

Florian Berding & Heike Jahncke

## **Anwendung, Nutzung und Grenzen von Beobachtungsbögen als Instrument zur Professionalisierung im Hochschulstudium**

### **Zusammenfassung**

Der Lehrerberuf zeichnet sich professionstheoretisch dadurch aus, dass ähnlich wie bei Mediziner\*innen eine Einführung in die Handlungspraxis, den sogenannten Habitus, notwendig ist. Studien zur Expertiseforschung zeigen dabei, dass für hochwertige Leistungen Berufserfahrung allein nicht ausreicht, sondern stetige (Selbst-)Reflexion des eigenen Handelns erforderlich ist (vgl. z. B. Helsper, 2011; Palmer et al., 2005). Für das Hochschulstudium bedeutet dies, dass eine Einführung in den professionellen Habitus durch (Selbst-)Reflexionsgelegenheiten angeboten werden sollte. Eine Möglichkeit hierzu bieten Beobachtungsbögen, die als Grundlage für (Selbst-)Reflexionsprozesse dienen. Diese werden von Lehrenden ausgefüllt und erfassen das konkrete Handeln angehender Professionsmitglieder. Bislang ist jedoch kaum bekannt, wie Beobachtungsbögen konkret in der Hochschullehre eingesetzt werden, welche Faktoren die Bewertung beeinflussen und ob Beobachtungsbögen tatsächlich ein geeignetes Instrument zur Professionalisierung Studierender darstellen. Der vorliegende Beitrag widmet sich diesem Forschungsdesiderat anhand einer Stichprobe von 390 Studierenden aus Bachelor- und Mastermodulen mit 840 Beobachtungen, die von 9 Dozent\*innen durchgeführt wurden. Die Ergebnisse zeigen, dass zu stark ausdifferenzierte Beobachtungsbögen nicht zielführend sind, die Handhabung der Lehrenden variiert und eine Optimierung individuellen Lernens, nicht jedoch die Leistungsbewertung, mittels Beobachtungsbögen sinnvoll ist.

### **Schlüsselwörter**

Professionalisierung, Beobachtungsbögen, (Selbst-)Reflexion, Habitus

## **Application, use and limitations of observation forms as a tool for professionalization in higher education**

### **Abstract**

Professional theory claims that for teachers, as for medical professionals, an introduction to the practice of action, the so-called habitus, is necessary. Expertise research shows that

professional experience alone is not enough for high-quality services, but that perpetual (self-)reflection on one's own actions is required (see, for example, Helsper, 2011, Palmer et al., 2005). For higher education, this means that an introduction to the professional habitus through (self-)reflection opportunities is required. One possibility for this are observation forms, which serve as a basis for (self-)reflection processes. These are filled out by professors to capture concrete actions of prospective members of the profession. So far, however, little is known about how observation forms are practically used in higher education, which factors influence the assessments, and whether observation forms are a suitable instrument to professionalize students. This article analyses this based on a sample of 390 students from Bachelor's and Master's modules with 840 observations conducted by 9 lecturers. Results show that too granular observation forms are ineffective, that the professors' handling varies and that an optimization of individual learning, but not performance evaluation by means of grades, makes sense.

### Keywords

Professionalization, observation forms, (self-) reflection, habitus

## 1 Einleitung

Die Einordnung eines Berufes als Profession ist oft mit der Annahme verbunden, dass Professionen besondere gesellschaftliche Leistungen erbringen (vgl. z. B. Reinisch, 2009, S. 36). Dabei umfassten die im spätmittelalterlichen Europa entstandenen Professionen ursprünglich nur die Theologie, Medizin und Jurisprudenz (vgl. Stichweh, 2005, S. 31). Im Sinne der Professionstheorie von Oevermann (1996) wird die Leistung eines bzw. einer Professionellen immer dann notwendig, wenn ein Problem mit alltäglichen Mitteln und Methoden nicht mehr gelöst werden kann (vgl. Helsper 2011, S. 149). Zu einer Problemlösung trägt der bzw. die Professionelle dadurch bei, indem bewährtes Wissen in einer besonderen Art und Weise eingesetzt wird, einerseits ingenieurl, d. h. durch Deduktion bzw. Subsumtion aus standardisiertem Wissen, andererseits interventionspraktisch durch Rekonstruktion des individuellen Einzelfalls in der Art, dass der Einzelfall durch Deutung, Interpretation und Auslegung unter das standardisierte Wissen subsumiert werden kann (vgl. Helsper, 2011, S. 150; Oevermann, 2008, S. 58 f.). Reinisch (2009, S. 34) fasst diese Grundidee prägnant zusammen:

„Dementsprechend besteht professionelles Handeln aus zwei in sich widersprüchlichen Elementen: Anwendung universeller Regeln und Verstehen des je besonderen Falls. Deren Einheit herzustellen, macht die besondere Leistung des Professionellen aus.“

Während die Kenntnis universeller und standardisierter Regeln durch den Erwerb wissenschaftlichen Wissens z. B. im Rahmen eines Studiums möglich wird, kann die nicht-standardisierbare Rekonstruktion des Einzelfalls nicht durch ein wissenschaftliches Studium erlernt werden, sondern bedarf einer gesonderten Einführung in die professionelle Handlungspraxis (vgl. Helsper, 2011, S. 150).

Die Professionstheorie von Oevermann (1996) knüpft an der Medizin an und leitet aus einem Vergleich der Tätigkeiten von Lehrkräften mit denen von Ärzt(inn)en die Professionsbedeutung des Lehrerberufs ab. Beispielsweise zeigt sich die *ingenieurale Wissensanwendung* im Lehrerberuf darin, dass Lehrkräfte im Rahmen des Studiums wissenschaftliche Erkenntnisse darüber erwerben, welche *universellen* Regeln zur Wirkung bestimmter Methoden und didaktischer Entscheidungen bestehen. So zeigt die Selbstbestimmungstheorie der Motivation beispielsweise, dass ein Unterricht, der das Autonomieerleben, das Kompetenzerleben und die soziale Eingebundenheit der Lernenden fördert, grundsätzlich eine höhere Chance auf motivierte Lernende besitzt als ein Unterricht, der diese Bedürfnisse nicht befriedigt (vgl. z. B. Deci & Ryan, 1993; 2012; Ryan & Deci, 2002). Die Meta-Analyse von Sung, Chang und Liu (2016) zeigt darüber hinaus, dass die Verwendung mobiler Endgeräte (z. B. Smartphones, Tablets) grundsätzlich eine höhere Chance auf einen guten Lernerfolg besitzt als die Verwendung von Laptops.

Die *interventionspraktische Rekonstruktion des Einzelfalls* zeigt sich im Lehrerberuf hingegen darin, dass eine Lehrkraft die Eigenschaften und Voraussetzungen jedes einzelnen Lernenden verstehen sollte, sodass er oder sie die *universellen* Regeln durch Deutung und Interpretation auf den einzelnen Lernenden abgestimmt anwenden kann. Die Rekonstruktion des Einzelfalls hat aber noch weitere Konsequenzen:

„Aus (der) rekonstruktiven Handlungslogik resultiert die Anfälligkeit professionellen Handelns für Fehler, die konstitutiv zum professionellen Handeln gehören und deren offene, kollegiale Reflexion ein Kennzeichen der professionalisierten Handlungspraxis ist“ (Helsper, 2011, S. 150).

Die besondere Leistung einer Lehrkraft besteht folglich darin, das Universelle mit dem Individuellen sinnvoll in Einklang zu bringen. Konkret heißt dies beispielsweise, dass eine Lehrkraft für den Einzelfall klären sollte, wie das Autonomieerleben für eine(n) bestimmte(n) Schüler(in) unterstützt werden kann (z. B. durch das Angebot verschiedener Aufgaben im Unterricht) oder wie ein(e) Lernende(r) Smartphones in Abhängigkeit seiner/ihrer Vorkenntnisse im Unterricht nutzen könnte. Das Erlernen dieser besonderen Leistungsfähigkeit erfolgt neben dem Erwerb wissenschaftlichen Wissens im Studium (erste Professionalisierung) zusätzlich durch eine Einführung in den professionellen Habitus, welche die zweite Professionalisierung darstellt (vgl. Helsper, 2011, S. 150).

In der Lehrerbildung in Deutschland wird diese doppelte Professionalisierung durch eine gestufte Ausbildung erzielt. Die erste Phase besteht dabei aus dem Studium und dient vor allem der ersten Professionalisierung. Das sich anschließende Referendariat stellt die zweite Phase der Lehrerbildung dar und setzt durch Einführung in die Handlungspraxis die zweite Professionalisierung um. Spätestens seit den Bologna-Prozessen kann die Einführung in die Handlungspraxis jedoch nicht mehr allein der zweiten Phase der Lehrerbildung zugeordnet werden. Vielmehr gewinnt im Rahmen von Bologna eine Berufsfeld- und Kompetenzorientierung im Studium deutlich an Bedeutung (vgl. z. B. Schubarth et al., 2011, S. 75). In der Folge sind Ansätze für die Hochschullehre notwendig, die Studierende beim Erwerb des professionellen Habitus unterstützen, da das Studium der wissenschaftlichen Erkenntnisse dafür nicht ausreichend ist (vgl. Helsper, 2011, S. 150). In der Lehrerbildung werden zu diesem Zweck seit langem Beobachtungsbögen eingesetzt. Hierbei beobachten Professionelle die angehenden Lehrkräfte bei der tatsächlichen Umsetzung beruflicher

Handlungen und geben ihnen auf der Grundlage des Beobachtungsbogens eine individuelle Rückmeldung zu ihrem beruflichen Handeln. Neben einer Professionalisierung des Handelns soll damit in der Regel gleichzeitig die (Selbst-)Reflexionskompetenz der angehenden Professionellen gestärkt werden, die eine grundlegende Fähigkeit für professionelles Handeln darstellt (vgl. dazu Helsper, 2011, S. 150).

Während die Förderung der (Selbst-)Reflexionsfähigkeit im Studium durch Feedback auf der Grundlage von Beobachtungsbögen bereits empirisch belegt werden konnte (vgl. Jahncke et al., 2018), ist über die konkrete Nutzung von Beobachtungsbögen durch Lehrende an Hochschulen und die damit verbundenen Chancen und Grenzen bislang wenig bekannt. Ziel des vorliegenden Beitrages ist es daher zu klären, wie Beobachtungsbögen im Lehramtsstudium verwendet werden, welche Chancen und Risiken sich aus der Nutzung dieser ergeben und wo sinnvolle Einsatzgebiete von Beobachtungen bzw. Beobachtungsbögen für die Professionalisierung von angehenden Lehrkräften bestehen.

## 2 Theoretischer Hintergrund

### 2.1 (Selbst-)Reflexion als Teil professionellen Handelns

Professionelles Handeln ist aufgrund der notwendigen individuellen Fallrekonstruktion stets fehleranfällig und fordert daher von den Mitgliedern einer Profession eine reflektierte Haltung gegenüber ihrem beruflichen Handeln (vgl. Helsper, 2011, S. 150). (Selbst-)Reflexion kann dabei verstanden werden als die Kompetenz eines/einer Professionellen, das eigene Handeln sowie die äußeren Rahmenbedingungen insbesondere dann in Frage zu stellen, wenn dieses nicht die gewünschten Effekte bewirkt. Die (Selbst-)Reflexionskompetenz lässt sich dabei über folgende vier Facetten genauer beschreiben (vgl. Jahncke, 2019):

- Der (1) *Blick zurück in die Vergangenheit* ermöglicht es Lehrkräften, die notwendigen Erkenntnisse zu sammeln, die für ihre eigenen Lernprozesse oder Veränderungen der Rahmenbedingungen notwendig sind (vgl. z. B. Dewey, 1910; Göhlich, 2011; Leonhard & Abels, 2017; Schön 1987).
- Die explizite Berücksichtigung des (2) *Blicks nach vorne in die Zukunft* führt dazu, aus den generierten Erkenntnissen auch tatsächliche Konsequenzen für Veränderungen zu formulieren (vgl. z. B. Dewey, 1910; Korthagen, 1999; Schön, 1987). Auf diese Weise wird sichergestellt, dass die gesammelten Erkenntnisse nicht folgenlos bleiben.
- Der Einbezug des (3) *Blicks nach innen in das Individuum* selbst erlaubt es, die eigene Person und ihre Handlungen zum Gegenstand von Reflexionen bzw. Selbstreflexionen zu machen, dadurch Lernprozesse anzustoßen und zur Professionalisierung aus eigener Kraft beizutragen (vgl. z. B. Abels, 2011; Janík, 2003; Stock & Riebenbauer, 2014).

- Der Einbezug äußerer Rahmenbedingungen (4) (*Blick nach außen*) in ein Konzept der Reflexion erlaubt es schließlich, Veränderungen anzustoßen, die für Lernprozesse relevant sind, deren Ursprung aber nicht in der individuellen Lehrkraft selbst liegt (z. B. vorgegebene Stundeneinteilungen, Anschaffung bestimmter Lehr-/Lernmaterialien) (vgl. z. B. Bain, Ballantyne & Packer, 1999; Janík, 2003; Nguyen et al., 2014).

Die Modellierung von (Selbst-)Reflexion über diese vier Facetten ist über eine umfangreiche Metastudie erarbeitet worden (vgl. Jahncke, 2019). Sie bietet insgesamt die Chance, ein Verständnis des Konstrukts zu entwickeln, aus dem sich Verbesserungen für die allgemeine Lehrpraxis, aber auch für die individuelle und persönliche Entwicklung einer Lehrkraft ergeben. Sie knüpft an die Standards der KMK (2014) an, dabei insbesondere an die Kompetenzbereiche *Unterrichten* und *Innovieren*. Hiernach sollen Lehrkräfte Fachleute für das Lehren und Lernen sein, die ihre eigenen Kompetenzen ständig weiterentwickeln und eine aktive gestalterische Rolle für Lehr-/Lernprozesse einnehmen (vgl. KMK, 2014). Das Unterrichten ist dabei unbestritten das Kerngeschäft der Lehrertätigkeit (vgl. Baumert & Kunter, 2006; Graff, 2014). Unterrichten selbst ist ein facettenreiches Unterfangen in einem komplexen sozialen Gefüge, das durch Unvorhersehbares, Unwägbarkeiten, Unplanbarkeiten gekennzeichnet ist und nicht durch Rezeptwissen bewältigt werden kann (vgl. Gage, 1978). Dabei erschließt bereits die erste Phase der Lehrerbildung ausgehend von dem Schwerpunkt Theorie die pädagogische Praxis und hierbei zentral das Unterrichten (vgl. KMK, 2004).

Im Rahmen des Studiums konnte bereits nachgewiesen werden, dass die Verwendung von Feedback auf der Grundlage von Beobachtungen wesentlich zur (Selbst-)Reflexionsfähigkeit von angehenden Lehrkräften beiträgt (vgl. Jahncke et al., 2018). Unklar ist hingegen noch, wie die Beobachtungsbögen bzw. Beobachtungen konkret durchgeführt werden.

## 2.2 Beobachtungen und Beobachtungsbögen

Beobachtungsbögen können aus Perspektive der empirischen Sozialforschung als Teil von Beobachtungen betrachtet werden.

„Unter einer wissenschaftlichen Beobachtung (...) versteht man die zielgerichtete, systematische und regelgeleitete Erfassung, Dokumentation und Interpretation von Merkmalen, Ereignissen oder Verhaltensweisen mithilfe menschlicher Sinnesorgane und/oder technischer Sensoren zum Zeitpunkt ihres Auftretens“ (Döring & Bortz, 2016, S. 324).

Ziel von Beobachtungen ist die Erfassung der Abläufe, aber auch der Bedeutung von Handlungen und Handlungszusammenhängen (vgl. Kromrey, Roose & Strübing, 2016, S. 325). Beobachtungen weisen gegenüber Befragungen eine Reihe von Vorteilen auf, wie z. B. die Möglichkeit, automatisierte und unbewusste Verhaltensweisen zu erfassen und Handlungen im Zeitablauf abbilden zu können (vgl. Döring & Bortz, 2016, S. 325). Im Vergleich zu Befragungen sind Beobachtungen jedoch deutlich zeit- und kostenintensiver und auf Phänomene beschränkt, die durch eine Fremdbeobachtung zugänglich sind, d. h. die Datenermittlung ist auf sinnlich wahrnehmbares Verhalten beschränkt (vgl. Atteslander, 2010, S. 78 f.; Döring & Bortz, 2016, S. 325).

Wissenschaftliche Beobachtungen lassen sich anhand der sechs Merkmale (1) Strukturierungsgrad der Beobachtung, (2) Gegenstand der Beobachtung, (3) Direktheit der Beobachtung, (4) Ort der Beobachtung, (5) Involviertheitsgrad der Beobachterrolle und (6) Transparenz der Beobachtung klassifizieren (vgl. Döring & Bortz, 2016, S. 328 f.). Der Strukturierungsgrad der Beobachtung ist dabei eng mit dem verwendeten Beobachtungsschema bzw. Beobachtungsbogen verbunden (vgl. Döring & Bortz, 2016, S. 328; Schnell, Hill & Esser, 2013, S. 382). So gelingt es, über den Einsatz eines voll- bzw. teilstandardisierten Beobachtungsbogens, die interessierenden Phänomene festzulegen und die Voraussetzungen für die Überprüfbarkeit der Beobachtungsergebnisse herzustellen (vgl. Häder, 2015, S. 312).

Kromrey, Roose und Strübing (2016, S. 326 f.) stellen dementsprechend hohe Anforderungen an einen Beobachtungsbogen, da die Datenerfassung simultan zum beobachteten Geschehen erfolgen muss, in der Beobachtungssituation in der Regel eine Vielzahl von Aktivitäten gleichzeitig ausgeführt werden und verpasste Beobachtungen nicht mehr nachgeholt werden können. Konkret sollte ein Beobachtungsbogen so konstruiert sein, dass

1. die zu erfassenden Phänomene eindeutig und präzise mit Hilfe von Kategorien definiert sind.
2. die Kategorien so definiert werden, dass sie eine leichte Einschätzung in der Beobachtungsstation erlauben.
3. die Kategorien den Handlungszusammenhang beibehalten, d. h. den Handlungssinn nicht zerstören.
4. möglichst wenige Kategorien enthalten sind, damit der Beobachtungsbogen auch praktisch anwendbar bleibt.

Die Umsetzung dieser Anforderungen bei der Konstruktion des Beobachtungsbogens soll zudem sicherstellen, dass die beobachtete Situation auch intersubjektiv nachvollziehbar eingeschätzt und bewertet wird (vgl. Kromrey, Roose & Strübing, 2016, S. 326 f.). Verzerrungen, z. B. aus der Durchführung der Situation oder der Art der Ausfüllung des Beobachtungsbogens, sollen dagegen soweit wie möglich minimiert werden.

Die Sicherstellung einer intersubjektiv nachvollziehbaren Einschätzung ist ein wesentliches Merkmal, damit die gewonnenen Beobachtungsdaten für die Verbesserung individueller Lernprozesse, aber auch für die Bewertung professioneller Leistungen im Sinne der Qualifikationenvergabe (z. B. für Zeugnisse, Noten usw.) genutzt werden können. Daher sollten die mit dem Beobachtungsbogen erzielten Diagnosen die klassischen Gütekriterien der Objektivität, Reliabilität und Validität erfüllen (vgl. Helmke, 2015, S. 124). Die Objektivität beschreibt hierbei, „in welchem Ausmaß die Testergebnisse vom Testanwender unabhängig sind“ (Bortz & Döring, 2009, S. 195). Sie lässt sich weiter untergliedern in (1) die Durchführungsobjektivität, welche die Invarianz der Ergebnisse von der Testdurchführung fordert, (2) die Auswertungsobjektivität, welche die Ermittlung gleicher Testwerte für gleiche Resultate sicherstellen soll, und (3) die Interpretationsobjektivität, die zur Gewährleistung gleicher Beurteilungen bzw. Interpretationen von Testwerten dient (vgl. Bortz & Döring, 2009, S. 195; Bühner, 2011, S. 59 f.). Die Reliabilität beschreibt die Fähigkeit eines Instruments, eine zuverlässige bzw. genaue Messung zu ermöglichen (vgl. Bortz & Döring,

2009, S. 196), während die Validität das Ausmaß beschreibt, in dem ein Instrument das intendierte/zu messende Konstrukt tatsächlich erfasst (vgl. Bortz & Döring, 2009, S. 200; Bühner, 2011, S. 61).

Aus Sicht der pädagogisch-psychologischen Diagnostik ist der Genauigkeitsanspruch jedoch differenzierter zu betrachten. So sollten diese drei Gütekriterien immer dann umfänglich erfüllt sein, wenn die Daten zur Bewertung im Sinne der Vergabe von Qualifikationen dienen, da fehlerhafte Bewertungen mit weitreichenden Konsequenzen für die bewerteten Personen verbunden sind (vgl. z. B. Helmke, 2015, S. 124; Hesse & Latzko, 2011, S. 29). Für die Verbesserung individueller Lernprozesse können jedoch geringere Ansprüche geltend gemacht und Verzerrungen gegebenenfalls auch bewusst in Kauf genommen werden, da hier die schnelle Schaffung von Orientierungswissen im Vordergrund steht und Auswirkungen auf die Lernenden erst bei häufigem Auftreten von Fehlern zum Tragen kommen (vgl. Helmke, 2015, S. 124). So argumentieren beispielsweise Weinert und Schradler (1986, S. 18 f.), dass die Qualität didaktischer Maßnahmen nicht durch eine sehr hohe Genauigkeit gleichermaßen ansteigt. Sie fordern vielmehr, die ggf. weniger genauen Diagnosen ständig zu überprüfen, sensibel auf Veränderungen im Handeln der Lernenden zu reagieren, die individuellen Fähigkeiten der Lernenden zum Ausgangspunkt zu nehmen (individuelle Bezugsnorm) und dass sie sich durch pädagogisch sinnvolle Verzerrungen auszeichnen dürfen (z. B. leichte Überschätzung der Leistungsfähigkeit, um einen leistungsförderlichen Unterricht zu gestalten). Solche pädagogisch sinnvollen Verzerrungen stellen zwar aus Sicht der Messtheorie *Fehler* dar, allerdings haben sie vorteilhafte Konsequenzen für die Lernenden.

Aus diesen Überlegungen ergeben sich die folgenden Fragestellungen:

1. Kann der Einsatz von Beobachtungsbögen in der Lehrerbildung zur Unterstützung der professionellen Entwicklung als objektiv betrachtet werden?
2. Wie zuverlässig sind die Diagnosen mittels Beobachtungsbögen?
3. Trägt der Einsatz der Beobachtungsbögen zur professionellen Entwicklung bei?
4. Lassen sich bestimmte pädagogisch sinnvolle Verzerrungen beim Einsatz von Beobachtungsbögen feststellen?
5. Eignen sich Beobachtungsbögen zur Vergabe von Qualifikationen oder eher zur Verbesserung individueller Lernprozesse von Studierenden?

Das zur Beantwortung dieser Fragestellungen genutzte Forschungsdesign wird im folgenden Abschnitt detailliert dargestellt.

### 3 Methodik und Stichprobe

Zur Klärung der Fragestellungen wird auf einen Beobachtungsbogen zurückgegriffen, der an der Carl von Ossietzky Universität Oldenburg im Rahmen des Studiums für das Lehramt an berufsbildenden Schulen zwischen dem Wintersemester 2014/2015 und dem Wintersemester 2018/19 eingesetzt wurde (vgl. Abbildung 1 und Abbildung 2).

<b>Seminargestaltung: Beobachtungskriterien<sup>1</sup></b>					
Studierende/r		_____			
Thema der Sitzungsgestaltung		_____			
Zielorientierung und Strukturierung der Seminarsitzung					
	++	+	-	--	<b>Bemerkung</b>
Roter Faden ist erkennbar.					
Lernzielverfolgung ist erkennbar.					
Transparenz über das Seminargeschehen wird hergestellt.					
Zeitplanung wird eingehalten.					
Gestaltung der Lernprozesse					
	++	+	-	--	<b>Bemerkung</b>
Methoden sind geeignet ( <i>Inhalt, Ziel, Klientel</i> ).					
Methoden werden sicher eingesetzt.					
Medien- und Materialeinsatz ist sinnvoll ( <i>Inhalt, Ziel, Klientel</i> ).					
Medien- und Materialeinsatz gelingt sicher.					
Medien- und Materialeinsatz sind von guter Qualität.					
Fachlicher Gehalt der Seminarsitzung					
	++	+	-	--	<b>Bemerkung</b>
Fachlicher Inhalt ist korrekt.					
Lehrkraft antwortet auf Fragen angemessen.					
Lehrkraft verdeutlicht Praxisrelevanz.					
Lehrkraft sorgt für Klarheit und Angemessenheit der fachlichen Aufgaben.					
Lehrkraft verdeutlicht den Bezug zum Makrokontext.					
„Lehrerverhalten“ der Seminarverantwortlichen					
	++	+	-	--	<b>Bemerkung</b>
Lehrkraft achtet auf Kooperation und Respekt.					
Lehrkraft achtet auf Beteiligungschancen und koordiniert Studierendenbeiträge.					
Lehrkraft wählt eine angemessene Fragetechnik ( <i>zielorientiert, offen, aktivierend</i> ).					
Lehrkraft zeigt angemessene Reaktion auf Studierendenverhalten.					
<b>Weitere Anmerkungen:</b>					
<i>Aufreten vor der Gruppe:</i> z. B. Haltung, Ausstrahlung, Gestik, Mimik/Bewegungsverhalten im Seminarraum					
<i>Sprache:</i> z. B. Sprachniveau, Lautstärke, Deutlichkeit, Modulation, Tempo/Sprachumfang (Sprechanteil, Lehrerecho)					

Abb. 1: Eingesetzter Beobachtungsbogen der Studie (Seite 1)

<b>Zielvereinbarungsprozess</b>
Umsetzung der Zielvereinbarung(en) der letzten Sitzungsgestaltung, Zielvereinbarung für die nächste Sitzungsgestaltung
<p><b>Mögliche Leitfragen zur Reflexion</b></p> <p>(1) Wie habe ich mich in meiner Rolle als Lehrkraft gefühlt?</p> <p>(2) Welche Dinge/Aspekte sind mir gut gelungen? Womit bin ich richtig zufrieden?</p> <p>(3) Was würde ich so wieder machen? Warum?</p> <p>(4) Welche Dinge/Aspekte würde ich ändern? Warum?</p> <p>(5) Wo gab es Abweichungen von meiner ursprünglichen Planung? Warum sind diese aufgetreten und wie bin ich damit umgegangen? Wie habe ich mich dabei gefühlt?</p> <p>(6) Welche Bereiche des Planens und Unterrichtens fallen mir noch schwer? Was könnte ich tun, um mich diesbezüglich sicherer zu fühlen? Wo und wie kann ich mir ggf. Hilfe holen?</p> <p>(7) Welche Ziele setze ich mir für meine nächste Sitzungsgestaltung? An was möchte ich bis dahin arbeiten? Was möchte ich dort anders/besser machen? Etc.</p>

Abb. 2: Eingesetzter Beobachtungsbogen der Studie (Seite 2)

Der Beobachtungsbogen wurde mit einer Ausbildungsinstitution der zweiten Phase der Lehrerbildung abgestimmt und erfasst Merkmale, die auch in der zweiten Phase der Lehrerbildung beobachtet werden. Die konkreten Beobachtungskategorien sind dabei keiner bestimmten Theorien entlehnt, sondern stellen Kategorien dar, die in der zweiten Phase der Lehrerbildung in dieser oder ähnlicher Form zur Anwendung kommen. Hinter dieser Übernahme der Kategorien aus der Praxis stand die Überlegung, eine bessere Verzahnung der beiden Phasen der Lehrerbildung zu erzielen und Impulse aus der Praxis für die Theoriebildung nutzbar zu machen. Im Studium wurde das Instrument konkret für

Rückmeldungen zu Seminargestaltungen von Studierenden genutzt. Hierbei haben Studierende in Einzelarbeit oder Gruppen von maximal vier Personen eine *Unterrichtsstunde* für bis zu 90 Minuten konzipiert und durchgeführt. Während der Vorbereitung sprachen sie Inhalte, Verlauf, Materialien und einzusetzende Medien mit dem jeweiligen Lehrenden des Moduls ab. Gleichzeitig wurden die Studierenden während der Seminargestaltung von der Dozentin bzw. dem Dozenten beobachtet und für jeden Studierenden ein individueller Beobachtungsbogen ausgefüllt, auf dessen Basis im Anschluss eine Rückmeldung in Form eines Feedbackgesprächs erfolgte. Die Handhabung des Instruments wurde ganz zu Beginn unter den Lehrenden abgesprochen.

Aus diesem Vorgehen liegen insgesamt die Datensätze von 241 Studentinnen und 149 Studenten vor, die sich auf 840 Beobachtungen von drei Dozenten und sechs Dozentinnen aus vier Modulen (zwei Master- und zwei Bachelormodule) verteilen. Auf das Modul BIW111 (Strukturelle Aspekte sowie Personal- und Unterrichtsentwicklung in berufs- und wirtschaftspädagogischen Handlungsfeldern) entfallen 278, auf das Modul PB027/029 (Ausgewählte Probleme in Berufs- und wirtschaftspädagogischen Handlungsfeldern) 312, auf das Modul WIR181/182 (Ausgewählte Probleme in wirtschaftsdidaktischen Handlungsfeldern) 233 und auf das Modul WIR731 (Gestaltung wirtschaftsdidaktischer Lernsituationen) 17 Beobachtungen. Dozent 1 stellt insgesamt 86, Dozent 2 35, Dozent 3 51, Dozentin 1 131, Dozentin 2 297, Dozentin 3 63, Dozentin 4 32, Dozentin 5 122 und Dozentin 6 23 Beobachtungen. Mit Ausnahme des Moduls WIR181/182 verlaufen alle Module über zwei Semester. Die Daten werden in der nachfolgenden Tabelle zusammengefasst:

Tab. 1: Verteilung der Beobachtungseinheiten der Studierenden auf Lehrende und Module

	BIW111	PB027/029	WIR181/182	wir731	Gesamt
Dozent 1	0	65	21	0	86
Dozent 2	35	0	0	0	35
Dozent 3	32	19	0	0	51
Dozentin 1	0	46	68	17	131
Dozentin 2	155	53	89	0	297
Dozentin 3	3	60	0	0	63
Dozentin 4	0	0	32	0	32
Dozentin 5	53	69	0	0	122
Dozentin 6	0	0	23	0	23
Gesamt	278	312	233	17	840

Zur Abbildung der Nutzung der Items wird eine Skala von 0 für -- bis 3 für ++ verwendet. Für den Fall, dass eine Markierung zwischen zwei Kategorien lag, wurde der Mittelwert gebildet. Die handschriftlichen Notizen sind nicht Gegenstand dieser Analyse.

## 4 Analysen

In einem ersten Schritt wird mittels der konfirmatorischen Faktorenanalyse (CFA) geprüft, ob sich die Beobachtungstems zu den vier Kategorien des Beobachtungsbogens verdichten lassen. Grundlage hierfür sind die vorliegenden 840 Beobachtungen. Gleichzeitig wird geprüft, ob sich die vier Kategorien weiter zu einem Generalfaktor verdichten lassen, der die Gesamtqualität der Seminargestaltungen charakterisiert. Das dazugehörige Modell ist in Abbildung 3 dargestellt.

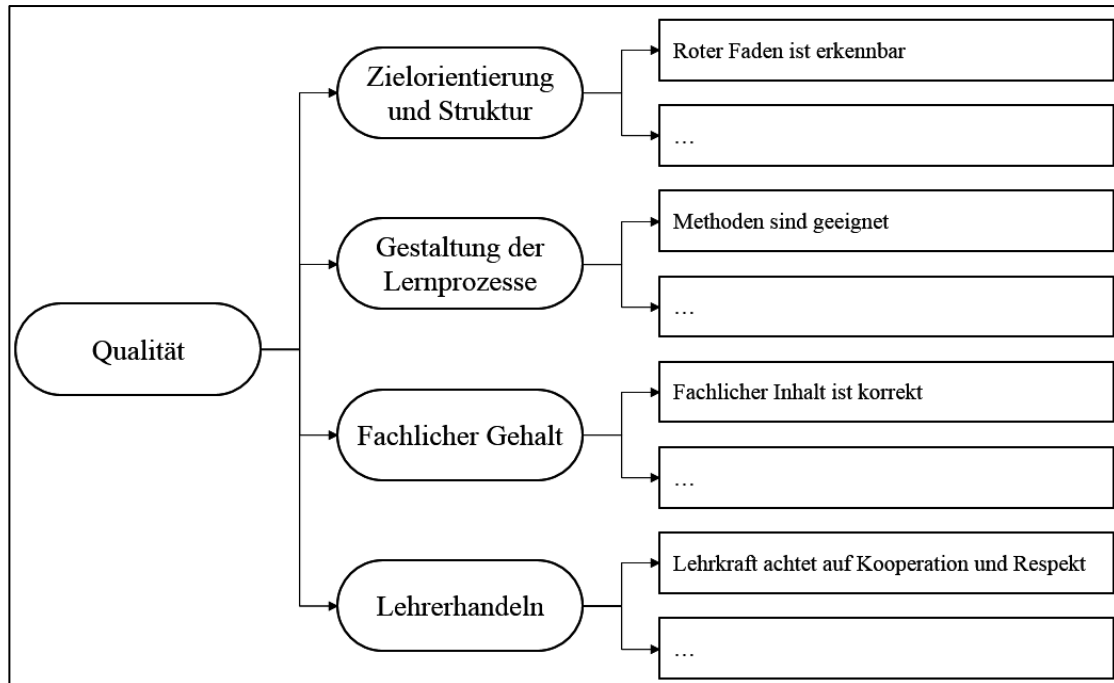


Abb. 3: Modell der Seminargestaltungen

Zunächst wird das Ergebnis einer CFA zweiter Ordnung mit dem ML-Schätzer dargestellt. Dies hat den Vorteil, dass MPlus fehlende Werte mit Hilfe der Full-Information-Maximum-Likelihood-Schätzung ergänzen kann. Solche fehlenden Werte resultieren im vorliegenden Datensatz z. B. daraus, dass bei einem Beobachtungstems keine Kategorie angekreuzt wurde (vgl. dazu Tabelle 2). Die Rechnung kommt für die globale Passung des in Abbildung 1 dargestellten Modells zu den folgenden Werten:  $N = 806$ ,  $\chi^2(131) = 566.624$ ,  $p < .01$ ; RMSEA: .064; CI90 [.059; .070]; CFI: .924; SRMR: .040). Gemessen an der Kombinationsregel von Hu und Bentler (1999, S. 27 f.) zeigt ein RMSEA unter .06 und ein SRMR unter .09 eine globale Modellpassung an. Die Modellpassung wird zudem untermauert, wenn der CFI einen Wert von mindestens .950 annimmt. Im vorliegenden Fall verfehlt der RMSEA-Wert knapp diese Grenze, während der SRMR-Wert erfüllt ist. Allerdings liegt der benötigte Wert für RMSEA im 90 Prozent Konfidenzintervall, sodass ggf. noch von einer Modellpassung ausgegangen werden kann.

Um dieses Ergebnis weiter abzusichern, wird zusätzlich eine Rechnung mit Hilfe des MLM-Schätzers in MPlus vorgenommen. Dieser Schätzer hat den Vorteil, dass er Abweichungen von der Normalverteilung berücksichtigt und so korrigierte Kennwerte zur Verfü-

gung stellt. Allerdings kann diese Schätzung nur für komplett vollständige Datensätze genutzt werden, sodass sich der Stichprobenumfang, welcher der Schätzung zu Grunde liegt, reduziert. Diese Berechnung ergibt die folgende Werte:  $N = 407$ ; korrigiertes  $\chi^2(131) = 295.908$ ;  $p < .01$ ; RMSEA: .056; CI90 [.047; .064]; CFI: .946; SRMR: .041. Diese Analyse erfüllt die Grenzwerte von Hu und Bentler (1999, S. 27 f.) für den RMSEA und SRMR und auch der CFI-Wert ist nahezu erfüllt. Vor dem Hintergrund dieser Ergebnisse wird für die weiteren Analysen von einer Passung des Modells ausgegangen, die jedoch in weiteren Studien zu prüfen ist. Zentrale Kennwerte der lokalen Modellpassung zeigt Tabelle 2.

Tab. 2: Standardisierte Faktorladungen, Mittelwerte und Standardabweichungen sowie Cronbachs  $\alpha$

Item	SF	M	SD	FW
<b>Zielorientierung und Struktur (<math>\alpha = .690</math>)</b>				
Roter Faden ist erkennbar.	.628	2.505	.622	4.5
Lernzielverfolgung ist erkennbar.	.708	2.301	.650	8.0
Transparenz über das Semingergeschehen wird hergestellt.	.634	2.482	.600	8.9
Zeitplanung wird eingehalten.	.434	1.962	.830	8.2
<b>Gestaltung der Lernprozesse (<math>\alpha = .825</math>)</b>				
Methoden sind geeignet ( <i>Inhalt, Ziel, Klientel</i> ).	.765	2.621	.561	6.2
Methoden werden sicher eingesetzt.	.778	2.321	.638	6.3
Medien- und Materialeinsatz ist sinnvoll ( <i>Inhalt, Ziel, Klientel</i> ).	.779	2.518	.586	6.2
Medien- und Materialeinsatz gelingt sicher.	.786	2.503	.626	6.7
Medien- und Materialeinsatz sind von guter Qualität.	.663	2.311	.644	6.7
<b>Fachlicher Gehalt (<math>\alpha = .846</math>)</b>				
Fachlicher Inhalt ist korrekt.	.692	2.460	.594	8.8
Lehrkraft antwortet auf Fragen angemessen.	.772	2.403	.638	25.7
Lehrkraft verdeutlicht Praxisrelevanz.	.730	2.300	.678	11.8
Lehrkraft sorgt für Klarheit und Angemessenheit der fachlichen Aufgaben.	.669	2.252	.626	12.4
Lehrkraft verdeutlicht den Bezug zum Makrokontext.	.713	2.221	.733	17.7
<b>Lehrerhandeln (<math>\alpha = .816</math>)</b>				
Lehrkraft achtet auf Kooperation und Respekt.	.772	2.517	.600	23.8
Lehrkraft achtet auf Beteiligungschancen und koordiniert Studierendenbeiträge.	.743	2.393	.646	16.3
Lehrkraft wählt eine angemessene Fragetechnik	.738	2.240	.679	21.2
Lehrkraft zeigt angemessene Reaktion auf Studierendenverhalten.	.743	2.377	.650	16.7
<b>Generalfaktor Qualität (<math>\alpha = .860</math>)</b>				
Zielorientierung und Struktur	.733	2.325	.491	
Gestaltung der Lernprozesse	.934	2.448	.470	
Fachlicher Gehalt	.987	2.331	.514	
Lehrerhandeln	.945	2.404	.549	

Hinweise: SF = Standardisierte Faktorladung, FW = Anzahl fehlender Werte in %.

Alle Faktorladungen sind auf dem 1%-Niveau signifikant.

Wie Tabelle 2 verdeutlicht, sind alle Faktorladungen auf dem 1 %-Niveau signifikant und besitzen ein positives Vorzeichen. Damit liegt insgesamt eine lokale Passung des Modells aus Abbildung 3 vor, sodass auch diese Analyse die obige Annahme einer Modellpassung bestätigt.

Tabelle 2 berichtet darüber hinaus auch die Werte für Cronbachs  $\alpha$ , die einen Wert von ca. .700 übersteigen sollten, um eine zuverlässige Messung anzuzeigen (vgl. Adams & Lawrence, 2015). Diesen Grenzwert erfüllen alle Skalen mit Ausnahme der *Zielorientierung und Struktur*, welche den Grenzwert knapp verfehlt. Um weiterhin zu prüfen, ob die Skalen unterschiedliche Konstrukte erfassen, wird das Fornell-Larcker-Kriterium herangezogen. Es fordert, dass die quadrierte Korrelation zwischen zwei Konstrukten geringer ist als die durchschnittlich erfasste Varianz der Konstrukte (vgl. Fornell & Larcker, 1981). Die notwendigen Informationen weist Tabelle 3 aus.

Tab. 3: Korrelation, quadrierte Korrelation und durchschnittlich erfasste Varianz

	Zielorientierung und Struktur	Gestaltung der Lernprozesse	Fachlicher Gehalt	Lehrerhandeln
Zielorientierung und Struktur	.373	.476	.593	.368
Gestaltung der Lernprozesse	.690	.571	.823	.810
Fachlicher Gehalt	.770	.907	.513	.870
Lehrerhandeln	.607	.900	.933	.561

*Hinweise:* Werte auf der Diagonalen sind die durchschnittlich erfasste Varianz, Werte unterhalb der Diagonalen sind die Korrelationen und Werte oberhalb der Diagonalen sind die quadrierten Korrelationen. Alle Korrelationen sind auf dem 1 %-Niveau signifikant.

Die durchschnittlich erfasste Varianz sollte für eine gute Messung mindestens .500 betragen (vgl. Bagozzi & Yi, 1988, S. 80). Diesen Grenzwert erfüllen alle Skalen mit Ausnahme der *Zielorientierung und Struktur*. Das Fornell-Larcker-Kriterium ist hingegen nur im Fall der Konstruktpaare *Lehrerhandeln* und *Zielorientierung und Struktur* erfüllt. Dies bedeutet, dass in allen anderen Fällen wesentliche Überschneidungen zwischen den Konstrukten vorliegen. Dies zeigt sich beispielsweise auch an den hohen Korrelationen, die in drei Fällen Werte von mindestens .900 aufweisen. Damit ist insgesamt die diskriminante Validität der Skalen als fraglich zu bewerten.

Auffallend sind zudem die Unterschiede in den fehlenden Werten. Während die Lehrenden nur in wenigen Fällen keine Kategorie des Fragebogens für die Faktoren *Zielorientierung und Struktur* sowie *Gestaltung der Lernprozesse* nutzen, weisen vor allem die Faktoren *Fachlicher Gehalt* und *Lehrerhandeln* hohe fehlende Werte auf, deren Anteil überwiegend zwischen 11,8 % und 25,7 % schwankt.

Um tiefergehende Einblicke in die Nutzung des Beobachtungsbogens zu erhalten, stellt Tabelle 4 die Ergebnisse multivariater Varianzanalysen auf der Grundlage der einzelnen Beobachtungen dar.

Tab. 4: Zusammengefasste Ergebnisse der multiplen Varianzanalysen

	Zielorientierung und Struktur	Gestaltung der Lernprozesse	Fachlicher Gehalt	Lehrerhandeln	Generalfaktor Qualität
Geschlecht	F(1;765) = 1.975 $\eta^2 = .00$	F(1;758) = .000 $\eta^2 = .00$	F(1;754) = .074 $\eta^2 = .00$	F(1;727) = .437 $\eta^2 = .00$	F(1;766) = .460 $\eta^2 = .00$
Modul	F(5;765) = 3.356** $\eta^2 = .02$	F(5;758) = 2.632* $\eta^2 = .01$	F(5;754) = 2.775* $\eta^2 = .01$	F(5;727) = 1.901 $\eta^2 = .01$	F(5;766) = 3.677** $\eta^2 = .02$
Dozent/-in	F(8;765) = 12.878** $\eta^2 = .11$	F(8;758) = 15.593** $\eta^2 = .13$	F(8;754) = 14.047** $\eta^2 = .12$	F(8;727) = 11.150** $\eta^2 = .10$	F(8;766) = 14.314* * $\eta^2 = .12$
Geschlecht* Modul	F(5;765) = .900 $\eta^2 = .00$	F(5;758) = .944 $\eta^2 = .01$	F(5;754) = .952 $\eta^2 = .01$	F(5;727) = 1.264 $\eta^2 = .01$	F(5;766) = .978 $\eta^2 = .01$
Geschlecht* Dozent/-in	F(8;765) = .744 $\eta^2 = .01$	F(8;758) = 1.387 $\eta^2 = .01$	F(8;754) = 2.094* $\eta^2 = .02$	F(8;727) = 2.532* $\eta^2 = .02$	F(8;766) = 2.358* $\eta^2 = .02$
Modul* Dozent/-in	F(6;765) = 2.644* $\eta^2 = .02$	F(6;758) = 4.781** $\eta^2 = .03$	F(6;754) = 5.466** $\eta^2 = .03$	F(6;727) = 3.474** $\eta^2 = .02$	F(6;766) = 4.884** $\eta^2 = .03$
Geschlecht* Modul* Dozent/-in	F(6;765) = 3.601** $\eta^2 = .02$	F(6;758) = .369 $\eta^2 = .00$	F(6;754) = .754 $\eta^2 = .00$	F(6;727) = .462 $\eta^2 = .00$	F(6;766) = 1.201 $\eta^2 = .01$

Hinweise:  
\*\*  $p < .01$ ; \*  $p < .05$

Wie Tabelle 4 zeigt, hat der Hauptfaktor *Geschlecht* der Studierenden keinen Einfluss auf die Bewertung der einzelnen Faktoren des Beobachtungsbogens. Das gleiche gilt auch für die Interaktion aus *Modul* und *Geschlecht*. Im Gegensatz dazu lassen sich Interaktionen zwischen dem *Geschlecht* der Studierenden und dem *Dozenten bzw. der Dozentin* für die Bewertungen beim fachlichen Gehalt, dem Lehrerhandeln und dem Generalfaktor *Qualität* feststellen, die jeweils ca. 2 % der Gesamtvarianz aufklären. Gemäß den Richtlinien von Cohen (1988, S. 283 ff.) liegt ab einem  $\eta^2$  von 1 % ein schwacher, ab 6 % ein mittlerer und ab 14 % ein starker Effekt vor. Demnach ist dieser Effekt als eher schwach zu klassifizieren.

In eine ähnliche Richtung deutet auch der Haupteffekt *Modul*, der mit Ausnahme des Lehrerhandels signifikant wird, aber dessen Wert für  $\eta^2$  2 % nicht übersteigt. Erklärungsmächtiger ist hingegen die Interaktion aus *Modul* und *Dozentin bzw. Dozent*, die zwischen 2 % und 3 % der Gesamtvarianz für alle Kategorien erklären kann.

Den wichtigsten Einflussfaktor stellt jedoch der Haupteffekt aus *Dozentin bzw. Dozent* dar, der in jedem Fall signifikant wird und eine Varianzaufklärung für alle Faktoren des Beobachtungsbogens zwischen 10 % und 13 % erlaubt. Gemessen nach Cohen (1988, S. 283 ff.) ist dieser Effekt als mittel bis stark zu klassifizieren. Um einen besseren Eindruck von diesem Effekt zu erhalten, stellt die folgende Abbildung 4 die Randmittelwerte der Lehrenden dar.

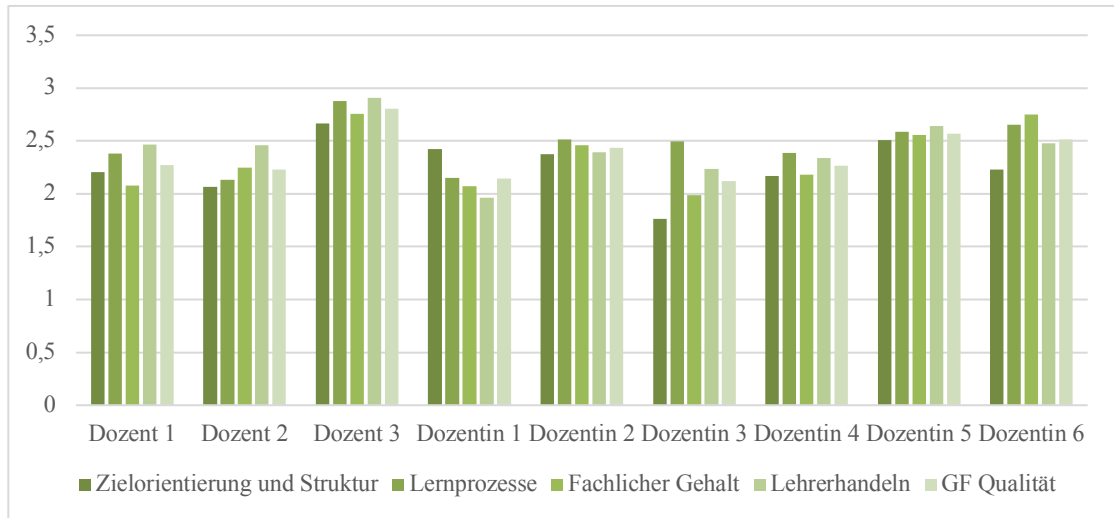


Abb. 4: Randmittelwerte für den Hauptfaktor Dozent/-in

Wie Abbildung 4 zeigt, lassen sich in den Mittelwerten Verwendungstendenzen der Beobachtungssitems erkennen. So weist Dozent 3 im Vergleich zu den anderen Lehrenden auffallend positive Bewertungen in allen Faktoren auf. Im Gegensatz dazu sind die Dozentinnen 1 und 3 eher streng in ihren Bewertungen. Um die Unterschiede in der Handhabung der Lehrenden noch deutlicher zu machen, zeigt Abbildung 5 die z-Werte für die Beobachtungen. z-Werte sind so gestaltet, dass der Mittelwert der Stichprobe 0 entspricht und die Standardabweichung stets 1 beträgt. Somit lässt sich ein Wert im Bereich zwischen -1 und +1 als durchschnittlich interpretieren (vgl. Bühner & Ziegler, 2017, S. 77 f.).

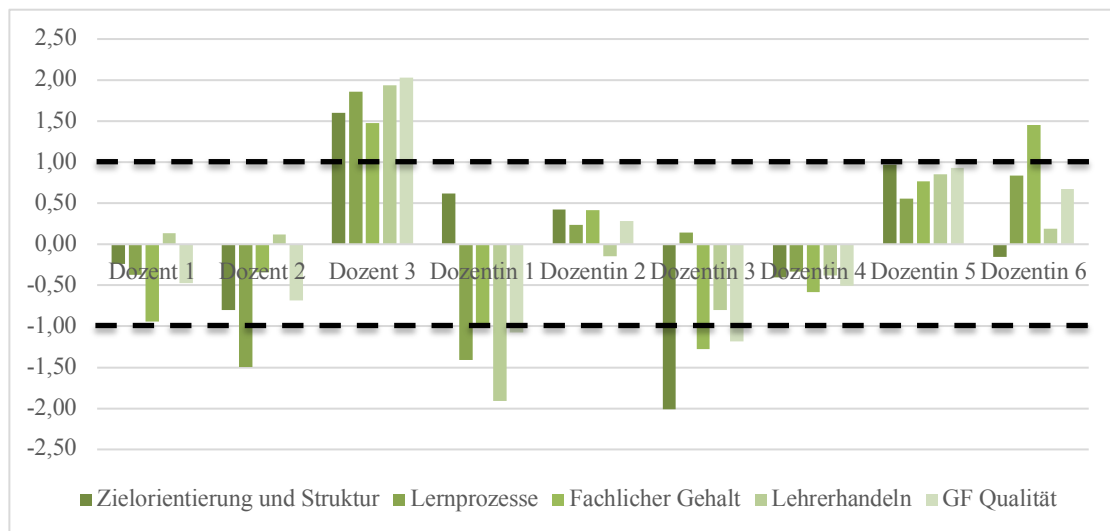


Abb. 5: z-Werte für den Hauptfaktor Dozent/-in

Demnach bewertet Dozent 3 die Studierenden mit Hilfe des Beobachtungsbogens überdurchschnittlich gut im Vergleich zu den anderen Lehrenden. Das gleiche gilt für Dozentin 6 bei dem Kriterium *fachliches Gehalt*. Dozentin 1 vergibt auf dem Beobachtungsbogen hingegen auf fast allen Faktoren eher unterdurchschnittliche Werte, insbesondere für die Kriterien *Zielorientierung und Struktur* sowie *Fachlicher Gehalt*. Ähnlich sind auch die Be-

wertungen von Dozent 2 und Dozentin 3 einzuordnen. Dozent 1 sowie Dozentin 2 und Dozentin 4 vergeben hingegen eher durchschnittliche Bewertungen mit Hilfe des Beobachtungsbogens.

Um darüber hinaus Informationen über die Entwicklung der Studierenden zu erhalten stellt Tabelle 5 die zentralen Ergebnisse dreifaktorieller Varianzanalysen mit Messwiederholung getrennt für die Module dar.

Tab. 5: Zusammengefasste Ergebnisse der multiplen Varianzanalysen mit Messwiederholung für die Entwicklung des Generalfaktors *Qualität*

Modul	PB	WIR	BIW
Zeit	F (1, 52) = 4.844* $\eta^2 = .02$	F (1, 38) = .032 $\eta^2 = .00$	F (1, 44) = 6.637* $\eta^2 = .01$
Zeit*Geschlecht	F (1, 52) = .000 $\eta^2 = .00$	F (1, 38) = .942 $\eta^2 = .01$	F (1, 44) = 1.568 $\eta^2 = .00$
Zeit*Dozent/-in (t0)	F (4, 52) = 3.776** $\eta^2 = .06$	F (2, 38) = .666 $\eta^2 = .01$	F (3, 44) = 9.542** $\eta^2 = .04$
Zeit*Dozent/-in (t1)	F (5, 52) = 10.499** $\eta^2 = .21$	F (1, 38) = 1.610 $\eta^2 = .02$	F (4, 44) = 11.279** $\eta^2 = .07$
Zeit*Geschlecht*Dto	F (3, 52) = .576 $\eta^2 = .01$	F (1, 38) = .515 $\eta^2 = .01$	F (1, 44) = .544 $\eta^2 = .00$
Zeit*Geschlecht*Dt1	F (3, 52) = 1.019 $\eta^2 = .01$	F (1, 38) = .023 $\eta^2 = .00$	F (1, 44) = 1.629 $\eta^2 = .00$
Zeit*Dto*Dt1	F (4, 52) = 1.666 $\eta^2 = .03$	-/-	-/-
Zeit*Geschlecht*Dto*Dt1	F (1, 52) = .059 $\eta^2 = .00$	-/-	-/-
Geschlecht	F (1, 52) = 2.238 $\eta^2 = .01$	F (1, 38) = 1.347 $\eta^2 = .02$	F (1, 44) = .022 $\eta^2 = .00$
Dozent/-in (t0)	F (4, 52) = 2.237 $\eta^2 = .05$	F (1, 38) = .499 $\eta^2 = .01$	F (3, 44) = 1.145 $\eta^2 = .05$
Dozent/-in (t1)	F (5, 52) = 2.912** $\eta^2 = .07$	F (1, 38) = 2.645 $\eta^2 = .03$	F (4, 44) = 1.095 $\eta^2 = .07$
Geschlecht*Dozent/-in (t0)	F (3, 52) = .030 $\eta^2 = .00$	F (1, 38) = 2.325 $\eta^2 = .03$	F (1, 44) = .232 $\eta^2 = .00$
Geschlecht*Dozent/-in (t1)	F (3, 52) = .303 $\eta^2 = .00$	F (1, 38) = .731 $\eta^2 = .01$	F (1, 44) = .042 $\eta^2 = .00$
Dozent/-in (t0)* Dozent/-in (t1)	F (4, 52) = 2.062 $\eta^2 = .04$	-/-	-/-
Geschlecht* Dozent/-in (t0)* Dozent/-in (t1)	F (1, 52) = .055 $\eta^2 = .00$	-/-	-/-

Hinweise:

\*\*  $p < .01$ ; \*  $p < .05$

Wie Tabelle 5 zeigt, wird in den Modulen PB und BIW der Hauptfaktor *Zeit* signifikant. Dies bedeutet, dass über die zwei Semester eine Entwicklung stattfindet, die zwischen 1 % und 2 % der gesamten Varianz erklären kann. Eine Inspektion der Randmittelwerte zeigt, dass die Werte für den Generalfaktor für BIW von 2.259 auf 2.433 ansteigen und für PB von 2.485 auf 2.271 sinken. Ein Unterschied zwischen den beiden Messzeitpunkten im Seminar WIR konnten hingegen nicht ermittelt werden. Damit liegt insgesamt ein höchstens schwacher Effekt dieses Hauptfaktors vor.

Darüber hinaus wird für die Module PB und BIW die Interaktion aus Zeit und Dozent/-in signifikant. Diese Interaktion erklärt zum Zeitpunkt  $t_0$  zwischen 4 % und 6 % der Varianz und für den Zeitpunkt  $t_1$  zwischen 7 % und 21 % der Varianz. Nach Cohen (1988, S. 283 ff.) sind diese Effekte zu  $t_0$  als schwach bis mittel und für  $t_1$  als mittel bis stark einzuordnen. Eine detaillierte Analyse der Interaktion ist mit Abbildung 6 möglich.

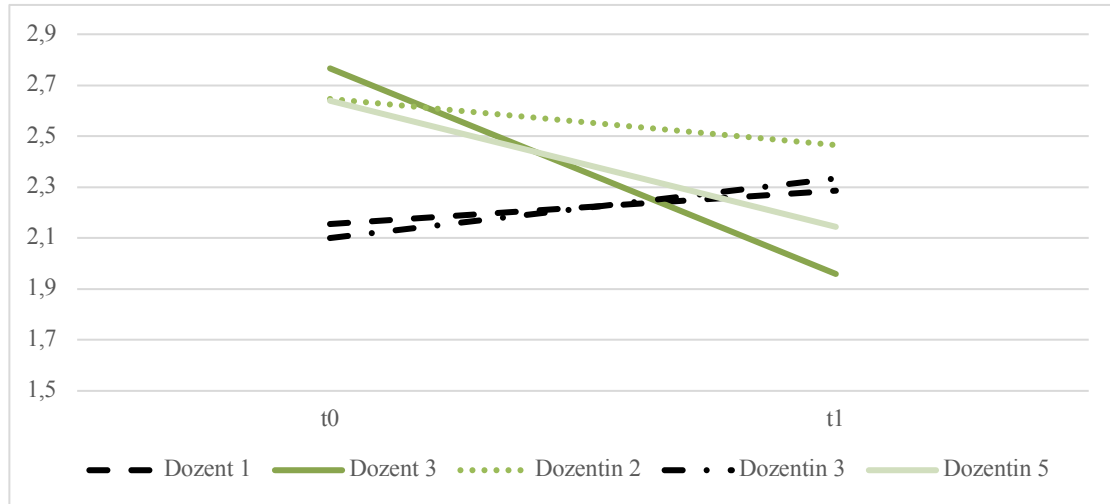


Abb. 6: Randmittelwerte für die Interaktion aus der Entwicklung und den Lehrenden zum Zeitpunkt  $t_0$  im Modul PB

Wie Abbildung 6 zeigt, erzielten Studierende bei Dozentin 5 und Dozent 3 zum Zeitpunkt  $t_0$  relativ hohe Werte auf dem Generalfaktor Qualität. Zum Zeitpunkt  $t_1$  erzielten die Studierenden dann jedoch deutlich geringere Werte. Im Gegensatz dazu erzielten die Studierenden bei Dozent 1 und Dozentin 3 zum ersten Messzeitpunkt relativ niedrige und zum Zeitpunkt  $t_1$  relativ hohe Werte. Wird die Abweichung von der Einschätzung aus Abbildung 5 genutzt, wonach Dozentin 2 und 5 sowie Dozent 3 zu eher milden Bewertungen tendieren und Dozentin 3 und Dozent 1 zu eher strengen Einschätzungen gelangen, so zeigt sich ein eindeutiger Zusammenhang. Die strengen Lehrenden sind in Abbildung 6 in schwarz, die milden Lehrenden in grün dargestellt. Es wird deutlich, dass die Studierenden bei den milden Lehrenden eine Verringerung der Einschätzungen erzielen, während die Studierenden mit strengen Lehrenden eine Verbesserung der Einschätzungen erreichen. Damit nehmen die Werte der Studierenden bei Lehrenden, die eher überdurchschnittlich gute Bewertungen abgegeben, mit der Zeit ab, während Studierende bei eher strengen Lehrenden mit der Zeit bessere Werte erzielen. Weitere Informationen hierzu zeigt Abbildung 7 für die Interaktion aus Entwicklung und Dozent/-in zum Zeitpunkt  $t_1$ .

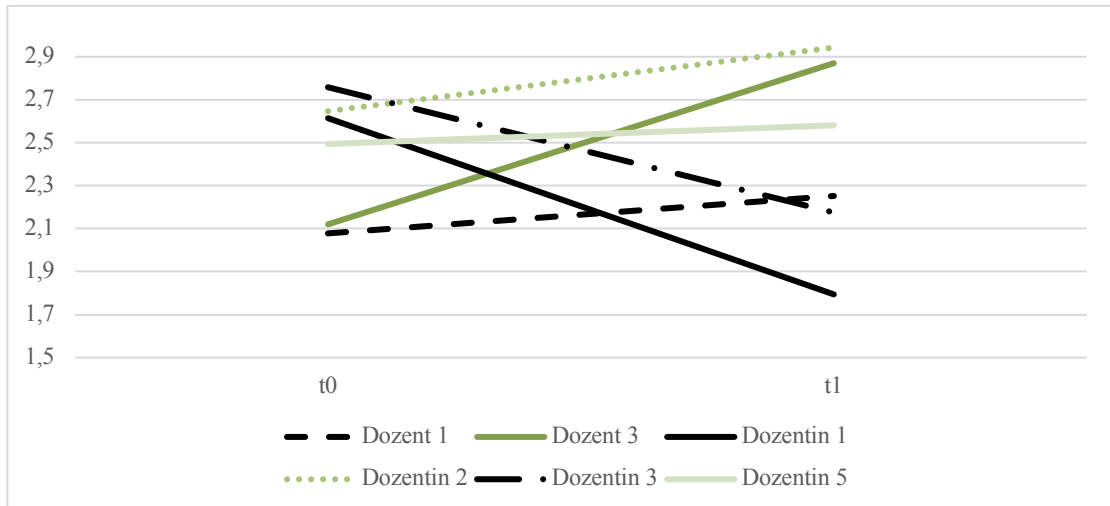


Abb. 7: Randmittelwerte für die Interaktion aus der Entwicklung und den Lehrenden zum Zeitpunkt  $t_1$  im Modul PB

Wie Abbildung 7 zeigt, erzielen die Lernenden im zweiten Messzeitpunkt bessere Werte, wenn die Lehrenden zum zweiten Zeitpunkt Dozentin 2 und Dozent 3 sind. Die Werte verschlechtern sich für die Lehrenden Dozentin 3 und Dozentin 1. Eher keine Entwicklung ist bei Dozentin 5 und Dozent 1 festzuhalten. Damit hängt die Entwicklung der Seminarqualität, die über den Beobachtungsbogen gemessen werden soll, stark von den Lehrenden ab. Dies deckt sich mit der vorhergehenden Analyse, dass Dozentin 3 und Dozentin 1 eher streng bewerten, während Dozent 3 im Durchschnitt eher hoch bewertet.

Die Interaktion aus Zeit und Lehrenden für das Modul BIW wird in den folgenden Abbildungen für eine detaillierte Analyse dargestellt.

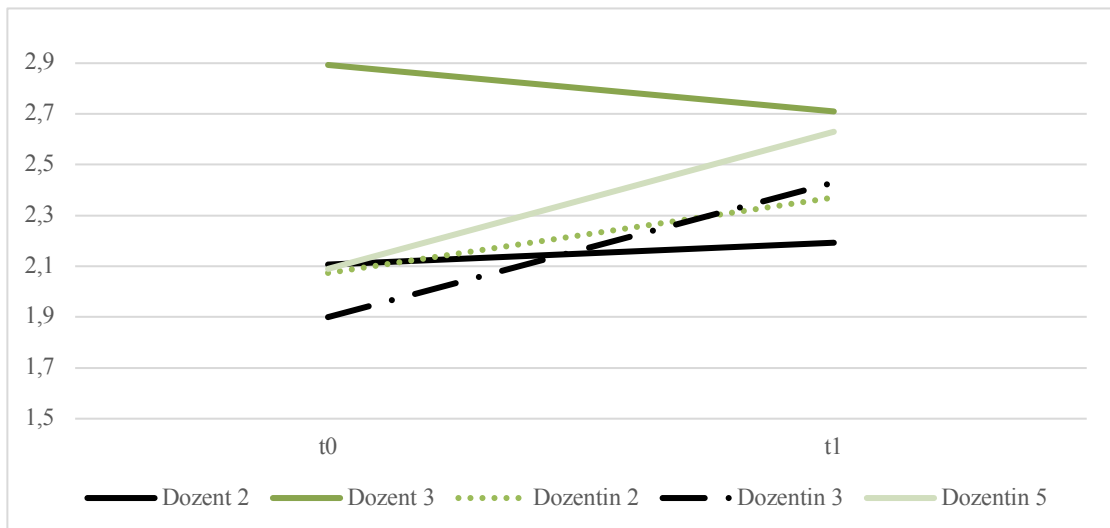


Abb. 8: Randmittelwerte für die Interaktion aus der Entwicklung und den Lehrenden zum Zeitpunkt  $t_0$  im Modul BIW

Wie Abbildung 8 zeigt, lässt sich mit Ausnahme von Dozent 3 eine positive Entwicklung festhalten, wobei die Zuwachsrates für Dozent 2 relativ gering ist. Damit besteht der Interaktionseffekt vor allem darin, dass die verschiedenen Lehrenden unterschiedlich stark zur Verbesserung der Werte für die Qualität der Seminargestaltung beitragen.

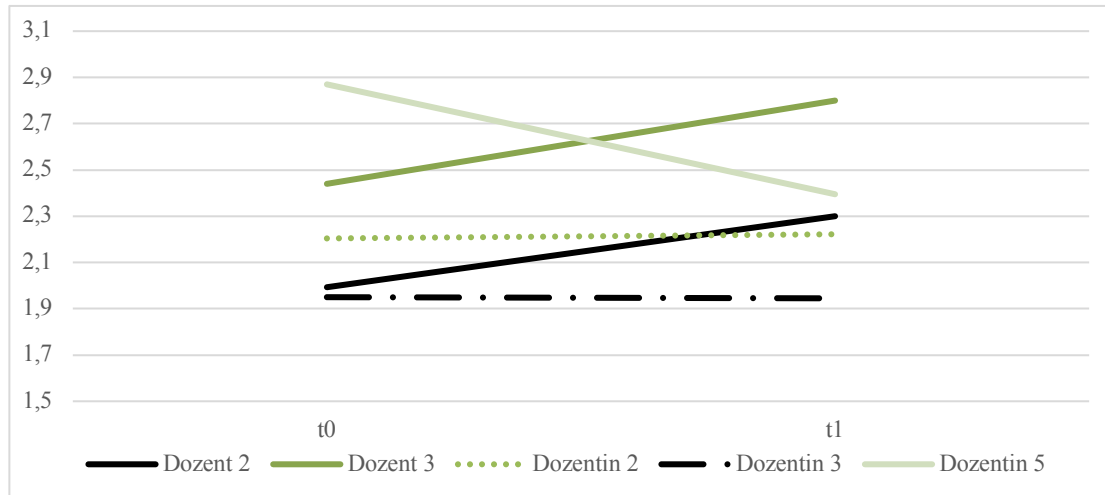


Abb. 9: Randmittelwerte für die Interaktion aus der Entwicklung und den Lehrenden zum Zeitpunkt  $t_1$  im Modul BIW

Wie Abbildung 9 zeigt, erzielten die Lernenden geringere Werte auf dem Beobachtungsbogen, wenn Dozentin 5 die Bewertungen zum Zeitpunkt  $t_1$  vornimmt. Kaum eine Veränderung in den Werten ist für Dozentin 2 und Dozentin 3 festzustellen. Bei Dozenten 2 und 3 kommt es hingegen zu einer Steigerung der mit dem Beobachtungsbogen erzielten Werte. Hier lässt sich keine systematische Abhängigkeit zwischen *strengen* und *weniger strengen* Lehrenden erkennen.

## 5 Diskussion

Die Einführung in die Besonderheiten beruflichen Handelns stellt für Professionen ein konstitutives Element dar, welches zudem auf einen (selbst-)reflexiven Umgang mit beruflichen Situationen angewiesen ist (vgl. z. B. Helsper, 2011, S. 150). Während die Wirkung von Feedback auf der Grundlage von Beobachtungen nachweislich zu einer Steigerung der (Selbst-)Reflexionsfähigkeit von angehenden Lehrkräften führt (vgl. Jahncke et al., 2018), ist bislang wenig darüber bekannt, wie Beobachtungsbögen konkret zur Unterstützung professionellen Handelns im Hochschulstudium eingesetzt werden. Auf der Grundlage dieser Studie lassen sich die zu Beginn ermittelten Fragestellungen wie folgt diskutieren.

Die Ergebnisse zeigen zunächst, dass eine Entwicklung nur in den Modulen festzustellen ist, die über mindestens zwei Semester verlaufen (konkret PB und BIW). Dies deutet darauf hin, dass die Verbesserung professionellen Handelns nicht unmittelbar nach einem Feedback erfolgt, sondern Zeit und Raum benötigt. Dies entspricht Ergebnissen aus der Expertiseforschung, die davon ausgehen, dass für den Erwerb eines Expertenstatus mindestens 10 Jahre Berufspraxis notwendig sind (vgl. Gruber, 2001, S. 5146) und erklärt auch die geringe Varianzaufklärung des Faktors *Zeit* von gerade einmal 1 % bis 2 %. Für die Gestaltung von Studiengängen hat dies weitreichende Konsequenzen. So steht das Ziel, bereits nach drei Jahren einen berufsqualifizierenden Bachelorabschluss zu erwerben, d. h. also deutlich *früher* als in den alten Diplom- und Magisterstudiengängen, dem Zeitbedarf entgegen, der für die Einführung in die professionelle Handlungspraxis benötigt wird. Diese

Problematik spiegelt sich auch in den Daten wider. So profitieren die Studierenden im Bachelor-Modul PB nicht von einer Steigerung der Bewertungen auf dem Beobachtungsbogen. Eine solche Steigerung ist erst in dem Mastermodul BIW festzustellen. Für die Gestaltung von Lehramtsstudiengängen hat dieser Befund zur Konsequenz, dass die Beobachtungsbögen künftig nur noch im Masterstudium eingesetzt werden sollten, da die Studierenden im Bachelorstudium für eine ausreichend differenzierte Nutzung der Beobachtungsergebnisse noch nicht weit genug fortgeschritten sind. Beobachtungsbögen sind vor allem dann sinnvoll, wenn es weniger um pauschales, sondern um nuanciertes, hoch individuelles Feedback geht, wie es sich erst während des Masterstudiums sinnvoll an die angehenden Lehrkräfte geben lässt.

Obwohl die Skalen des Beobachtungsbogens Werte für Cronbachs  $\alpha$  erzielen, die mit Ausnahme der Skala für *Struktur und Zielorientierung* über dem in der pädagogischen Psychologie geforderten Grenzwert von .700 liegen (vgl. Adam & Lawrence, 2015, S. 90), d. h. auf den ersten Blick reliable Messungen möglich sind, darf dieser Befund nicht darüber hinweg täuschen, dass die erzielten Messergebnisse stark von den Dozentinnen und Dozenten abhängig sind. Hier ermitteln die Varianzanalysen die Lehrenden als stärksten Einflussfaktor. Entsprechend der Definition von Objektivität sind die Messergebnisse damit nicht unabhängig vom Testanwender, d. h. die Objektivität deutlich eingeschränkt (vgl. dazu z. B. Bortz & Döring, 2009, S. 195; Bühner, 2011, S. 59 f.). Aufgrund dieser Einschränkungen bei den Hauptgütekriterien, sollten die mit Hilfe des Beobachtungsbogens erzielten Ergebnisse nicht für die Bewertung von Leistungen im Sinne der Vergabe von Qualifikationen genutzt werden, sondern nur zur Unterstützung individueller Lernprozesse (vgl. dazu Helmke, 2015, S. 124). Einem Einsatz zur Bewertung steht ebenfalls entgegen, dass die meisten Lehrenden in ihren Bewertungen, soweit sie im Rahmen dieser Arbeit analysiert worden sind, einen offensichtlichen Diagnosefehler aufweisen. Die Art dieses Fehlers unterscheidet sich dabei jedoch: es sind sowohl zu große milde (z. B. Dozent 3) als auch strenge Effekte (z. B. Dozentinnen 1 und 3) im Datensatz zu finden.

Für mögliche pädagogisch vorteilhafte Verzerrungseffekte kann vor diesem Hintergrund das in Abbildung 6 dargestellte Resultat als Beispiel dienen. Die Ergebnisse zeigen, dass eher strenge oder durchschnittliche Rückmeldungen zum ersten Zeitpunkt zu Verbesserungen in den Beobachtungswerten zum zweiten Zeitpunkt führen, während eine zu milde Bewertung zum ersten Zeitpunkt (Dozent 3) keinen oder einen eher leicht abnehmenden Effekt zum zweiten Zeitpunkt zur Folge hat. Die strengeren Bewertungen schaffen vor diesem Hintergrund offenbar mehr Anknüpfungspunkte für die Studierenden, ihr eigenes professionelles Handeln zu verbessern. Vor dem Hintergrund, dass die Ergebnisse nicht zur Bewertung verwendet werden sollten, lässt sich auf der Grundlage der analysierten Daten im Zweifelsfall eine eher strenge Bewertung empfehlen, um angehende Lehrkräfte in ihrem Kompetenzerwerb zu unterstützen. Pädagogisch bedeutet dies, dass die Tendenz zu einer Unterschätzung der Qualität der Unterrichtsgestaltung sich anscheinend in einem ausführlicheren Feedback auf der Grundlage des Beobachtungsbogens äußert und damit bessere Voraussetzungen für die Entwicklung der Professionalität bietet als eine Tendenz zur Überschätzung der Qualität der Unterrichtsgestaltung durch die Lehrenden. Eine solche eher strenge Bewertung sollte aber nur zur Förderung und Unterstützung der

Professionalisierung erfolgen. Sobald mit der Einschätzung die Vergabe von Qualifikationen, z. B. in Form von Noten, Beurteilungen von Leistungen für die Personalakte oder die Vergabe von Ressourcen gekoppelt sind, ist eine solche Bewertung nicht zulässig. Hier sind möglichst genaue Einschätzungen notwendig.

Schließlich zeigen die Daten hohe Korrelationen zwischen den Skalen, die in der Regel das Fornell-Larcker-Kriterium nicht erfüllen. Demnach ist die diskriminante Validität des Beobachtungsbogens fraglich und von deutlichen Überschneidungen auszugehen. Vor dem Hintergrund, dass Beobachtungsbögen grundsätzlich möglichst wenig Items beinhalten sollten, damit sie anwendbar bleiben (vgl. dazu Kromrey, Roose & Strübing, 2016, S. 326), empfiehlt sich daher eine deutliche Reduzierung der Beobachtungsitems. Zudem sollte geprüft werden, ob stattdessen andere Items in den Fragebogen aufgenommen werden können, um eine umfassende und vollständige Unterstützung bei der Einführung in die professionelle Handlungspraxis zu gewährleisten. Neben der fehlenden Trennschärfe der Items selber ist hier allerdings auch die Nutzung durch die Lehrenden kritisch zu hinterfragen, da diesen potentiell noch die notwendige Routine und Erfahrung in der Anwendung des Beobachtungsbogens fehlte. Somit sollten als weitere Konsequenz künftig auch die Lehrenden in Bezug auf den Umgang mit Beobachtungsbögen gezielter geschult und vorbereitet werden.

Zusammenfassend lassen sich damit die zu Beginn des Beitrages herausgearbeiteten Fragestellungen wie folgt beantworten:

Der Einsatz von Beobachtungsbögen in der Lehrerbildung kann *nicht* als objektiv betrachtet werden (Fragestellung 1), obwohl statistische Maße für einen relativ geringen Zufallsfehler und somit für eine zuverlässige Einschätzung sprechen (Fragestellung 2). Da sich die Güte eines Instruments stets an dem schwächsten Merkmal bemessen lassen muss, d. h. hier an der Objektivität, sind die Einschätzungen mittels der Beobachtungsbögen nicht zur Vergabe von Qualifikationen, sondern nur zur Förderung des individuellen Lernens geeignet (Fragestellung 5). Hier zeigen die Daten, dass Rückmeldungen mit einem Beobachtungsbogen ihre Wirkung weniger im Bachelorstudium, sondern vor allem im Master entfalten (Fragestellung 3). Dabei verdeutlichen die Daten, dass eine Verzerrung im Sinne einer strengen Rückmeldung der Lehrenden die Entwicklung der Studierenden besonders unterstützt (Fragestellung 4).

Die Resultate dieser Studien besitzen den Vorteil, dass die Daten *im Feld* erhoben werden konnten. Gleichzeitig limitiert dieser Umstand den Geltungsumfang der ermittelten Ergebnisse. Beispielsweise konnten in den Varianzanalysen aus Tabelle 5 nicht alle Formen der Interaktion analysiert werden. Auch eine umfassende Darstellung der Entwicklung der Studierenden über das gesamte Studium ist zum jetzigen Zeitpunkt noch nicht möglich, da noch zu wenige Studierende über das gesamte Studium mit dem Beobachtungsbogen begleitet wurden. Hier bestehen Ansatzpunkte für zukünftige Studien. Langfristig bietet es sich darüber hinaus an, historische Daten aus den Beobachtungsbögen mit Blick auf eine künftig noch zu erfolgende Messung von beruflichem Erfolg zu analysieren. Einen passenden Rahmen hierfür könnte beispielsweise eine Langzeitstudie über Unterrichtsqualität bieten, die klärt, in welchem Ausmaß die Ergebnisse auf den Beobachtungsbögen tatsächlich den späteren beruflichen Erfolg der Studierenden vorhersagen können.

## Literatur

- Abels, S. (2011). *LehrerInnen als "Reflective Practitioner" – Reflexionskompetenz für einen demokratieförderlichen Naturwissenschaftsunterricht*. Wiesbaden: Springer.
- Adams, K. A. & Lawrence, E. K. (2015). *Research methods, statistics, and applications*. Thousand Oaks: Sage.
- Atteslander, P. (2010). *Methoden der empirischen Sozialforschung* (13. Aufl.). Berlin: Erich Schmidt.
- Bagozzi, R. P. & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94.
- Bain, J. D., Ballantyne, R. & Packer, J. (1999). Teachers and teaching. Theory and practise using journal writing to enhance student teachers' Reflectivity during field experience placements. *Teachers and Teaching. Theory and Practice*, 5(1), 51–73.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 4, 469–520.
- Bortz, J. & Döring, N. (2009). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson.
- Bühner, M. & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* (2. Aufl.). München: Pearson.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Deci, E. L. & Ryan, R. M. (2012). Motivation, personality, and development within embedded social contexts: An overview of Self-Determination Theory. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 85–107). Oxford: Oxford University Press.
- Dewey, J. (1910). *How we think*. Lexington: D.C. Heath.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer.
- Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement errors. *Journal of Marketing Research*, 18(1), 39–50.
- Gage, N. L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press.
- Göhlich, M. (2011). Reflexionsarbeit als pädagogisches Handlungsfeld. Zur Professionalisierung der Reflexion und zur Expansion von Reflexionsprofessionellen in Supervision, Coaching und Organisationsberatung. *Zeitschrift für Pädagogik, Beiheft 57*, 138–152.
- Graff, T. (2014). „Wann soll ich das noch machen?“ (Selbst)Reflexion von Lehrerinnen und Lehrern im Sachunterricht. URL: [https://www.uni-hildesheim.de/media/forschung/fff/PDFs/Schriftenreihe/Graff\\_05-2014.pdf](https://www.uni-hildesheim.de/media/forschung/fff/PDFs/Schriftenreihe/Graff_05-2014.pdf) [06.03.2018].
- Gruber, H. (2001). Expertise, acquisition of. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 8; pp. 5145–5150). Oxford: Elsevier.
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (6. Aufl.). Seelze: Kallmeyer.

- Helsper, W. (2011): Lehrerprofessionalität – der strukturtheoretische Professionsansatz zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.): *Handbuch zur Forschung zum Lehrerberuf* (S. 149-170). Münster: Waxmann.
- Häder, M. (2015). *Empirische Sozialforschung: Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hesse, I. & Latzko, B. (2011). *Diagnostik für Lehrkräfte* (2. Aufl.). Opladen: Budrich.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling*, 6(1), 1-55.
- Jahncke, H. (2019). *(Selbst-)Reflexionsfähigkeit – Modellierung, Differenzierung und Beförderung mittels eines Kompetenzentwicklungsportfolios*. München: Hampp.
- Jahncke, H., Berding, F., Porath, J. & Magh, K. (2018). Einfluss von Feedback auf die (Selbst-)Reflexion von Lehramtsstudierenden. *die hochschullehre*, 4, online unter: <http://www.hochschullehre.org/?m=201809>.
- Janík, T. (2003). Zur reflexiven LehrerInnenbildung. In R. Seebauer (Hrsg.), *Erste Schritte in die Unterrichtspraxis – Texte, Materialien, Berichte* (S. 78–84). Brno: Paido.
- KMK (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (Hrsg.) (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. URL: [http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung.pdf](http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf) [18.01.2016].
- KMK (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (Hrsg.) (2014). URL: [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf) [27.12.2017].
- Korthagen, F. (1999). Linking reflection and technical competence: the logbook as an instrument in teacher education. *European Journal of Teacher Education*, 22(2/3), 191–207.
- Kromrey, H., Roose, J. & Strübing, J. (2016). *Empirische Sozialforschung. Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung*. (13. Aufl.). Stuttgart: utb.
- Leonhard, T. & Abels, S. (2017). Der „reflective practitioner“. Leitfigur oder Kategorienfehler einer reflexiven Lehrerinnen- und Lehrerbildung? In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Reflexive Lehrerbildung revisited. Traditionen – Zugänge – Perspektiven* (S. 46–55). Bad Heilbrunn: Klinkhardt.
- Nguyen, Q. D., Fernandez, N., Karsenti, T. & Charlin, B. (2014). What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Medical Education*, 48(12), 1176–1189.
- Oevermann, U. (1996). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. In A. Combe & W. Helsper (Hrsg.), *Pädagogische Professionalität. Untersuchungen zum Typus pädagogischen Handelns* (S. 70–182). Frankfurt a. M.: Suhrkamp.
- Oevermann, U. (2008). Profession contra Organisation? In W. Helsper, S. Busse, M. Hummrich & R.-T. Kramer (Hrsg.), *Pädagogische Professionalität in Organisationen* (S. 55–77). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Palmer, D. J., Stough, L. M., Burdinski, J. & Gonzales, M. (2005). Identifying Teacher

- Expertise: An Examination of Researchers' Decision Making. *Educational Psychologist*, 40(1), 13–25.
- Reinisch, H. (2009). „Lehrprofessionalität“ als theoretischer Term. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehr-professionalität* (S. 33–43). Weinheim: Beltz.
- Ryan, R. M. & Deci, E. L. (2002). Overview of Self-Determination Theory: An organismic dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3–33). Rochester: University of Rochester Press.
- Schnell, R., Hill, P. B. & Esser, E. (2013). *Methoden der empirischen Sozialforschung* (10. Aufl.). München: Oldenbourg.
- Schubarth, W., Speck, K., Seidel, A. Gottmann, C., Kamm, C., Kopp, A. & Krohn, M. (2011). Nach Bologna: Warum das Neue (manchmal) nicht in die Hochschule kommt. Das Beispiel Praxisphasen im Studium. *Zeitschrift für Hochschulentwicklung* (6)3, 74–88.
- Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.
- Stichweh, R. (2005). Wissen und die Professionen in einer Organisationsgesellschaft. In T. Klatetzki & V. Tacke (Hrsg.), *Organisation und Profession*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stock, M. & Riebenbauer, E. (2014). *Wegweiser durch das Thema (Selbst-)Reflexion*. URL: [https://static.unigraz.at/fileadmin/sowiinstitute/Wirtschaftspaedagogik/Neuigkeiten/Wegweiser\\_durch\\_die\\_Reflexion\\_Kr%C3%A4mer\\_Strassegger.pdf](https://static.unigraz.at/fileadmin/sowiinstitute/Wirtschaftspaedagogik/Neuigkeiten/Wegweiser_durch_die_Reflexion_Kr%C3%A4mer_Strassegger.pdf) [01.06.2016].
- Sung, Y.-T., Chang, K.-E. & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, 94(2016), 252–275.
- Weinert, F. E. & Schrader, F.-W. (1986). Diagnose des Lehrers als Diagnostiker. In H. Petillon, J. W. L. Wagner & B. Wolf (Hrsg.), *Schülergerechte Diagnose* (S. 11–29). Weinheim: Beltz.

## Autor/-innen

PD Dr. Florian Berding (Universität Oldenburg), Fachgebiet Berufs- und Wirtschaftspädagogik, Oldenburg, Deutschland, [florian.berding@uni-oldenburg.de](mailto:florian.berding@uni-oldenburg.de)

Dr. Heike Jahncke (Universität Oldenburg), Fachgebiet Berufs- und Wirtschaftspädagogik, Oldenburg, Deutschland, [heike.jahncke@uni-oldenburg.de](mailto:heike.jahncke@uni-oldenburg.de)



**Zitiervorschlag:** Berding, F. & Jahncke, H. (2019). Anwendung, Nutzung und Grenzen von Beobachtungsbögen als Instrument zur Professionalisierung im Hochschulstudium. *die hochschullehre*, Jahrgang 5/2019, online unter: [www.hochschullehre.org](http://www.hochschullehre.org)