor (vgl. Schrader et al. 2016; Widany und Gerhards 2022).

Ausgehend davon positioniert sich die vorliegende Studie im bestehenden Forschungsfeld. Sie folgt einer primär governance- und berichtorientierten Perspektive im Sinne der empirischen Bildungsforschung und fokussiert auf die Frage, unter welchen psychometrischen Voraussetzungen Evaluationsdaten in belastbare und vergleichsfähige Indikatoren überführt werden können.

Die Untersuchung setzt an dieser Forschungslücke an und leistet einen Beitrag zur empirischen Fundierung indikatorengestützter Bildungsberichterstattung in der Hochschulweiterbildung. Ziel ist es zu prüfen, inwieweit die im Evaluationsinstrument eingesetzten Skalen reliabel, valide und messinvariant sind und somit als Grundlage für standardisierte und vergleichsfähige Indikatoren genutzt werden können (vgl. Moosbrugger und Kelava 2020; Widany und Gerhards 2022).

Vor diesem Hintergrund werden folgende Forschungsfragen adressiert:

1. Weisen die Skalen des Evaluationsinstruments eine zuverlässige interne Konsistenz auf, und bildet sich die theoretisch erwartete Struktur der Qualitätsdimensionen Input, Prozess, Output in den empirischen Daten ab?

2. Lassen sich die Skalen als valide Indikatoren der Qualitätsdimensionen Input, Prozess und Output belegen und voneinander abgrenzen?
3. Bleiben die Skalen über verschiedene Zertifikatslehrgänge und Zeitpunkte hinweg stabil, sodass Vergleiche von Mittelwerten und Entwicklungen über die Zeit methodisch zulässig sind?
4. Eignen sich die Skalen zur Bildung standardisierter, vergleichsfähiger Indikatoren, die sowohl für die interne Qualitätssicherung als auch für die Bildungsberichterstattung nutzbar sind?

### 3 Methodik

#### 3.1 Forschungsdesign und Datenbasis

Aufbauend auf den in Kapitel 2 dargestellten theoretischen und empirischen Überlegungen sowie den formulierten Forschungsfragen untersucht die Studie die psychometrische Qualität der im Evaluationsinstrument der PH FHNW eingesetzten Skalen.

Die Untersuchung verfolgt einen quantitativ-psychometrischen Forschungsansatz, der auf der statistischen Analyse von Evaluationsdaten beruht und die Gütekriterien Reliabilität, Validität und Messinvarianz prüft (vgl. Döring und Bortz 2016; Moosbrugger und Kelava 2020).

Als Datengrundlage diente eine Sekundäranalyse institutioneller Evaluationsdaten, die zwischen dem Herbstsemester 2019 und dem Frühjahrssemester 2025 im Rahmen von Zertifikatslehrgängen der PH FHNW erhoben wurden. Die Analyse umfasste standardisierte Teilnehmendenbefragungen, die am Ende von Lehrveranstaltungen im Rahmen des institutionellen Qualitätssicherungsprozesses durchgeführt wurden. Einbezogen wurden formal strukturierte Zertifikatslehrgänge der Hochschulweiterbildung (CAS, DAS, MAS), für die einheitliche Evaluationsprozesse vorliegen, die wiederholte Erhebungen unter vergleichbaren Bedingungen zulassen. Diese standardisierten Erhebungsbedingungen ermöglichten die Analyse der psychometrischen Eigenschaften und der Vergleichbarkeit der erhobenen Daten.

#### 3.2 Stichprobe

Die Datengrundlage umfasst insgesamt  $n = 7812$  ausgewertete Teilnehmendenbefragungen aus über 200 Zertifikatslehrgängen verschiedener Fachbereiche. Über den gesamten Erhebungszeitraum entfallen rund 97 % der Teilnehmenden auf CAS-Lehrgänge, etwa 2 % auf MAS- und rund 0,5 % auf DAS-Zertifikatslehrgänge. Die Teilnehmenden stammen überwiegend aus pädagogischen und sozialwissenschaftlichen Berufsfeldern, darunter Lehrpersonen, Schulleitungen, pädagogische Fachpersonen sowie Fachpersonen der Erwachsenenbildung. Die Stichprobe ist damit durch unterschiedliche berufliche Hintergründe, Qualifikationsniveaus und institutionelle Kontexte gekennzeichnet. Weitere demografische Merkmale wurden im Rahmen des Evaluationsprozesses nicht systematisch erfasst und konnten daher in der Analyse nicht berücksichtigt werden.

Die Erhebung erfolgte online über die Evaluationsplattform EvaSys. Der Befragungslink wurde jeweils am Ende eines Moduls bereitgestellt. Die durchschnittliche Rücklaufquote lag bei rund 85 %.

Der Anteil fehlender Werte ist bei der Mehrzahl der Items unter 1,5 %. Lediglich zwei Items weisen höhere Ausfallraten zwischen rund 3 % und 5 % auf. Da der Anteil fehlender Werte insgesamt gering ist und keine systematischen Ausfallmuster erkennbar sind, werden Fälle mit fehlenden Angaben mittels Listwise Deletion aus den Analysen ausgeschlossen (Little und Rubin 2019). Aufgrund des insgesamt niedrigen Anteils fehlender Werte ist davon auszugehen, dass alternative Imputationsverfahren die Ergebnisse nur geringfügig beeinflussen würden.

Die Evaluationen sind Bestandteil eines institutionell verankerten Qualitätssicherungsprozesses und weisen eine hohe Rücklaufquote auf, sodass das Risiko einer ausgeprägten Selbstselektionsverzerrung als begrenzt eingeschätzt werden kann.

Insgesamt umfasste die Stichprobe unterschiedliche Disziplinen, Zielgruppen und Zertifikatslehrgänge (CAS, MAS, DAS) und bot damit eine geeignete Grundlage für die psychometrischen Analysen.

### 3.3 Erhebungsinstrument

Die Datenerhebung erfolgte mit dem standardisierten Evaluationsinstrument der PH FHNW, das in allen Zertifikatslehrgängen einheitlich eingesetzt wird. Der Fragebogen umfasst 15 geschlossene Items auf einer vierstufigen Likert-Skala (1 = stimme nicht zu, 4 = stimme zu). Die Entwicklung der Items war theoriegeleitet und orientierte sich an bestehenden Ansätzen zur Gestaltung von Hochschul- und Weiterbildungsangeboten sowie an institutionellen Vorgaben der PH FHNW. Dabei wurden zentrale Aspekte wie etwa didaktische Gestaltung, Praxisbezug und organisationale Rahmenbedingungen mit hochschulischen Anforderungen und Evaluationszielen verknüpft und in konkrete Aussagen überführt, die von den Teilnehmenden eingeschätzt wurden (vgl. Salland 2018; Schrader et al. 2016; Widany und Gerhards 2022).

Die Operationalisierung folgte einem deduktiv-pragmatischen Vorgehen, bei dem theoretisch fundierte Qualitätsdimensionen mit institutionellen Anforderungen der Qualitätssicherung verbunden und in beobachtbare Indikatoren (Items) überführt wurden. Dabei wurde das Konstrukt der Weiterbildungsqualität als wahrgenommene Qualität aus Teilnehmendensicht konzeptualisiert und entlang der Input-Prozess-Output-Logik differenziert operationalisiert. Für jede Dimension wurden inhaltliche Aspekte identifiziert und in mehrere, konkret formulierte Aussagen überführt, die typische Merkmale von Zertifikatslehrgängen abbilden (z. B. Passung der Inhalte zur beruflichen Praxis, didaktische Gestaltung, organisatorische Abläufe, wahrgenommener Nutzen).

Die Items wurden so formuliert, dass sie von Teilnehmenden auf Basis ihrer konkreten Erfahrungen im Zertifikatslehrgang beurteilt werden können und gleichzeitig eine standardisierte, vergleichbare Erfassung über unterschiedliche Lehrgänge hinweg ermöglichen. Die Zuordnung der Items zu den Skalen erfolgte theoriegeleitet und wurde im Rahmen der anschließenden psychometrischen Analysen empirisch überprüft. Ziel war es, zentrale Aspekte wahrgenommener Weiterbildungsqualität

abzubilden und zugleich eine standardisierte, organisationsweit einsetzbare Erhebung zu gewährleisten.

Die Qualitätsdimensionen wurden über Skalen operationalisiert, die aus mehreren inhaltlich verwandten Items bestanden. Die Zuordnung folgte der Input-Prozess-Output-Logik (Ditton 2002) und knüpfte an aktuelle Qualitätsansätze der Hochschul- und Weiterbildungsforschung an (vgl. Salland 2018; Schrader et al. 2016; Widany und Gerhards 2022).

- Inputqualität bezieht sich auf die Passung der Lehrinhalte zu den Lernzielen und zur beruflichen Praxis der Teilnehmenden. Dazu gehört die Skala *Inhaltliche Relevanz* (z. B. „Die behandelten Themen waren für meine berufliche Praxis relevant.“).
- Zur Prozessqualität zählen die Skalen *Didaktische Gestaltung* und *Organisation* (z. B. „Die Lehrperson förderte aktives Lernen“ oder „Die organisatorische Abwicklung war reibungslos.“).
- Outputqualität wird über die Skalen *Praxisbezug*, *Zufriedenheit* und *Nutzen* erfasst (z. B. „Ich kann das Gelernte in meinem beruflichen Alltag anwenden.“ oder „Ich würde diese Weiterbildung weiterempfehlen.“).

Für jede der sechs Skalen wurden Skalenmittelwerte berechnet. Eine Skala wurde gebildet, wenn mindestens 50 % der zugehörigen Items beantwortet waren. Dieses Vorgehen entspricht gängigen Verfahren der Skalenbildung und ermöglicht eine stabile Schätzung der zugrunde liegenden Konstrukte trotz einzelner fehlender Werte (vgl. Döring und Bortz 2016; Moosbrugger und Kelava 2020).

### 3.4 Datenanalyse

Die Datenanalyse orientierte sich an etablierten Verfahren der quantitativen Skalenprüfung (vgl. Brown 2015; Creswell 2020; Döring 2023) und hatte zum Ziel, die psychometrische Qualität sowie die indikatorische Verwendbarkeit der eingesetzten Skalen zu beurteilen.

#### 1. Reliabilität

Die interne Konsistenz der Skalen wurde mittels Cronbach's Alpha bestimmt. Ergänzend wurden Item-Trennschärfen analysiert, um die Homogenität der Skalen und die Qualität der einzelnen Items zu beurteilen.

#### 2. Strukturvalidität

Die Strukturvalidität wurde mithilfe einer konfirmatorischen Faktorenanalyse (CFA) geprüft. Es wurde ein dreifaktorielles Messmodell spezifiziert, das die theoretisch postulierten Qualitätsdimensionen Input-, Prozess- und Outputqualität abbildet. Die Schätzung erfolgte mit WLSMV (Weighted Least Squares Mean and Variance adjusted), um der ordinalen Skalierung der vierstufigen Likert-Items Rechnung zu tragen.

Die Modellgüte wurde anhand der Fit-Indizes CFI, TLI, RMSEA und SRMR beurteilt (Orientierungswerte: CFI/TLI  $\geq 0,90$ – $0,95$ ; RMSEA  $\leq 0,08/0,06$ ; SRMR  $\leq 0,08$ ). Zur Bewertung der Messstruktur wurden standardisierte Faktorladungen herangezogen (Richtwert  $\geq 0,40$ ).

### 3. Empirische Validität

Die konvergente und diskriminante Validität wurde anhand der Interkorrelationen der Skalen sowie der Ladungsmuster in den CFA-Modellen geprüft. Dabei wurde erwartet, dass Skalen innerhalb einer IPO-Dimension stärker miteinander korrelieren als zwischen unterschiedlichen Dimensionen.

### 4. Messinvarianz

Die Vergleichbarkeit der Skalen über Zertifikatslehrgänge und Erhebungszeitpunkte wurde mittels Multi-Group-CFA geprüft. Invarianz wurde angenommen, wenn sich der Comparative Fit Index zwischen Modellstufen um weniger als  $\Delta\text{CFI}=0,01$  veränderte.

### 5. Indikatorentauglichkeit

Zur Einschätzung der Eignung der Skalenwerte als Bildungsindikatoren wurden deskriptive Kennwerte (Mittelwert, Standardabweichung, Spannweite) analysiert. Die Ergebnisse wurden herangezogen, um die Stabilität und Vergleichbarkeit der Indikatoren über Zertifikatslehrgänge und Zeitpunkte hinweg einzuschätzen.

Alle Analysen wurden mit RStudio (Version 2025.09.2) durchgeführt.

## 4 Ergebnisse

Die Darstellung der Ergebnisse orientiert sich an den vier Forschungsfragen und umfasst die Prüfung der Reliabilität und Strukturvalidität (F1), die empirische Validität (F2), die Messinvarianz über Zertifikatslehrgänge und Zeitpunkte hinweg (F3) sowie die Indikatorentauglichkeit der Skalen (F4). Grundlage der Analysen bilden die sechs Skalen des Evaluationsinstruments, die den Qualitätsdimensionen Input, Prozess und Output zugeordnet sind.

### 4.1 Reliabilität und Strukturvalidität (F1)

Zur Beurteilung der internen Konsistenz wurden für alle sechs Skalen Cronbach's  $\alpha$  berechnet. Die Reliabilitäten liegen insgesamt im guten bis sehr guten Bereich (siehe Tab. 1). Die Skala der Inputqualität (Inhaltliche Relevanz) weist bei zwei Items eine akzeptable interne Konsistenz auf ( $\alpha=0,72$ ). Die Prozessqualität umfasst zwei Skalen mit insgesamt sieben Items: Didaktische Gestaltung ( $\alpha=0,88$ ) und Organisation ( $\alpha=0,83$ ), beide mit sehr guten internen Konsistenzen. Die drei Skalen der Output-

**Tab. 1** Interne Konsistenz der Skalen (Cronbach's  $\alpha$ , Trennschärfenbereich)

Dimension	Skala	Items	$\alpha$	r_items (Min–Max)
Input	Inhaltliche Relevanz	2	0,72	0,57–0,63
Prozess	Didaktische Gestaltung	4	0,88	0,49–0,76
Prozess	Organisation	3	0,83	0,52–0,70
Output	Praxisbezug	2	0,81	0,55–0,67
Output	Zufriedenheit	2	0,85	0,58–0,71
Output	Nutzen	2	0,87	0,60–0,74

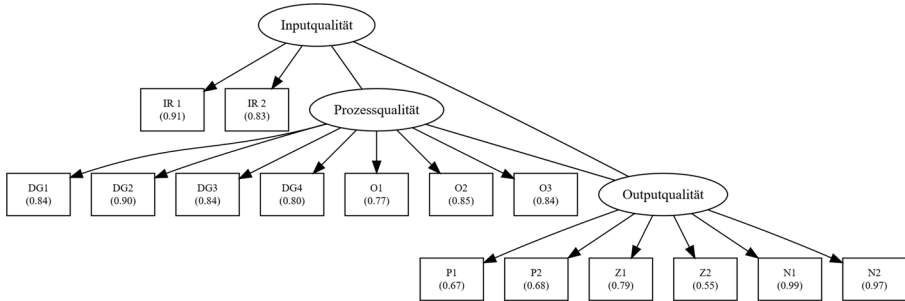
Qualität mit jeweils zwei Items Praxisbezug ( $\alpha=0,81$ ), Zufriedenheit ( $\alpha=0,85$ ) und Nutzen ( $\alpha=0,87$ ) erreichen ebenfalls gute bis sehr gute Reliabilitäten.

Die Item-Trennschärfen liegen für alle Skalen im mittleren bis hohen Bereich ( $r \approx 0,49-0,76$ ), und kein Item unterschreitet einen kritischen Schwellenwert.

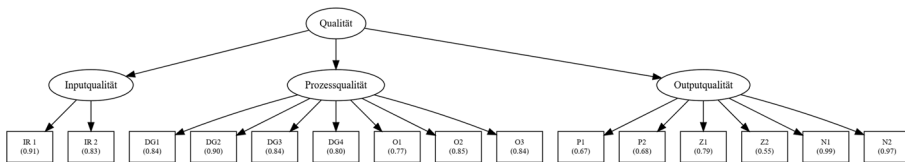
Zur Prüfung der Strukturvalidität wurde eine konfirmatorische Faktorenanalyse (CFA) mit WLSMV-Schätzung durchgeführt. Das Modell umfasst die drei theoretisch begründeten Dimensionen der Inputqualität, Prozessqualität und Outputqualität mit jeweils zugeordneten Item-Bündeln. Die Items wurden als ordinale Indikatoren spezifiziert (Brown 2015).

Das Drei-Faktoren-Modell zeigt eine gute Passung ( $\chi^2(87)=2503,56, p<0,001$ ; CFI=0,997; TLI=0,996; RMSEA=0,063, 90%-KI [0,061, 0,065]; SRMR=0,050). Die standardisierten Faktorladungen sind durchweg hoch und signifikant, wie in Abb. 2 dargestellt (alle Ladungen  $p<0,001$ ). Die Interfaktorkorrelationen liegen zwischen 0,90 und 0,93.

Das ergänzend geschätzte Second-Order-Modell mit einem übergeordneten Qualitätsfaktor zeigt eine vergleichbare Modellpassung (CFI=1,00; TLI=0,99; RMSEA=



**Abb. 2** Strukturgleichungsmodell zur dreidimensionalen Qualitätskonzeption. Dargestellt ist ein konfirmatorisches Dreifaktorenmodell mit den Dimensionen Inputqualität, Prozessqualität und Outputqualität. Rechtecke repräsentieren die beobachteten Items mit ihren standardisierten Faktorladungen, Ellipsen die latenten Faktoren. Zwischen den drei Faktoren sind Korrelationen zugelassen. Höhere Faktorladungen verweisen auf eine stärkere Zuordnung des jeweiligen Items zur entsprechenden Qualitätsdimension. Die Itembezeichnungen setzen sich aus der Abkürzung der Skala und der Itemnummer zusammen (z. B. IR1 = Inhaltliche Relevanz 1).



**Abb. 3** Second-Order-Modell mit übergeordnetem Qualitätsfaktor. Dargestellt ist ein Second-Order-Modell, in dem die drei Qualitätsdimensionen Inputqualität, Prozessqualität und Outputqualität als Faktoren erster Ordnung durch einen übergeordneten Qualitätsfaktor zweiter Ordnung zusammengefasst werden. Rechtecke repräsentieren die beobachteten Items mit ihren standardisierten Faktorladungen, Ellipsen die latenten Faktoren. Die Pfade vom übergeordneten Qualitätsfaktor zu den drei Dimensionen geben an, in welchem Ausmass diese durch eine gemeinsame latente Qualitätsdimension erklärt werden. Die Itembezeichnungen setzen sich aus der Abkürzung der Skala und der Itemnummer zusammen (z. B. IR1 = Inhaltliche Relevanz 1).

**Tab. 2** Deskriptive Kennwerte und Interkorrelationen der Skalen

Skala	M	SD	1	2	3	4	5	6
1. Inhaltliche Relevanz	3,56	0,47	–	–	–	–	–	–
2. Didaktische Gestaltung	3,68	0,42	0,41	–	–	–	–	–
3. Organisation	3,74	0,38	0,33	0,57	–	–	–	–
4. Praxisbezug	3,61	0,45	0,38	0,64	0,49	–	–	–
5. Zufriedenheit	3,79	0,35	0,36	0,69	0,58	0,62	–	–
6. Nutzen	3,67	0,41	0,32	0,59	0,52	0,55	0,68	–

$N=7755$ ; alle Korrelationen  $p < 0,001$  (Pearson)

0,063; SRMR=0,050). Der übergeordnete Faktor erklärt rund 91 % der gemeinsamen Varianz der drei Qualitätsdimensionen (siehe Abb. 3).

## 4.2 Empirische Validität (F2)

Zur Prüfung der empirischen Validität wurden deskriptive Kennwerte und Interkorrelationen der sechs Skalen berechnet (siehe Tab. 2). Die Skalenmittelwerte liegen durchgängig im oberen Bereich der vierstufigen Antwortskala. Die Standardabweichungen zeigen eine ausreichende Streuung der Antworten.

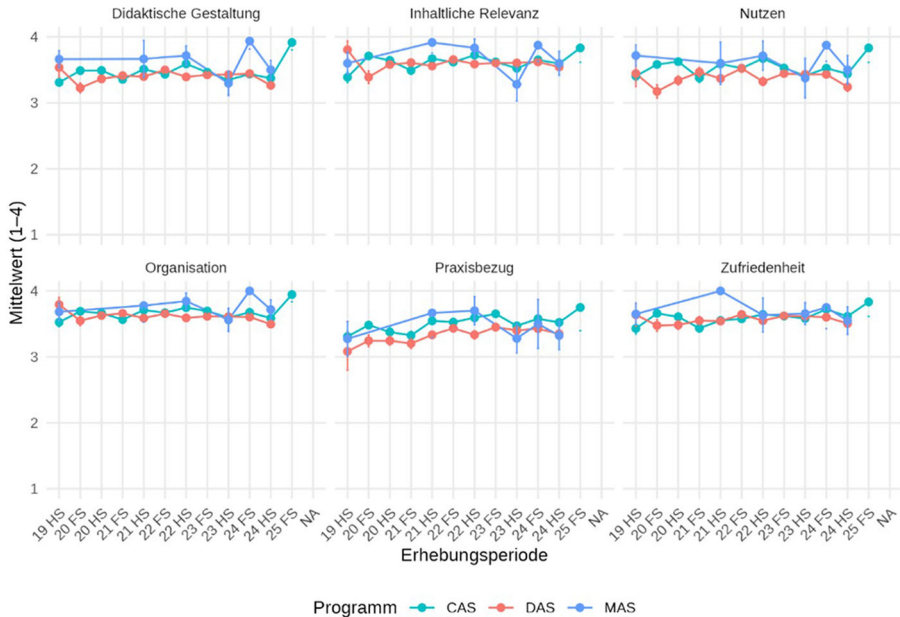
Die Interkorrelationen der sechs Skalen liegen im mittleren Bereich und spiegeln die inhaltliche Nähe der Qualitätsdimensionen wider. Die höchsten Zusammenhänge ergeben sich zwischen den Skalen Didaktische Gestaltung und Zufriedenheit sowie Nutzen und Zufriedenheit. Die Skala Inhaltliche Relevanz, die der Input-Dimension zugeordnet ist, weist hingegen geringere Korrelationen zu den auf Prozess und Output bezogenen Skalen auf. Insgesamt zeigen die Skalen ein differenziertes Korrelationsmuster.

## 4.3 Messinvarianz (F3)

Zur Beurteilung der Vergleichbarkeit der Skalen über Typen von Zertifikatslehrgängen und Erhebungszeitpunkte hinweg wurden sequenzielle Multi-Group-CFA-Analysen durchgeführt (konfigurale, metrische und skalare Stufe). Die Bewertung der Invarianz erfolgte anhand der Fit-Indizes (CFI, TLI, RMSEA, SRMR) sowie der Veränderungswerte ( $\Delta$ CFI) zwischen den Modellstufen. Eine Veränderung von  $|\Delta$ CFI  $< 0,01$  wurde als Kriterium für Invarianz herangezogen (vgl. Brown 2015).

Über Zertifikatslehrgänge hinweg (CAS, DAS, MAS) zeigt das konfigurale Modell eine gute Modellanpassung (CFI=0,954; TLI=0,935; RMSEA=0,076; SRMR=0,042). Die Gleichsetzung der Faktorladungen im metrischen Modell führt zu keiner relevanten Verschlechterung der Modellgüte ( $\Delta$ CFI=–0,001), ebenso wenig die Gleichsetzung der Schwellenwerte im skalar invarianten Modell ( $\Delta$ CFI=–0,003). Damit kann von metrischer und skalarer Invarianz über die Zertifikatslehrgänge hinweg ausgegangen werden.

Über die Erhebungszeitpunkte 2019 bis 2025 zeigt sich ein vergleichbares Muster. Das konfigurale Modell ergibt ebenfalls eine gute Passung (CFI=0,962; TLI=



**Abb. 4** Skalenmittelwerte nach Zertifikatslehrgang und Erhebungsperiode. Dargestellt sind die Mittelwerte der sechs Skalen getrennt nach Zertifikatslehrgang und Erhebungsperiode. Die vertikalen Linien zeigen 95%-Konfidenzintervalle und veranschaulichen die Unsicherheit der Mittelwertschätzungen.

0,946; RMSEA = 0,071; SRMR = 0,046). Die metrische ( $\Delta\text{CFI} = -0,004$ ) und die skalare ( $\Delta\text{CFI} = -0,007$ ) Invarianz werden bestätigt. Damit bleibt die Messstruktur der Skalen über die Zeit stabil, wodurch Veränderungen in den Mittelwerten als inhaltliche Entwicklungen und nicht als Messartefakte interpretierbar sind.

Deskriptive Mittelwerte zeigen ein konsistentes Muster. Die Antworten bleiben über die Zertifikatslehrgänge und Zeiträume hinweg weitgehend konstant. Leichte Unterschiede zeigen sich insbesondere in den Skalen Praxisbezug und Nutzen.

Abb. 4 visualisiert die Entwicklung der Skalenmittelwerte der Typen von Zertifikatslehrgängen über die Erhebungsperioden hinweg.

#### 4.4 Indikatorentauglichkeit (F4)

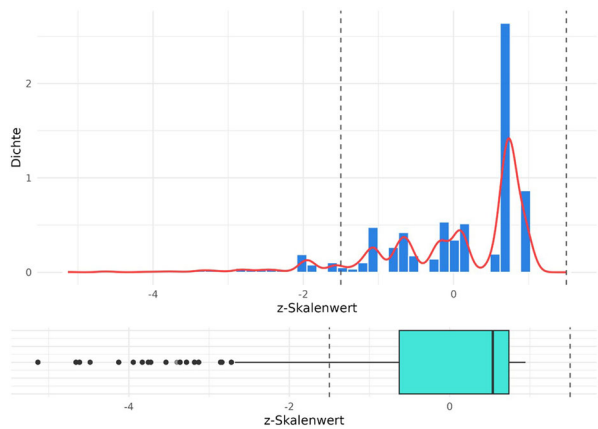
Zur Prüfung der Indikatorentauglichkeit wurden zusätzliche beschreibende Kennwerte der Skalen herangezogen (siehe Tab. 3).

Fehlende Werte liegen bei allen Skalen unter 1%. Zur besseren Vergleichbarkeit wurden z-standardisierte Skalenwerte berechnet, auf deren Grundlage Zertifikatslehrgänge und Zeitvergleiche möglich sind. Der Wertebereich der z-Werte liegt zwischen  $-1,06$  und  $+2,18$ . Die Verteilung zeigt eine Konzentration im mittleren Bereich mit wenigen Ausreißern. Die Korrelationen der z-Werte zwischen den Skalen bewegen sich im mittleren bis hohen Bereich ( $r = 0,32$ – $0,71$ ,  $p < 0,001$ ), ohne Hinweise auf übermäßige Redundanz.

**Tab. 3** Deskriptive Kennwerte der Skalen (Mittelwert, Standardabweichung, Spannweite, fehlende Werte)

Skala	N	Fehlende (%)	M	SD	Min	Max	Range
1. Inhaltliche Relevanz	7755	0,72	3,56	0,47	1	4	3
2. Didaktische Gestaltung	7755	0,72	3,68	0,42	1	4	3
3. Organisation	7754	0,73	3,74	0,38	1	4	3
4. Praxisbezug	7750	0,78	3,61	0,45	1	4	3
5. Zufriedenheit	7740	0,91	3,79	0,35	1	4	3
6. Nutzen	7737	0,95	3,67	0,41	1	4	3

**Abb. 5** Verteilung der z-standardisierten Skalenwerte über alle Zertifikatslehrgänge. Das Histogramm mit überlagerter Dichtekurve zeigt die Verteilung der z-standardisierten Skalenwerte über alle Zertifikatslehrgänge. Die zentrale Hälfte der Werte liegt zwischen  $-0,55$  und  $+0,65$ ; Werte ausserhalb von  $\pm 1,5$  treten nur in einem kleinen Anteil der Fälle auf



Analysen nach Zertifikatslehrgang (CAS, DAS, MAS) zeigen keine signifikanten Mittelwertunterschiede ( $\eta^2 < 0,01$ ), und auch über die Erhebungszeitpunkte hinweg variieren die Werte nur geringfügig ( $\Delta M \leq 0,08$ ).

Die Verteilung der z-standardisierten Skalenwerte ist in Abb. 5 dargestellt. Die Grafik zeigt eine kompakte, leicht linksschiefe Verteilung mit nur wenigen Ausreissern. Die mittleren 50% der Werte liegen zwischen  $-0,55$  und  $+0,65$ , Werte ausserhalb  $\pm 1,5$  treten in weniger als 3% der Fälle auf.

## 5 Diskussion

Im Mittelpunkt des Beitrags steht die Prüfung der psychometrischen Qualität jener Skalen, mit denen Teilnehmende Zertifikatslehrgänge beurteilen. Im Folgenden werden die Ergebnisse im Hinblick auf Reliabilität, Validität und Messinvarianz zusammengeführt, theoretisch im Input-Prozess-Output-Modell verortet und hinsichtlich ihrer praktischen Bedeutung für Qualitätsentwicklung und Berichterstattung diskutiert.

## 5.1 Interpretation der zentralen Befunde

Die Ergebnisse deuten insgesamt auf eine konsistente und stabile Struktur der gemessenen Qualitätsdimensionen. Die Reliabilitätsanalysen stützen die Annahme, dass die Skalen zentrale Aspekte der wahrgenommenen Weiterbildungsqualität gemäss IPO-Modell zuverlässig erfassen.

Die durchweg guten Trennschärfen zeigen, dass die Items zuverlässig zwischen unterschiedlichen Ausprägungen der Zustimmung unterscheiden. Zugleich weisen die homogenen Skalen darauf hin, dass zentrale Qualitätsaspekte differenziert erfasst werden, aber als zusammengehörige Facetten eines umfassenden Qualitätsverständnisses wahrgenommen werden.

Die Interkorrelationen zwischen den Skalen bestätigen, dass Qualität in der hochschulischen Weiterbildung als mehrdimensionales, zugleich aber zusammenhängendes Konstrukt beschrieben werden kann. Trotz der teilweise hohen Zusammenhänge bleibt die diskriminante Validität erhalten, da die Faktoren im CFA unterscheidbar bleiben.

Auffällig ist, dass Zufriedenheit besonders eng mit der didaktischen Gestaltung und dem wahrgenommenen Nutzen verknüpft ist, während die inhaltliche Relevanz als Inputdimension etwas stärker abgesetzt bleibt. Dies kann darauf hindeuten, dass Gesamtzufriedenheit stärker mit Prozess- und Output-Merkmalen verknüpft ist, während die inhaltliche Passung der Themen eine eigenständige Rolle für die Qualitätswahrnehmung spielt.

Dieses Muster lässt sich auch theoretisch im IPO-Modell einordnen, wonach Input als rahmende Voraussetzung verstanden wird, während Prozess- und Output-Aspekte die erlebte Qualität stärker prägen (vgl. Ditton 2002; Schrader et al. 2016).

Besonders hervorzuheben ist die empirisch belegte Stabilität der Skalen über Zertifikatslehrgänge und Erhebungszeitpunkte hinweg. Die nachgewiesene Messinvarianz belegt, dass die Skalen über verschiedene Zertifikatsformate (CAS, DAS, MAS) und über mehrere Jahre hinweg (2019–2025) vergleichbare latente Konstrukte erfassen.

Die  $\Delta$ CFI-Werte lagen in allen Invarianztests unter dem Schwellenwert von 0,01, was die Stabilität der Messstruktur zusätzlich untermauert. Damit können Unterschiede in den Mittelwerten als Unterschiede in der wahrgenommenen Qualität interpretiert werden und nicht primär durch strukturelle Verzerrungen der Messung erklärbar sind.

Darüber hinaus deuten die geringen Varianzen und die kompakte Verteilung der z-standardisierten Skalenwerte auf eine relativ hohe Ähnlichkeit der Bewertungen innerhalb der untersuchten Kontexte hin. Die kompakten z-Werte (zwischen  $-1,06$  und  $+2,18$ ) bestätigen die geringe Ausreisseranfälligkeit und unterstützen die Vergleichbarkeit der Indikatoren.

Die Bewertungen fallen über Zertifikatslehrgänge hinweg ähnlich aus, was sowohl auf ein vergleichbares Qualitätsniveau als auch auf die Robustheit des Evaluationsinstruments hinweist. Gleichzeitig bleibt ausreichend Streuung erhalten, sodass Unterschiede zwischen einzelnen Zertifikatslehrgängen erkennbar bleiben. Dies spricht dafür, dass die Skalen grundsätzlich für Zwecke der Qualitätsentwicklung genutzt werden können, da Unterschiede identifizierbar bleiben.

Insgesamt liefern die Befunde Hinweise darauf, dass die eingesetzten Skalen eine tragfähige Grundlage für die empirische Erfassung wahrgenommener Weiterbildungsqualität darstellen. Sie können dazu beitragen, Entwicklungen über Zeit und Zertifikatslehrgänge hinweg abzubilden und Evaluationen nicht nur als punktuelle Rückmeldungen, sondern auch im Kontext organisationaler Beobachtungs- und Reflexionsprozesse zu nutzen.

## 5.2 Theoretische Rückbindung und forschungsbezogene Einordnung

Die Ergebnisse lassen sich gut in den theoretischen Rahmen des Input-Prozess-Output-Modells einordnen. Sie bestätigen, dass Qualität durch das Zusammenspiel von Input-, Prozess- und Output-Merkmalen beschrieben werden kann.

Das dreidimensionale Modell erweist sich gegenüber einfacheren Modellvarianten als theoretisch plausibel, da es ermöglicht, Input-, Prozess- und Output-Qualität analytisch voneinander zu unterscheiden. Gleichzeitig legt die gute Passung des Second-Order-Modells nahe, dass die einzelnen Dimensionen auf einem gemeinsamen, übergeordneten Qualitätskonstrukt beruhen (siehe Abb. 2 und 3). Qualität wird von den Teilnehmenden ganzheitlich bewertet, aber entlang bestimmter, inhaltlich interpretierbarer Facetten strukturiert.

Insgesamt zeigte sich, dass Qualität in den Zertifikatslehrgängen nicht auf Zufriedenheit oder einzelne Aspekte reduziert werden kann, sondern als mehrdimensionales Konstrukt zu verstehen ist.

Gleichzeitig ist hervorzuheben, dass diese Modellierung eine analytische Reduktion komplexer Weiterbildungsrealitäten darstellt. Aus erwachsenenbildungswissenschaftlicher Perspektive bleiben insbesondere biografische Kontexte, individuelle Motivlagen sowie subjektive Bedeutungszuschreibungen unterbelichtet, sodass die Ergebnisse als komplementär zu stärker subjektorientierten Forschungszugängen zu verstehen sind (Jütte und Rohs 2020).

## 5.3 Methodische Reflexion

Die psychometrische Analyse weist auf mehrere, methodische Aspekte hin, die für zukünftige Forschung zur Evaluation hochschulischer Weiterbildung relevant sind. Erstens zeigen die ausgeprägten Ceiling-Tendenzen, dass Zertifikatslehrgänge insgesamt positiv bewertet werden (siehe Tab. 3). Dies entspricht Befunden aus der Evaluationsforschung, weist aber zugleich darauf hin, dass die Differenzierung im oberen Qualitätsbereich eingeschränkt ist (vgl. Courey und Lee 2021). Gleichwohl ist die verbleibende Varianz ausreichend, um Unterschiede zwischen Zertifikatslehrgängen und Erhebungszeitpunkten abzubilden.

Zweitens legen die hohen Interfaktorkorrelationen nahe, Modelle höherer Ordnung in Betracht zu ziehen. Die gute Passung des getesteten Second-Order-Modells deutet darauf hin, dass die drei Qualitätsdimensionen unterscheidbare Facetten darstellen, zugleich aber auf einem gemeinsamen übergeordneten Qualitätskonstrukt beruhen. Dies verweist auf die Frage nach geeigneten Modellierungen, etwa Second-Order- oder bifaktoriellen Modellen.

Drittens weist die geringe Fehlwertquote darauf hin, dass das Instrument im Erhebungskontext gut akzeptiert und die Items weitgehend vollständig beantwortet wurden. Die Antwortverteilungen, Trennschärfen und Faktorladungen zeigen ein konsistentes und plausibles Muster auf, sodass keine offensichtlichen Hinweise auf ausgeprägte Antworttendenzen oder strukturelle Messprobleme vorliegen.

Viertens zeigen die Ergebnisse zur Messinvarianz, dass die Skalen über Zertifikatslehrgänge und Erhebungszeitpunkte hinweg vergleichbare latente Konstrukte erfassen. Damit lassen sich Unterschiede in Skalenwerten als tatsächliche Unterschiede in der wahrgenommenen Qualität interpretieren und nicht als Messartefakte.

Insgesamt belegen die Ergebnisse, dass die eingesetzten Skalen ein robustes und theoretisch fundiertes Messinstrument darstellen. Sie erfassen die zentralen Dimensionen wahrgenommener Weiterbildungsqualität zuverlässig, zeigen eine konsistente Struktur und weisen eine methodisch abgesicherte Messstabilität auf.

## 6 Fazit und Implikationen

Die Ergebnisse zeigen, dass die eingesetzten Skalen wahrgenommene Weiterbildungsqualität reliabel, valide und über Zertifikatslehrgänge sowie Erhebungszeitpunkte hinweg stabil erfassen. Damit bilden sie eine tragfähige Grundlage für standardisierte, vergleichsfähige Indikatoren in Qualitätsentwicklung und Bildungsberichterstattung.

### 6.1 Praktische Relevanz und Implikationen für die Hochschulpraxis

Für Hochschulen ergeben sich aus den Befunden vor allem drei praktische Implikationen. Erstens können die geprüften Skalen für die systematische Qualitätsentwicklung von Zertifikatslehrgängen genutzt werden, da sie Veränderungen über die Zeit vergleichbar abbilden. Zweitens unterstützen sie die Nutzung von Evaluationen als Feedbackinstrumente für Lehrende und Verantwortliche, indem Rückmeldungen strukturiert ausgewertet und in die Weiterentwicklung von Lehrangeboten einbezogen werden können. Drittens eröffnen aggregierte Skalenwerte Möglichkeiten für lehrgangsübergreifende Vergleiche und für die strategische Steuerung im Rahmen hochschulischer Governance. Zugleich bleibt zu berücksichtigen, dass diese Anwendungen auf subjektiven Einschätzungen beruhen und daher nur eine spezifische Perspektive auf Qualität abbilden.

### 6.2 Bedeutung für die Bildungsberichterstattung

Die Ergebnisse zeigen, dass die untersuchten Skalen als Bausteine einer indikatoren-gestützten Bildungsberichterstattung in der Hochschulweiterbildung genutzt werden können. Sie bilden strukturierte Aspekte wahrgenommener Weiterbildungsqualität ab und ermöglichen Vergleiche über Zertifikatslehrgänge und Zeitpunkte hinweg. Institutionelle Unterschiede können sich dabei etwa in curricularer Struktur, organisatorischer Einbettung, Zusammensetzung der Teilnehmendengruppen und fachlichen Profilen der Zertifikatslehrgänge zeigen. Zugleich wird deutlich, dass eine tragfähige

ge Bildungsberichterstattung auf verlässliche und standardisierte Datengrundlagen angewiesen ist. Dies gilt in der wissenschaftlichen Weiterbildung in besonderem Mass, da vorhandene Datenquellen bislang nur Teilaspekte des Feldes abbilden und Monitoringansätze weiterhin mit begrenzter Vergleichbarkeit und einer hybriden Verortung zwischen Hochschulbildung, Weiterbildung und beruflicher Bildung umgehen müssen (Widany et al. 2025). Die Befunde zeigen damit, wie Evaluationsdaten systematisch in Berichtssysteme integriert werden können.

### 6.3 Grenzen der Studie und Ausblick auf weitere Forschung

Trotz der insgesamt positiven Befunde weist die Studie einige Einschränkungen auf. Erstens beruhen die Analysen auf subjektiven Selbstauskünften und erlauben daher keine direkten Aussagen zu tatsächlichen Kompetenzgewinnen oder längerfristigen Wirkungen. Zweitens ist die Untersuchung auf einen institutionellen Kontext beschränkt, sodass die Übertragbarkeit auf andere Hochschulen und Formate offenbleibt. Drittens enthält der Datensatz keine individuellen Verlaufsdaten, weshalb Aussagen zu Veränderungen auf Individualebene nicht möglich sind. Künftige Forschung sollte daher weitere Datenquellen, Replikationsstudien und längsschnittliche Designs einbeziehen. Darüber hinaus besteht Bedarf an weiterentwickelten Dateninfrastrukturen, die Evaluation, Qualitätsmanagement und Bildungsberichterstattung stärker verknüpfen. Dies entspricht auch aktuellen Befunden zum Monitoring wissenschaftlicher Weiterbildung, die auf Lücken in Datenverfügbarkeit, Abgrenzung und institutionenübergreifender Vergleichbarkeit hinweisen (Widany et al. 2025).

### 6.4 Gesamtbewertung und Ausblick

Insgesamt zeigt die Studie, dass standardisierte Evaluationsskalen ein zentrales Instrument zur evidenzorientierten Weiterentwicklung der Hochschulweiterbildung darstellen können. Sie ermöglichen die Erfassung wahrgenommener Qualität entlang theoretisch fundierter Dimensionen sowie deren Vergleich über Zeit und Zertifikatslehrgänge hinweg.

Für Hochschulen eröffnet dies die Möglichkeit, Weiterbildung systematisch in Qualitätsmanagement und strategische Steuerung einzubinden.

Die Untersuchung leistet damit einen Beitrag zu einer dateninformierten Weiterbildungsberichterstattung. Zugleich zeigt sie, dass Dateninfrastrukturen und Forschungsperspektiven weiterentwickelt werden müssen, um das Potenzial indikatorengestützter Berichterstattung vollständig auszuschöpfen.

**Funding** Open Access funding enabled by Deutsches Institut für Erwachsenenbildung (DIE).

**Data Availability** The datasets analysed during the current study are not publicly available due to data protection and institutional confidentiality restrictions.

### Declarations

**Conflict of Interests** J. Marchal gibt 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

- Autor:innengruppe Bildungsberichterstattung (Hrsg.). (2024). *Bildung in Deutschland 2024: Ein indikatorengestützter Bericht mit einer Analyse zu beruflicher Bildung*. <https://doi.org/10.3278/6001820iw>
- Bilger, F., & Koubek, E. (2024). *Weiterbildungsverhalten in Deutschland 2022: Ergebnisse des Adult Education Survey – AES-Trendbericht*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2. Aufl.). New York: Guilford.
- Courey, K. A., & Lee, M. D. (2021). A model-based examination of scale effects in student evaluations of teaching. *AERA Open*. <https://doi.org/10.1177/23328584211040083>.
- Creswell, J. W. (2020). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6. Aufl.). London: Pearson Education.
- Ditton, H. (2002). Unterrichtsqualität – Konzeptionen, methodische Überlegungen und Perspektiven. *Unterrichtswissenschaft*, 30(3), 197–212. <https://doi.org/10.25656/01:7685>.
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-64762-2>.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-41089-5>.
- Echarti, N., Reichart, E., & Gerhards, P. (2021). Die Abbildung von Wirkungen beruflicher Weiterbildung in der Bildungsberichterstattung: Bestand und Perspektiven. *Zeitschrift für Weiterbildungsforschung*, 44(3), 311–331. <https://doi.org/10.1007/s40955-021-00196-1>.
- Fischer, A. (2014). *Hochschulweiterbildung in einem heterogenen Feld. Bericht zu Händen der Geschäftsstelle des Schweizerischen Wissenschafts- und Innovationsrats*. Bern: Schweizerischer Wissenschafts- und Innovationsrat (SWIR). Arbeitsdokument Geschäftsstelle SWIR 3/2014
- Hanft, A., & Knust, M. (2007). *Internationale Vergleichsstudie zur Struktur und Organisation der Weiterbildung an Hochschulen*. Oldenburg: Carl von Ossietzky Universität.
- Hüther, O., & Krücken, G. (2018). *Higher education in Germany—recent developments in an international perspective*. Bd. 49. Cham: Springer. <https://doi.org/10.1007/978-3-319-61479-3>.
- Jütte, W. (2025). Wissenschaftliche Weiterbildung. In P. Pasternack, G. Reinmann & C. Schneijderberg (Hrsg.), *Hochschulforschung: Forschung über Hochschule und Wissenschaft* (S. 411–420). Baden-Baden: Nomos. <https://doi.org/10.25656/01:33927>.
- Jütte, W., & Rohs, M. (Hrsg.). (2020). *Handbuch Wissenschaftliche Weiterbildung* (1. Aufl.). Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-17643-3>.
- Köller, O., Hasselhorn, M., Hesse, F. W., Maaz, K., Schrader, J., Solga, H., Spieß, C. K., & Zimmer, K. (Hrsg.). (2019). *Das Bildungswesen in Deutschland: Bestand und Potenziale* (1. Aufl.). Stuttgart: utb. <https://doi.org/10.36198/9783838547855>.
- Kuper, H., Widany, S., & Kaufmann, K. (2016). Empfehlungen für eine Statistik zur Bildung Erwachsener in Deutschland: Strategien, Indikatoren und Datengewinnung. In *Wissenschaftliche Diskussionspapiere* (Bd. 176, S. 74–106).
- Little, R., & Rubin, D. (2019). *Statistical analysis with missing data* (3. Aufl.). Wiley. <https://doi.org/10.1002/9781119482260>.
- Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion* (3. Aufl.). Wiesbaden: Springer. <https://doi.org/10.1007/978-3-662-61532-4>.
- OECD (2008). *Handbook on constructing composite indicators: Methodology and user guide* (1. Aufl.). OECD. <https://doi.org/10.1787/9789264043466-en>.
- OECD (2019). *Getting skills right: Future-ready adult learning systems*. OECD. <https://doi.org/10.1787/9789264311756-en>.

- Rädiker, S. (2013). *Evaluation von Weiterbildungsprozessen. Status quo, Herausforderungen, Kompetenzanforderungen*. Marburg: Tectum.
- Rindermann, H. (2003). Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. *Zeitschrift für Evaluation*, 3(2), 233–256.
- Salland, C. (2018). Studiengangevaluation als Element der Qualitätssicherung in der wissenschaftlichen Weiterbildung. Konzept und Ergebnisse. In W. Seitter, M. Friese & P. Robinson (Hrsg.), *Wissenschaftliche Weiterbildung zwischen Implementierung und Optimierung* (S. 147–175). Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-19652-3\\_7](https://doi.org/10.1007/978-3-658-19652-3_7).
- Schrader, J., Kuper, H., & Behringer, F. (2016). *Entwicklung von Indikatoren und einer Datengewinnungsstrategie für die Weiterbildungsstatistik in Deutschland*. Bonn: BIBB.
- SKBF (2023). *Bildungsbericht Schweiz 2023*. Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.
- swissuniversities. (2019). Das schweizerische Hochschulsystem. [https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Lehre/ENIC/2019\\_Hochschulsystem\\_d\\_e.pdf](https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Lehre/ENIC/2019_Hochschulsystem_d_e.pdf). Zugegriffen: 20. April 2026.
- swissuniversities (2021). *Zukunft der Hochschulweiterbildung*. [www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Lehre/Themen/211026\\_Zukunft\\_HS-WB\\_DE.pdf](http://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Lehre/Themen/211026_Zukunft_HS-WB_DE.pdf). Zugegriffen: 25. Nov. 2025.
- swissuniversities (2025). *Qualifikationsrahmen für den schweizerischen Hochschulbereich*. <https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Lehre/NQR/nqf-ch-HS-d.pdf>. Zugegriffen: 10. Nov. 2025.
- Widany, S., & Gerhards, P. (2022). „Wer braucht diese Äpfel und Birnen?“ Wahrnehmung und Nutzung von Daten der Weiterbildungsstatistik durch weiterbildungspolitische Akteure. *Zeitschrift für Bildungsforschung*, 12(1), 145–163. <https://doi.org/10.1007/s35834-022-00339-5>.
- Widany, S., Reichart, E., Christ, J., & Echarti, N. (Hrsg.). (2021). *Trends der Weiterbildung – DIE-Trendanalyse 2021*. <https://www.die-bonn.de/doks/2021-DIE-Trendanalyse.pdf>
- Widany, S., Wolter, A., & Dollhausen, K. (2025). Monitoring wissenschaftlicher Weiterbildung: Status quo und Perspektiven. In W. Jütte & M. Rohs (Hrsg.), *Handbuch Wissenschaftliche Weiterbildung*. Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-17674-7\\_35-2](https://doi.org/10.1007/978-3-658-17674-7_35-2).
- Zimmermann, T. E. (2025). Wissenschaftliche Weiterbildung in der Schweiz. In W. Jütte & M. Rohs (Hrsg.), *Handbuch Wissenschaftliche Weiterbildung*. Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-17674-7\\_33-2](https://doi.org/10.1007/978-3-658-17674-7_33-2).

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