

Ger J Exerc Sport Res 2017 · 47:246–254
 DOI 10.1007/s12662-017-0453-4
 Eingegangen: 4. Juli 2016
 Angenommen: 5. April 2017
 Online publiziert: 26. April 2017
 © Der/die Autor(en) 2017. Dieser Artikel ist
 eine Open-Access-Publikation.



Matthias Baumgartner

Eidg. Hochschule für Sport Magglingen, Magglingen, Schweiz

„Denn sie wissen nicht, was sie können ...!“ – die Qualität der Performanzen von angehenden Sportlehrkräften als Verzerrer der Selbstbeurteilung

Forschungslage und theoretischer Hintergrund

Integration der performativen Ebene in die Kompetenzdiagnostik: eine realistische Wende

In den vergangenen Jahren hat sich die evidenzbasierte Datenlage verdichtet, die den bedeutsamen Einfluss von Lehrkräften auf die Schulleistungen der Schülerinnen und Schüler bestärken (Hattie, 2009; Scheerens & Bosker, 1997). Unter Berücksichtigung dieser Erkenntnisse wurde die Erforschung der Effektivität der Ausbildung von Lehrkräften zu einem zentralen Gegenstand der empirischen Bildungsforschung (Blomberg, Seidel, & Prenzel, 2011). Im Rahmen dieser Forschungsbemühungen hat sich in den letzten zwei Dekaden das Paradigma der kompetenzorientierten Forschung weitgehend etabliert (Blömeke, Felbrich, & Müller, 2008). Diese Entwicklung ist auch in der Forschung zur Wirksamkeit der Ausbildung von Sportlehrkräften festzustellen (Baumgartner, 2013a, 2013b; Meier, 2015). In der kompetenzorientierten Forschung werden u. a. Aspekte professioneller Kompetenz (z. B. Professionswissen, motivationale Orientierungen, Selbstregulation, Überzeugungen, Werthaltungen und Ziele; Baumert & Kunter, 2011; Terhart, 2002) oder professionelle Kompetenzen (z. B. lernträchtiges Feedback geben können; Baumgartner, 2013b; Oser, 2001) benannt, welche die angehenden Lehr-

kräfte zur Bearbeitung von beruflichen Anforderungssituationen benötigen. Mit der Setzung von solchen verpflichtenden Zielen wird die Wirksamkeit der Ausbildung von Lehrkräften überprüfbar (Czerwenka & Nölle, 2014; Darling-Hammond, 2000).

In der kompetenzorientierten Forschung werden verschiedene Begrifflichkeiten häufig als Synonyme oder nicht trennscharf verwendet. Im Folgenden sollen dementsprechend Begriffsbestimmungen vorgenommen werden. Der Ausdruck *professionelle Kompetenz* – hier in singularisierter Form – wird als ein latentes Konstrukt gedeutet (Blömeke, Gustafsson, & Shavelson, 2015; Hager & Hasselhorn, 2000; Shavelson, 2013), das auf verschiedenen Aspekten professioneller Kompetenz basiert (z. B. Professionswissen, motivationale Orientierungen etc.; Baumert & Kunter, 2011). Werden die Beiträge zur Erforschung der professionellen Kompetenz von Lehrkräften betrachtet, so wird in den Forschungsbemühungen in erster Linie der Zusammenhang zwischen den einzelnen Aspekten professioneller Kompetenz und den Schulleistungen der Schülerinnen und Schüler erforscht (Kunter et al., 2011). Die Kompetenzdiagnostik basiert in dieser *analytischen kompetenzorientierten Forschungstradition* auf Papier-Bleistift-Tests, was problematisch ist. Denn in einer derartigen peripheren Kompetenzdiagnostik wird die Umsetzung der Aspekte professioneller Kompetenz in einer konkreten

beruflichen Anforderungssituationen kaum berücksichtigt (Doyle, 2006; Shavelson, 2013). *Kompetenz* ist jedoch mehr als das summarische Zusammenfügen einzelner Aspekte professioneller Kompetenz zu deuten, sondern weist einen Bezug zur beruflichen *realen Welt* auf (Baumgartner, 2016; Blömeke et al., 2015; McClelland, 1973; Shavelson, 2013). Aus ökologischen Validitätszwecken wird dementsprechend in jüngster Zeit auch gefordert, in der Kompetenzdiagnostik von angehenden Lehrkräften die performative Ebene im Sinne *anforderungsspezifischen Könnens in realen beruflichen Situationen* stärker zu berücksichtigen (Blömeke et al., 2015; McClelland, 1973; Shavelson, 2010, 2013). Denn je enger die Kompetenzdiagnostik an das Handeln in realen Anforderungssituationen gekoppelt ist, desto präziser wird gemessen, was die angehenden Lehrkräfte in beruflichen Anforderungssituationen auch leisten können (Shavelson, 2010). In der *holistischen kompetenzorientierten Forschungstradition* wird daher versucht, die performative Ebene stärker zu berücksichtigen. Dazu werden aus der Analyse beruflicher Anforderungssituationen und unter Berücksichtigung der vorliegenden theoretischen und empirischen Erkenntnisse *professionelle Kompetenzen* – hier in pluralisierter Form – extrahiert (Baumgartner, 2013b; Heinzer & Baumgartner, 2013; McClelland, 1973; Oser, 2001; Vonken, 2005). Professionelle Kompetenzen werden als „can-do statements“ (z. B. lernträcht-

ges Feedback geben können, effektives Klassenmanagement realisieren können; Heinzer & Baumgartner, 2013) formuliert, wodurch der Bezug zur beruflichen *realen Welt* bereits in der Benennung von professionellen Kompetenzen hergestellt wird. Professionelle Kompetenzen stellen dementsprechend Anforderungsprofile für den Lehrberuf (Oser, 2001) resp. für den Sportlehrberuf (Baumgartner, 2013b) dar (Terhart, 2007). Die Qualität von solchen professionellen Kompetenzen (z. B. lernträchtiges Feedback geben können) lässt sich letzten Endes jedoch einzig auf performativer Ebene ökologisch valide diagnostizieren (Shavelson, 2013), d. h. anhand der Qualität der *kompetenzbereichsbezogenen Performanzen* (z. B. die beobachtbare Qualität des Feedbacks einer angehenden Sportlehrkraft in konkreten Unterrichtssituationen und im Sinne von Können).

Selbst- und Fremdbeurteilung in der Kompetenzdiagnostik

Obschon in der holistischen kompetenzorientierten Forschungstradition in jüngerer Zeit versucht wird, durch verschiedene Verfahren die performative Ebene in die Kompetenzdiagnostik zu integrieren (Blömeke, König, Suhl, Hoth, & Döhrmann, 2015; Forster-Heinzer & Oser, 2015), basiert die Messung von professionellen Kompetenzen bei angehenden Lehrkräften häufig auf allgemeinen Selbstbeurteilungsverfahren (Frey, 2006; Oser, 2001). Mittels Fragebogen oder Interviews bewerten die angehenden Lehrkräfte die Qualität ihrer eigenen kompetenzbereichsbezogenen Performanzen bzw. die Güte ihrer Ausbildung hinsichtlich der eigenen Performanzentwicklung (Oser, 2001). Der Vorteil dieser methodischen Vorgehensweise besteht darin, dass mit einem relativ geringen Aufwand eine große Stichprobe untersucht werden kann. Der Nachteil von Selbstbeurteilungsverfahren liegt jedoch in den absichtlichen oder unabsichtlichen Einschätzungsverfälschungen, was die Qualität der Daten gefährdet und dementsprechend als problematisch erscheint (König & Tachtsoglou, 2012; Terhart, 2007). Selbstbeurteilungsverfahren gelten demgemäß nicht als eine

zuverlässige Form der Kompetenzdiagnostik (Terhart, 2007). Folglich stehen einige Forschende kritisch zu dieser Methode (Hartig & Klieme, 2006; Kunter & Klusmann, 2010; Shavelson, 2013) oder lehnen die Selbstbeurteilung als kompetenzdiagnostische Vorgehensweise gänzlich ab (Hartig & Jude, 2007). Des Weiteren wird die performative Ebene durch die Selbstbeurteilung kaum berücksichtigt, was aus ökologischen Validitätszwecken als problematisch zu betrachten ist (Shavelson, 2010, 2013). Eher wird in Selbstbeurteilungsverfahren das Kompetenzselbstkonzept gemessen (Abs, 2007; Hartig & Jude, 2007) und nicht die professionelle Kompetenz bzw. die kompetenzbereichsbezogenen Performanzen von angehenden Lehrkräften. Unter Berücksichtigung dieser Kritik erscheint es als wünschenswert, in der Kompetenzdiagnostik intersubjektive Fremdeinschätzungsverfahren einzusetzen, um die Qualität der kompetenzbereichsbezogenen Performanzen von angehenden Lehrkräften eruieren zu können.

Wird die Datenlage zur Übereinstimmung von Selbst- und Fremdbeurteilung betrachtet, so ist diese als widersprüchlich zu interpretieren. Befunde aus der Unterrichtsforschung weisen darauf hin, dass angehende Lehrkräfte die Qualität der eigenen kompetenzbereichsbezogenen Performanzen im Vergleich zu ihren Praxislehrkräften im Allgemeinen kritischer bewerten (Bodensohn & Schneider, 2009; Boekhoff, Franke, Dietrich, & Arnold, 2008). In anderen Studien werden kaum feststellbare Unterschiede berichtet (Moser & Hascher, 2000). Da die Qualität der kompetenzbereichsbezogenen Performanzen der angehenden Lehrkräfte in den benannten Studien jedoch nicht durch ein intersubjektives Fremdbeurteilungsverfahren (z. B. Expertinnen- und Expertenratings) erfasst wurde, können Verzerrungseffekte in der Selbstbeurteilung der angehenden Lehrkräfte sowie in der (subjektiven) Fremdbeurteilung der Praxislehrkräfte nicht ausgeschlossen werden.

Als ein bedeutsamer kognitiver Verzerrer der Selbstbeurteilung gilt die Qualität der eigenen Performanzen. So gehen Kruger und Dunning (1999) unter

Berücksichtigung der Erkenntnisse aus der Expertiseforschung u. a. davon aus, dass *Inkompetente* mit geringen Performanzen im Vergleich zu *Kompetenten* mit hohen Performanzen vor dem Hintergrund ihres Nichtwissens, ihrer qualitativ geringfügig ausgeprägten Strategien und der fehlenden Metakognition sich ihrer eigenen Inkompetenz nicht bewusst sind.¹ *Inkompetente* verfügen sozusagen nicht über die notwendigen Metakognitionen, um die eigenen Performanzen adäquat beurteilen zu können bzw. sind nicht fähig, die Qualität des Handelns in komplexen Anforderungssituationen angemessen einzuschätzen. *Inkompetente* neigen nach Kruger und Dunning (1999) demzufolge dazu, die Qualität ihrer eigenen Performanzen zu überschätzen. Um diese forschungsleitenden Annahmen empirisch zu prüfen, führten Kruger und Dunning (1999) mit Studierenden ($n = 45$) der Cornell Universität (USA) eine empirische Studie durch. Mit der Verwendung eines logischen Denktests (Orton, 1993) wurden u. a. die testbezogenen Performanzen der Versuchspersonen erhoben. Nach der Durchführung des Tests wurden die Versuchspersonen dazu aufgefordert, ihre eigenen testbezogenen Performanzen zu beurteilen. Aus den Ergebnissen geht hervor, dass die *inkompetenten* Versuchspersonen aus dem untersten Leistungsquartil ($n = 11$) ihre eigenen testbezogenen Performanzen im Vergleich zu den objektiv diagnostizierten Performanzen hoch signifikant überschätzten ($t(10) = 7,66; p < 0,001$). Der Effekt des gefundenen Unterschieds ist als stark zu bezeich-

¹ *Inkompetenz* ist an dieser Stelle weder als stabiles Persönlichkeitskonstrukt noch als abwertende Begrifflichkeit von *Inkompetenten* zu deuten. Mit dem Begriff *Inkompetenz* wird der Fokus unter Berücksichtigung dessen, dass Kompetenz erwerbbar ist, auf das *Noch-nicht-Wissen* und/oder das *Noch-nicht-Können* gerichtet.

nen ($d = 3,26$).² Obschon der Unterschied knapp nicht signifikant ist, weisen die Daten darauf hin, dass die *kompetenten* Versuchspersonen ($n = 13$) aus dem obersten Leistungsquartil ihre eigenen testbezogenen Performanzen im Vergleich zu den diagnostizierten Performanzen tendenziell unterschätzen ($t(12) = -2,15$; $p < 0,06$; $d = 84$). Die Teilstudie nach Kruger und Dunning (1999) betrachtend, kann zusammenfassend festgehalten werden, dass a) *Inkompetente* in der Selbstbeurteilung ihre testbezogenen Performanzen bedeutsam überschätzten und b) *Kompetente* die Qualität ihrer Performanzen tendenziell unterschätzten. Diese Ergebnisse weisen darauf hin, dass die Zuverlässigkeit der Daten in Selbstbeurteilungsverfahren als kritisch zu betrachten ist.

Feedback von Sportlehrerinnen- und Sportlehrern: ein Herzstück des Sportlehrberufs

Für einen gelingenden Sportunterricht benötigen Sportlehrerinnen und Sportlehrer eine Vielzahl an unterschiedlichen professionellen Kompetenzen (Baumgartner, 2013a, 2013b, 2016; Meier, 2015). Ein relevanter Kompetenzbereich stellt das Feedback von Sportlehrerinnen und Sportlehrer an die Schülerinnen und Schüler dar, denn Feedback wird aus der Perspektive von tätigen Sportlehrkräften der Berufsfachschule als bedeutsam bewertet (Baumgartner, 2013a, 2013b). Darüber hinaus wird aus fachübergreifenden (Hattie & Timperley, 2007; Kluger & DeNisi, 1996) und fachspezifischen Studien (Lee, Keh, & Magill, 1993; Magill, 1993; Marschall & Daus, 2003; Silverman, Tyson, & Krampitz, 1992) deutlich, dass das Feedback von Lehrkräften und Sportlehrkräften ein gewichtiger Faktor hinsichtlich der Beein-

² In der Originalstudie wurden keine Standardabweichungen (SD) und Effektstärkenmasse (z. B. Cohens d) berichtet. Dementsprechend basiert der Effektstärkekoeffizient Cohens d auf eigenen Berechnungen. Da in der Originalstudie die Kennwerte der bivariaten Korrelation (r) jedoch nicht dargelegt wurden, resultiert die Berechnung des Cohens d auf einem $r = 0$, wodurch die Schätzung der Effektstärken konservativ zu betrachten ist.

Ger J Exerc Sport Res 2017 · 47:246–254 DOI 10.1007/s12662-017-0453-4
© Der/die Autor(en) 2017. Dieser Artikel ist eine Open-Access-Publikation.

M. Baumgartner

„Denn sie wissen nicht, was sie können ...!“ – die Qualität der Performanzen von angehenden Sportlehrkräften als Verzerrer der Selbstbeurteilung

Zusammenfassung

In der Kompetenzdiagnostik von angehenden Sportlehrkräften nimmt die Selbstbeurteilung einen prominenten Stellenwert ein. Bislang liegen jedoch keine Studien vor, mit Hilfe derer die Datenqualität von Selbstbeurteilungsverfahren bei angehenden Sportlehrkräften quantifiziert wurde. Dies erscheint als problematisch, denn so wird nicht klar, ob in Selbstbeurteilungsverfahren absichtliche oder unabsichtliche Einschätzungsverfälschungen wirksam werden. Befunde weisen diesbezüglich darauf hin, dass *Inkompetente* aufgrund ihres Nichtwissens oder Nichtkönnens die Qualität ihrer Performanzen im Vergleich zur objektiven Leistung überschätzen bzw. *Kompetente* sich unterschätzen. In der vorliegenden Studie wurde untersucht, ob sich dieser kognitive Verzerrungseffekt bei angehenden Sportlehrkräften nachweisen lässt. Dabei beurteilten angehende Sportlehrkräfte aus den zwei Untersuchungsgruppen der

Kompetenten ($n = 13$) bzw. der *Inkompetenten* ($n = 10$) mittels Videovignetten die eigene feedbackbezogenen Performanzen. Die Mittelwerte der beiden Untersuchungsgruppen wurden mit denjenigen eines intersubjektiven Fremdbeurteilungsverfahrens (Expertinnen- und Expertenrating) verglichen. Die Ergebnisse zeigen auf, dass die *Inkompetenten* ihre feedbackbezogene Performanz im Vergleich zum intersubjektiven Fremdurteil deutlich über- und sich die *Kompetenten* unterschätzten. Die Befunde weisen u. a. auf die hohe Bedeutsamkeit objektiver bzw. intersubjektiver kompetenzdiagnostischer Verfahren hin, womit der Bezug zur performativen Ebene der angehenden Sportlehrkräfte hergestellt wird.

Schlüsselwörter

Sportlehrkräfte · Selbstbeurteilung · Feedback · Professionelle Kompetenz · Kompetenzdiagnostik

“Because they don’t know what they are capable of...!” – the quality of the performance of trainee PE teachers as a distorting factor in self-assessment

Abstract

Self-assessment plays an important role in evaluating the competence of trainee PE teachers. Up to now, however, there have been no studies to help quantify the data quality of self-assessment procedures for trainee PE teachers. This poses a problem, because intentional or unintentional distortions of these assessments cannot be ruled out. Findings show that – due to a lack of knowledge or skills – *incompetent* staff tend to overestimate the quality of their performance compared with their objective performance while *competent* teachers underestimate themselves. This study examined whether this cognitive distortion effect can be demonstrated in the case of trainee PE teachers. Trainee PE teachers from two study groups – competent ($n = 13$) and incompetent ($n = 10$) – assessed their own feedback-related

performance using *video vignettes*. The means for the two groups were compared with an intersubjective external evaluation (expert rating). The results show that the *incompetent* trainee teachers clearly overestimate their feedback-related performance compared with the intersubjective external rating, while *competent* trainee teachers underestimate their performance. The findings point, inter alia, to the high level of significance of objective and intersubjective competence-diagnostic procedures, which is where the link to the trainee PE teachers’ performative levels comes into play.

Keywords

PE teachers · Self-assessment · Feedback · Professional competence · Competence diagnostics

flussung von Schulleistungen darstellt. Die Forschungsbefunde weisen darauf hin, dass das Feedback von Sportlehrkräften jedoch nicht *per se* einen positiven Einfluss auf das motorische Lernen der Schülerinnen und Schüler hat und mehr Feedback auch nicht automatisch zu höheren Schulleistungen führt (Lee et al., 1993; Marschall & Daus, 2003; Silverman et al., 1992). Es wird deutlich, dass ein Feedback von Sportlehrkräften hinsichtlich des motorischen Lernens dann effektiv ist, wenn ein (positiver) spezifischer und präziser Informationsgehalt transportiert wird, wodurch die Schülerinnen und Schüler die Dissonanz zwischen Ist- und Sollzustand hinsichtlich eines gesetzten Zieles verringern bzw. korrigieren können (*Spezifität von Feedback*; Lee et al., 1993; Marschall & Daus, 2003; Silverman et al., 1992). Dieses Ergebnis korrespondiert mit den Befunden aus fachübergreifenden Effektivitätsstudien (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). Des Weiteren wird ersichtlich, dass das Feedback von Lehrkräften (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008) bzw. Sportlehrkräften (Marschall & Daus, 2003; Silverman et al., 1992) dann eine hohe Wirkung hat, wenn der konkrete Leistungsstand der Schülerinnen und Schüler berücksichtigt wird und die Lernenden die erhaltenen Informationen somit auch umsetzen können (*Realisierbarkeit von Feedback*). Betreffend die Frequenz des Feedbacks von Sportlehrkräften ist die Datenlage inkonsistent (Lee et al., 1993; Marschall & Daus, 2003). Trivialerweise kann das hohe Potenzial des Feedbacks jedoch nur dann ausgeschöpft werden, wenn die Schülerinnen und Schüler von ihren (Sport-)Lehrkräften auch ein spezifisches und realisierbares Feedback erhalten, was häufig nicht der Fall ist (*Regelmäßigkeit von Feedback*; Hattie, 2009; Voerman, Meijer, Korthagen, & Simons, 2012). Die Ergebnisse aus fachübergreifenden Effektivitätsstudien weisen zudem darauf hin, dass zu häufiges Feedback den Lernprozess der Schülerinnen und Schüler bei komplexen Aufgaben oder Übungen unnötigerweise unterbrechen kann. Folglich sollte bei schwierigen Aufgabenstellungen nicht

zu häufig Feedback erteilt werden (Hattie & Timperley, 2007; Voerman et al., 2012; Widorski, Salzmann, Bauder, Heinzer, & Oser, 2012). Aufgrund der hohen Bedeutsamkeit von Feedback im Sportunterricht wird der Fokus im Folgenden beispielhaft auf die abhängige Variable (AV) der *feedbackbezogenen Performanzen* von angehenden Sportlehrkräften der Berufsfachschule gerichtet. Die Qualität der feedbackbezogenen Performanzen von Sportlehrkräften im Sportunterricht wird in vorliegender Studie anhand der latenten Variablen der *Spezifität*, der *Realisierbarkeit* und der *Regelmäßigkeit* eruiert (unterer Bereich).

Ziel, konkrete Fragestellungen und Hypothesen

Obschon die Datenqualität in Selbstbeurteilungsverfahren im Allgemeinen als kritisch zu betrachten ist (Hartig & Klieme, 2006; König & Tachtsoglou, 2012; Kruger & Dunning, 1999; Kunter & Klusmann, 2010; Shavelson, 2013; Terhart, 2007), wurde im Spezifischen die Qualität von Selbstbeurteilungsverfahren bei angehenden Sportlehrkräften nicht erforscht. Diese Forschungslücke erscheint als problematisch, weil die Kompetenzdiagnostik bei angehenden Sportlehrkräften häufig auf Selbstbeurteilungen basiert (z. B. Meier, 2015). Es wird u. a. nicht klar, ob und wie stark die Qualität der eigenen kompetenzbereichsbezogenen Performanzen von angehenden Sportlehrkräften die Selbstbeurteilung beeinflusst. In Anlehnung an die Ergebnisse der Studie von Dunning und Kruger (1999) erscheint es folglich als wünschenswert, den Einfluss der Qualität von kompetenzbereichsbezogenen Performanzen bei angehenden Sportlehrkräften in der Selbstbeurteilung im Vergleich zu einem intersubjektiven Fremdbeurteilungsverfahren (Bezugsnorm) zu erforschen. Die vorliegende Untersuchung soll einen kleinen Beitrag zur Bearbeitung dieser Forschungslücke leisten und zielt darauf ab, Wissen über die Beeinflussung der Qualität von kompetenzbereichsbezogenen Performanzen in Selbstbeurteilungsverfahren bei angehenden Sportlehrkräften zu generieren.

Ausgehend von der beschriebenen Problemlage, der Forschungslücke, des dargelegten Forschungsdesiderats, des Forschungsziels und der Spezifizierung hinsichtlich den feedbackbezogenen Performanzen werden in vorliegender Studie die *konkreten Fragestellungen* bearbeitet, 1) ob *inkompetente* angehende Sportlehrkräfte der Berufsfachschule die Qualität ihrer eigenen feedbackbezogenen Performanzen im Vergleich zu einer intersubjektiven Fremdbeurteilung (Referenzsystem) überschätzen und 2) ob *kompetente* angehende Sportlehrkräfte der Berufsfachschule die Qualität ihrer eigenen feedbackbezogenen Performanzen im Vergleich zu einer intersubjektiven Fremdbeurteilung unterschätzen. In Anlehnung an diese konkreten Fragestellungen und unter Berücksichtigung der vorliegenden Forschungsbefunde von Kruger und Dunning (1999) resultieren für die vorliegende Studie die beiden folgenden *einseitigen Unterschieds- bzw. Alternativhypothesen*:

Hypothese 1 (H₁). *Inkompetente* angehende Sportlehrkräfte der Berufsfachschule überschätzen die Qualität ihrer eigenen feedbackbezogenen Performanzen im Vergleich zu einer intersubjektiven Fremdbeurteilung (starker Effekt).

Hypothese 2 (H₂). *Kompetente* angehende Sportlehrkräfte der Berufsfachschule unterschätzen die Qualität ihrer feedbackbezogenen Performanzen im Vergleich zu einer intersubjektiven Fremdbeurteilung (starker Effekt).

Methode

Studiendesign, Untersuchungsinstrument und Stichprobe

Das Design der empirischen Untersuchung basiert auf zwei Schritten. In einem ersten Schritt wurde in Anlehnung an eine quasi-experimentelle Feldstudie (Pre-Post-Control-Design) von Baumgartner (2016) auf der Grundlage eines intersubjektiven Fremdbeurteilungsverfahrens, wodurch zwei Expertinnen bzw. Experten unabhängig voneinander die feedbackbezogenen Performanzen von angehenden Sportlehrkräften beurteil-

ten, die zwei Untersuchungsgruppen der *Kompetenten* und der *Inkompetenten* gebildet. Der Unterschied der intersubjektiv fremdbeurteilten Qualität der feedbackbezogenen Performanzen zwischen den beiden Untersuchungsgruppen *Kompetente* und *Inkompetente* sollte groß sein. Denn in der Studie von Dunning und Kruger (1999) wird ersichtlich, dass insbesondere die Versuchspersonen aus dem obersten (*Kompetente*) und dem untersten (*Inkompetente*) Leistungsquartil die Qualität ihrer testbezogenen Performanzen deutlich unter- bzw. überschätzten. In einem zweiten Schritt wurden die Versuchspersonen der beiden Untersuchungsgruppen dazu aufgefordert, die Qualität ihrer feedbackbezogenen Performanzen anhand eines Selbstbeurteilungsverfahrens einzuschätzen. Zur Beantwortung der konkreten Fragestellungen bzw. Hypothese wurden die arithmetischen Mittel (M) der Selbstbeurteilung der beiden Untersuchungsgruppen mit denjenigen des intersubjektiven Fremdbeurteilungsverfahrens (Expertinnen- und Expertenrating) verglichen. Das intersubjektive Fremdbeurteilungsverfahren stellte dabei die *objektive* Bezugsnorm der Beurteilung der Qualität der feedbackbezogenen Performanzen dar.

Die abhängige Variable (AV) der Qualität der feedbackbezogenen Performanzen der Versuchspersonen wurde im Selbst- sowie im intersubjektiven Fremdbeurteilungsverfahren anhand des validierten Fremdbeurteilungsinstruments nach Widorski et al. (2012) bzw. Oser et al. (2010) erhoben. Das Originalinstrument wurde vor dem Hintergrund der fachübergreifenden Forschungsbefunde (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008) entwickelt und zielt auf die Messung der Güte des absichtlichen, verbalen sowie gezielten Feedbacks von Lehrkräften der Berufsfachschule an die Schülerinnen und Schüler. Das Instrument erfasst die vier latenten Variablen *Spezifität*, *Lernklima*, *Realisierbarkeit* und *Regelmäßigkeit* (Widorski et al., 2012). Unter Berücksichtigung der vorliegenden fachspezifischen Forschungsbefunde (Lee et al., 1993; Marschall & Dausg, 2003; Silverman et al., 1992) wurde das

Instrument angepasst, um die Qualität der feedbackbezogenen Performanzen von Sportlehrkräften an der Berufsfachschule diagnostizieren zu können (Baumgartner, 2016). So wurde u. a. die latente Variable *Lernklima* ausgeschlossen, weil Lernklima nicht einzig ein Kriterium eines effektiven Feedbacks von Sportlehrkräften darstellt, sondern eher als ein kompetenzbereichsübergreifendes Effektivitätskriterium zu deuten ist (Brophy, 2000). Das adaptierte Fremdbeurteilungsinstrument konstituiert sich folglich aus den folgenden drei latenten Variablen:

- a) Spezifität (Itembeispiel: „In der Instruktionsphase wird klar, welche Kriterien einer Aufgabe/Übung in Beziehung zum Ziel als bedeutsam gelten“; 5 Items; jeweils vierstufige Likert-Skala von 0 = „trifft nicht zu“ bis 3 = „trifft zu“; Baumgartner, 2016; Widorski et al., 2012; Tab. 1),
- b) Realisierbarkeit (3 Items; Itembeispiel: „Das Feedback der Sportlehrkraft kann von den Lernenden mit einem angemessenen Aufwand umgesetzt werden“),
- c) Regelmäßigkeit (2 Items; Itembeispiel: „Die Sportlehrkraft gibt den Lernenden zu wenig Feedback“).

Bezüglich der Güte der Daten des Fremdbeurteilungsinstruments wies die konfirmatorische Faktorenanalyse zweiter Ordnung in der Studie von Baumgartner (2016) einen ausreichenden bis guten Modell-Fit auf ($\chi^2 = 44,57$; $df = 32$; $p = 0,07$; $\chi^2 / df = 1,39$; SRMR = 0,04; CFI = 0,98; RMSEA=0,07; $n = 70$). Die Zuverlässigkeit des Konstrukts ist als sehr hoch zu bezeichnen ($\alpha = 0,92$).

Die Stichprobe der vorliegenden Untersuchung bestand aus angehenden Sportlehrkräften der Berufsfachschule, welche an der Eidg. Hochschule für Sport Magglingen (CH) in ihrem dreijährigen Studium das zweite Ausbildungsjahr absolvierten. Die Extrahierung der beiden Untersuchungsgruppen wurde in Anlehnung an die Interventionsstudie von Baumgartner (2016) vorgenommen. Die Versuchspersonen in der Untersuchungsgruppe *Kompetente* ($n = 13$; 2 weiblich) haben vor der Durchführung der vorliegenden Studie an ei-

nem theorie- und praxisverknüpfenden Interventionsprogramm zur gezielten Verbesserung der feedbackbezogenen Performanzen partizipiert (Baumgartner, 2016). Im Anschluss an das Interventionsprogramm haben sich die Versuchspersonen freiwillig dazu bereit erklärt, an der vorliegenden Studie teilzunehmen. Das durchschnittliche Alter lag in der Untersuchungsgruppe *Kompetente* bei 24,91 Jahren ($SD = 2,18$). In Abgrenzung zur Untersuchungsgruppe *Kompetente* wurde in der Untersuchungsgruppe *Inkompetente* ($n = 10$; 3 weiblich) keine gezielte Intervention zur Verbesserung der feedbackbezogenen Performanzen durchgeführt. Die Versuchspersonen absolvierten lediglich ein allgemeines Schulpraktikum (Standardintervention; Baumgartner, 2016). Auch diese Versuchspersonen haben sich freiwillig dazu bereit erklärt, an der vorliegenden Studie teilzunehmen. Das durchschnittliche Alter lag in der Untersuchungsgruppe *Inkompetente* bei 24,28 Jahren ($SD = 1,23$).

Untersuchungsdurchführung

Die Qualität der feedbackbezogenen Performanzen der Versuchspersonen aus beiden Untersuchungsgruppen wurde in der Studie von Baumgartner (2016) nach dem Interventionsprogramm (Untersuchungsgruppe *Kompetente*) resp. dem allgemeinen Schulpraktikum (Untersuchungsgruppe *Inkompetente*) in einer Praxislektion an der Berufsfachschule festgehalten. Die Bestimmung der Inhalte der Lektionen fand zufallsbasiert durch die Praxislehrkräfte statt. Die Praxislektion wurden visuell durch eine Klassenkamera sowie eine Interaktionskamera aufgezeichnet (*Kameramodell Sony HDR-CX 700VE*). Die verbale Interaktion zwischen den Versuchspersonen und den Schülerinnen und Schüler wurde mittels Funkmikrofon erhoben (*Modell Sennheiser EK 100 G3*). Im Anschluss wurden die Daten synchronisiert und aus dem Datenmaterial die spezifische feedbackbezogene Unterrichtssequenz herausgeschnitten. Durch diesen Prozess wurden feedbackbezogene Videovignetten generiert, womit die Qualität der feedbackbezogenen Perfor-

manzen der Versuchspersonen möglichst authentisch widerspiegelt werden sollte (Baumgartner, 2016).

Zur Bewertung der Qualität der feedbackbezogenen Performanzen der Versuchspersonen anhand der generierten Videovignetten wurde sodann ein intersubjektives Fremdbeurteilungsverfahren (Expertinnen- und Expertenrating) realisiert. Die teilnehmenden Expertinnen und Experten ($n = 6$; 2 weiblich) verfügten über ein Diplom als Sportlehrkraft, unterrichteten seit mehreren Jahren ($M = 14,09$; $SD = 10,91$) an der Berufsfachschule das Fach Sport und betreuten angehende Sportlehrkräfte der Eidg. Hochschule für Sport Magglingen in ihren Schulpraktika an der Berufsfachschule. Im intersubjektiven Fremdbeurteilungsverfahren bewerteten jeweils zwei Expertinnen und Experten in Anlehnung an das Diagnoseinstrument unabhängig voneinander computerbasiert die feedbackbezogenen Performanzen der Versuchspersonen anhand der vorliegenden Videovignetten („double scoring“; Baumgartner, 2016). Für die inferenzstatistische Datenanalyse wurde sodann das arithmetische Mittel (M) der beiden Expertinnen- und Expertenurteile verwendet. Die Zuordnung der Expertinnen und Experten zu den Versuchspersonen fand zufallsbasiert statt. Um im Fremdbeurteilungsverfahren eine möglichst hohe absolute Übereinstimmung und eine hohe Interraterreliabilität der Expertinnen und Experten gewährleisten zu können, wurde ein Raterinnen- und Ratertraining durchgeführt. Dazu haben sich die Expertinnen und Experten vor der Durchführung des Ratings in das Diagnoseinstrument eingearbeitet. Vor dem Rating fand eine Einführung in das Fremdbeurteilungsinstrument statt. Die einzelnen latenten sowie manifesten Variablen wurden im Raterinnen- und Ratertraining erläutert und besprochen. Im Anschluss bewerteten die Expertinnen und Experten mittels Diagnoseinstrument in verschiedenen Tandemgruppen während drei Stunden die Qualität der feedbackbezogenen Performanzen von angehenden Sportlehrkräften anhand von feedbackbezogenen Videovignetten (Übungsdateien). Nach diesem Raterinnen- und Ratertraining

begann die Datenerhebungsphase. Dabei waren die Expertinnen und Experten nicht darüber informiert, welche Interventionsmaßnahmen realisiert wurden bzw. welcher Untersuchungsgruppe eine Versuchsperson angegliedert war. Durch diese Maßnahme sollten Erwartungseffekte der Expertinnen und Experten ausgeschlossen werden (Baumgartner, 2016).

Im Selbstbeurteilungsverfahren wurden die Versuchspersonen in das Diagnoseinstrument eingeführt und Unklarheiten besprochen. Im Anschluss wurde die eigene feedbackbezogene Videovignette abgespielt, und die Versuchspersonen bewerteten anhand des Diagnoseinstruments die Qualität ihrer eigenen feedbackbezogenen Performanzen.

Statistische Analysen

Um die Datenqualität des intersubjektiven Fremdbeurteilungsverfahrens quantifizieren zu können, wurde erstens die *absolute Übereinstimmung* der Raterinnen- und Raterurteile berechnet. Dazu wurde die prozentuale Übereinstimmung (PÜ) sowie das zufallskorrigierte Übereinstimmungsmaß Cohens Kappa (κ) gerechnet. Zweitens wurde die Güte der *Raterreliabilität* anhand des Reliabilitätskoeffizienten ICC („intra-class correlation coefficient“) eruiert. Weil nicht alle Expertinnen und Experten jede Versuchsperson bewertete und die Zuordnung Expertinnen und Experten und Versuchsperson zufallsbasiert stattfand, wurde dazu die $ICC_{unjustiert, einf., random}$ eruiert (Wirtz & Caspar, 2002). Die Quantifizierung der Datenqualität erfolgte über alle Raterinnen- und Raterurteile und über alle manifesten Variablen.³

³ Cicchetti (1994) empfiehlt folgende Richtwerte für den ICC-Koeffizienten: gering $<0,40$; ausreichend = $0,40-0,59$; gut = $0,60-0,74$; exzellent = $0,75-1,00$. Landis und Koch (1977) definieren folgende Güteausprägungen des Cohens Kappa: schlecht $<0,00$; gering = $0,00-0,20$; ausreichend = $0,21-0,40$; moderat = $0,41-0,60$; beachtlich = $0,61-0,80$; nahezu perfekt = $0,81-1,00$. Der p-Wert der Raterinnen- und Raterurteile sollte bei einer vierstufigen Rating-Skala deutlich über der zufälligen Übereinstimmung von 25 % liegen (Wirtz & Caspar, 2002).

Zur inferenzstatistischen Datenanalyse wurden t-Tests sowie zweifaktorielle Varianzanalysen mit Messwiederholung (gemischtes Design) gerechnet. Um die Effektstärke der gefundenen Unterschiede quantifizieren zu können, wurden bei t-Tests der Effektstärkekoeffizient Cohens d und bezüglich der zweifaktoriellen Varianzanalyse der Populationseffektschätzer $\hat{\omega}^2$ (Omega Quadrat) gerechnet. Im Gegensatz zum partiellen Eta-Quadrat (η^2) wird in Omega Quadrat ($\hat{\omega}^2$) die Anzahl der Versuchspersonen in der Berechnung der Varianzaufklärung mitberücksichtigt, wodurch hinsichtlich der aufgeklärten Varianz mit geringeren Verzerrungen zu rechnen ist.

Ergebnisse

Güte des intersubjektiven Fremdbeurteilungsverfahrens und Extrahierung der Gruppen

Im intersubjektiven Fremdbeurteilungsverfahren ist die *absolute Übereinstimmung* bzw. die *Interraterreliabilität* der Expertinnen und Experten als gut bis exzellent zu bezeichnen ($P_M = 0,63$; $\kappa_M = 0,44$; $ICC_M = 0,83$). Vor dem Hintergrund der hohen Güte des Fremdbeurteilungsverfahrens konnten die Daten für die weiteren inferenzstatistischen Analysen verwendet werden. Aus den Forschungsbefunden wird ersichtlich, dass die (intersubjektiv fremdbestimmte) Qualität der feedbackbezogenen Performanzen der Versuchspersonen in der Untersuchungsgruppe der *Kompetenten* ($M = 2,16$; $SD = 0,51$) im Vergleich zur Untersuchungsgruppe der *Inkompetenten* ($M = 0,74$; $SD = 0,38$) hoch signifikant besser eingestuft wurden. Der Effekt des gefundenen Unterschieds der Qualität der feedbackbezogenen Performanzen zwischen den beiden Untersuchungsgruppen ist stark ($t(21) = 7,36$; $p < 0,001$; $1 - \beta = 1,00$; $d = 3,16$). Die Extrahierung der beiden Untersuchungsgruppen der *Kompetenten* und der *Inkompetenten* konnte somit inferenzstatistisch nachgewiesen werden, wodurch die Voraussetzung zu weiteren Berechnungen gegeben war.

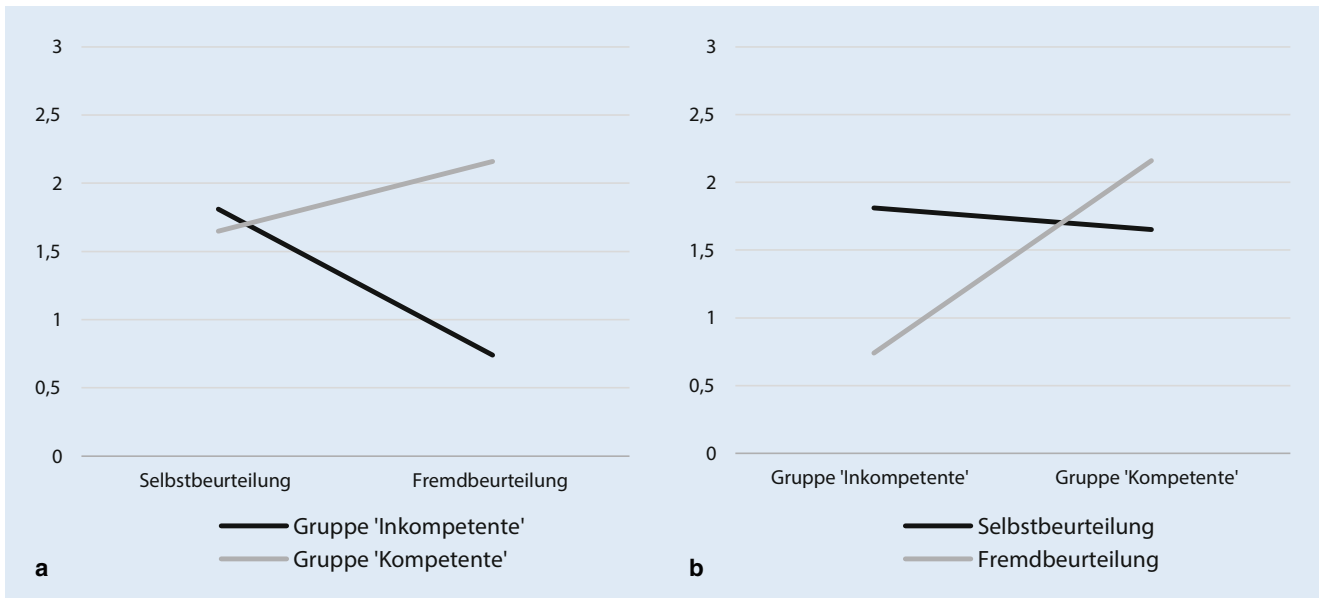


Abb. 1 ▲ Interaktionsdiagramme der Hauptfaktoren Beurteilungsverfahren und Untersuchungsgruppe bezüglich der feedbackbezogenen Performanz (AV)

Übereinstimmung der Selbst- und intersubjektiven Fremdbeurteilung

In Bezug zur Selbstbeurteilung der Versuchspersonen zeigt sich, dass die Untersuchungsgruppe der *Inkompetenten* die Qualität ihrer eigenen feedbackbezogenen Performanzen ($M = 1,81$; $SD = 0,30$) im Vergleich zur Fremdbeurteilung ($M = 0,74$; $SD = 0,38$) hoch signifikant besser einstufen. Der Effekt des gefundenen Unterschieds zwischen Selbst- und Fremdbeurteilung ist groß [$t(18) = 6,98$; $p = 0,001$; $1 - \beta = 1,00$; $d = 3,13$]. Im Vergleich dazu beurteilte die Untersuchungsgruppe der *Kompetenten* die Qualität ihrer feedbackbezogenen Performanzen ($M = 1,65$; $SD = 0,62$) im Vergleich zur Fremdbeurteilung ($M = 2,16$; $SD = 0,51$) signifikant schlechter. Der Effekt des gefundenen Unterschieds ist stark ($t(24) = 2,31$; $p = 0,011$; $1 - \beta = 0,90$; $d = 1,32$). Zur Visualisierung dieser Effekte wurde die zweifaktorielle Varianzanalyse mit Messwiederholung (gemischtes Design) gerechnet. Zwischen den beiden Hauptfaktoren *Beurteilungsverfahren* und *Untersuchungsgruppe* wird erwartungsgemäß ein hoch signifikanter Interaktionseffekt und ein starker Effekt sichtbar ($F(1;21) = 40,39$; $p = 0,001$; $1 - \beta = 1,00$; $\hat{\omega}^2 = 0,30$). Zwischen den

beiden Hauptfaktoren *Beurteilungsverfahren* und *Untersuchungsgruppe* besteht ein disordinaler Interaktionseffekt, d. h. keiner der beiden Haupteffekte kann generell interpretiert werden (■ **Abb. 1**). Infolgedessen unterschätzt sich die Untersuchungsgruppe der *Kompetenten* bzw. überschätzt sich die Untersuchungsgruppe der *Inkompetenten* im Selbstbeurteilungsverfahren hinsichtlich der intersubjektiv fremdbeurteilten Qualität der feedbackbezogenen Performanzen.

Diskussion und Fazit

Das Ziel der vorliegenden Studie bestand in der Generierung von Wissen hinsichtlich des Einflusses der Qualität der eigenen kompetenzbereichsbezogenen Performanzen in Selbstbeurteilungsverfahren bei angehenden Sportlehrkräften der Berufsfachschule. Unter Berücksichtigung der Ergebnisse der Studie von Kruger und Dunning (1999) wurde untersucht, 1) ob *inkompetente* angehende Sportlehrkräfte der Berufsfachschule die Qualität ihrer eigenen feedbackbezogenen Performanzen im Vergleich zu einer intersubjektiven Fremdbeurteilung überschätzen und 2) ob *kompetente* angehende Sportlehrkräfte der Berufsfachschule die Qualität ihrer eigenen feedbackbezogenen Performanzen

im Vergleich zu einer intersubjektiven Fremdbeurteilung unterschätzen.

Aus den datenbasierten Erkenntnissen wurde deutlich, dass die intersubjektiv fremdbeurteilte Qualität der feedbackbezogenen Performanzen in der Untersuchungsgruppe der *Kompetenten* deutlich höher ist als in der Untersuchungsgruppe der *Inkompetenten*. Die Extrahierung der beiden Untersuchungsgruppen der *Kompetenten* und der *Inkompetenten* konnte damit empirisch bestätigt werden, wodurch weitere Berechnungen zur Beantwortung der Hypothesen durchgeführt werden konnten. Es wurde ersichtlich, dass die Untersuchungsgruppe der *Inkompetenten* die Qualität ihrer feedbackbezogenen Performanzen im Vergleich zur intersubjektiven Fremdbeurteilung der Expertinnen und Experten deutlich überschätzte. *Die generierte Alternativhypothese H_1 wird damit nicht zurückgewiesen*. In Abgrenzung dazu unterschätzte die Untersuchungsgruppe der *Kompetenten* die Qualität ihrer feedbackbezogenen Performanzen in Relation zur intersubjektiven Fremdbeurteilung bedeutsam. *Folglich wird auch die Alternativhypothese H_2 nicht verworfen*.

In der vorliegenden Studie wird dementsprechend ersichtlich, dass die Datenqualität in kompetenzdiagnostischen

Selbstbeurteilungsverfahren von angehenden Sportlehrkräften zu hinterfragen ist. Es wird ersichtlich, dass die Qualität der feedbackbezogenen Performanzen bei den Untersuchungsteilnehmenden einen großen Einfluss auf die Selbstbeurteilung hat. Dieses Ergebnis korrespondiert mit den Forschungsbefunden von Kruger und Dunning (1999). Es weicht jedoch von den Nachweisen ab, dass angehende Lehrkräfte die Qualität ihrer eigenen kompetenzbereichsbezogenen Performanzen im Vergleich zur (subjektiven) Fremdbeurteilung der Praxislehrkräfte kritischer (Bodensohn & Schneider, 2009; Boekhoff et al., 2008) oder nicht unterschiedlich bewerten (Moser & Hascher, 2000). Eine derartige Verallgemeinerung konnte in der vorliegenden Studie nicht nachgewiesen werden. Denn es wurde deutlich, dass die Selbstbeurteilung bedeutsam von der Qualität der eigenen feedbackbezogenen Performanzen der angehenden Sportkräfte abhängig ist.

Die Generalisierbarkeit der Erkenntnisse der vorliegenden Studie wird durch verschiedene Faktoren eingegrenzt. So wurde die Untersuchung an einem Hochschulstandort und bei einem (geringen) Stichprobenumfang von 23 angehenden Sportlehrkräften durchgeführt. Zudem wurde der Fokus einzig auf die feedbackbezogenen Performanzen gerichtet. Dementsprechend können die Resultate nicht direkt auf andere Hochschulstandorte, andere Fächer und andere kompetenzbereichsbezogene Performanzen übertragen werden. Weitere Forschungsbemühungen erscheinen hierzu wünschenswert.

Das generierte Wissen über den Einfluss der Qualität der feedbackbezogenen Performanzen in der Selbstbeurteilung von angehenden Sportlehrkräften ist sowohl für die Kompetenzdiagnostik als auch für die Ausbildung von Sportlehrkräften von Nutzen. Die Ergebnisse weisen bezüglich der *Kompetenzdiagnostik* darauf hin, dass die Qualität der kompetenzbereichsbezogenen Performanzen die Selbstbeurteilung verzerren kann. Folglich erscheint es wünschenswert, dass in kompetenzdiagnostischen Verfahren zur Sicherung einer hohen Datengüte objektive und intersubjektiv

ve Vorgehensweisen (z. B. testbezogene Kompetenzdiagnostik in Anforderungssituationen, Fremdbeurteilungsverfahren) eingesetzt werden. Aus ökologischen Validitätszwecken bedarf die Kompetenzdiagnostik dabei der Integration der performativen Ebene. Dadurch kann der Kernauftrag des Sportlehrberufs – die Förderung der Entwicklung der Schülerinnen und Schüler in der Komplexität und Ungewissheit von Sportunterricht (Doyle, 2006) – in der Kompetenzdiagnostik berücksichtigt werden. Denn wenn eruiert werden soll, was angehende (Sport-)Lehrkräfte zu leisten vermögen, so sollte die Kompetenzdiagnostik verhaltensnah, domänenspezifisch und in enger Relation zu den konkreten Schulleistungen realisiert werden (Baumgartner, 2016; Seidel & Shavelson, 2007). Es wird damit deutlich, dass eine solide und ökologisch valide Kompetenzdiagnostik als aufwändig zu interpretieren ist (Shavelson, 2013) und dabei das Handeln in der situativen Komplexität von Unterricht berücksichtigt werden sollte.

Auf die *Ausbildung von Sportlehrkräften* bezogen weisen die generierten Erkenntnisse darauf hin, dass die Lehrerinnen- und Lehrerausbildenden einen bewussten Umgang mit dem dargelegten Verzerrungseffekt kultivieren sollten. Es bedarf der Generierung von Lehr-Lernarrangements, mit Hilfe derer die beiden Lernorte des hochschulischen Unterrichts und der eigenen Unterrichtspraxis durch methodische Elemente (z. B. eigene Unterrichtsvideos) gezielt verknüpft werden. Dadurch sollen die kompetenzbereichsbezogenen Performanzen der angehenden Sportlehrkräfte sichtbar werden, wodurch ein Diskurs über die selbst- und fremdbestimmte Qualität der eigenen Performanzen ermöglicht wird. Denn soll die Ausbildung die angehenden Sportlehrkräfte authentisch auf die spätere Berufstätigkeit vorbereiten, so sollten sie wissen, was sie (noch nicht) können.

Korrespondenzadresse



M. Baumgartner, M.A.
 Eidg. Hochschule für Sport
 Magglingen
 Hauptstraße 247,
 2532 Magglingen, Schweiz
 matthias_baumgartner79@
 gmx.ch

Danksagung. Ein großer Dank richtet sich an dieser Stelle an Dimitri Vogel, der im Rahmen des Selbstbeurteilungsverfahrens die Daten erhoben hat.

Interessenkonflikt. M. Baumgartner gibt an, dass kein Interessenkonflikt besteht.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) 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.

Literatur

- Abs, H. J. (2007). Überlegungen zur Modellierung diagnostischer Kompetenz bei Lehrerinnen und Lehrern. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung: Kompetenzentwicklung und Programmevaluation* (S. 63–84). Münster: Waxmann.
- Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Baumgartner, M. (2013a). Kompetenzprofile von Sportlehrpersonen der Berufsfachschule. In F. Oser, T. Bauder, P. Salzmann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzprofilen bei Lehrpersonen und Berufsverantwortlichen* (S. 96–126). Bad Heilbrunn: Klinkhardt.
- Baumgartner, M. (2013b). Professionelle Kompetenzprofile von Sportlehrpersonen. Zur empirischen Generierung eines kompetenzorientierten Referenzsystems für die Ausbildung von Sportlehrpersonen der Berufsschule. *Spectrum der Sportwissenschaften*, 25(2), 49–74.
- Baumgartner, M. (2016). Performanzentwicklung in der Lehrerinnen- und Lehrerausbildung. Unveröffentlichte Dissertation, Universität Flensburg.
- Blomberg, G., Seidel, T., & Prenzel, M. (2011). Neue Entwicklungen in der Erfassung pädagogisch-psychologischer Kompetenzen von Lehrpersonen. *Unterrichtswissenschaft*, 39(2), 98–101.
- Blömeke, S., Felbrich, A., & Müller, C. (2008). Theoretischer Rahmen und Untersuchungsdesign. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und*

- referendare. *Erste Ergebnisse zur Wirksamkeit der Lehrerbildung* (S. 15–48). Münster: Waxmann.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3–13.
- Blömeke, S., König, J., Suhli, U., Hoth, J., & Döhrmann, M. (2015). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift für Pädagogik*, 61(3), 310–327.
- Bodensohn, R., & Schneider, C. (2009). Vier Studien zu Blockpraktika als Ausgangsbasis für die Entwicklung standardbezogener Evaluation. In R. Bolle & M. M. Rotermund (Hrsg.), *Schulpraktische Studiengänge in gestuften Studiengängen* BaSS. Schriftenreihe der Bundesarbeitsgemeinschaft Schulpraktischer Studien. (S. 206–238). Leipzig: Universitätsverlag.
- Boekhoff, I., Franke, K., Dietrich, F., & Arnold, K.-H. (2008). *Effektivität der universitären Lehrerbildung an konsekutiven Studiengängen (EduLiks) unter besonderer Berücksichtigung schulpraktischer Studien*. Hildesheim: Universität Hildesheim.
- Brophy, J. E. (2000). *Teaching*. Brüssel: International Academy of Education.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.
- Czerwenka, K., & Nölle, K. (2014). Forschung zur ersten Phase der Lehrerbildung. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrberuf* (2. Aufl. S. 468–488). Münster: Waxmann.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1–44.
- Doyle, W. (2006). Ecological approaches to classroom management. Process-outcome approaches to classroom management and effective teaching. In C. Evertson & C. Weinstein (Hrsg.), *Handbook of classroom management: research, practice, and contemporary issues* (S. 97–125). Mahwah: Erlbaum.
- Forster-Heinzer, S., & Oser, F. (2015). Wer setzt das Mass? Eine kritische Auseinandersetzung mit dem Advokatorischen Ansatz. *Zeitschrift für Pädagogik*, 61(3), 361–376.
- Frey, A. (2006). *Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften – eine erste Standortbestimmung zu bereits publizierten Instrumenten*. Zeitschrift für Pädagogik, 51. Beiheft. (S. 30–46).
- Hager, W., & Hasselhorn, M. (2000). Psychologische Interventionsmassnahmen: Was sollen sie bewirken können? In W. Hager, J. L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmassnahmen* (S. 41–85). Bern: Huber.
- Hartig, J., & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17–36). Berlin: Bundesministerium für Bildung und Forschung.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Berlin: Springer.
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Heinzer, S., & Baumgartner, M. (2013). Bottom-up zur Qualität: Eine Quasi-Delphi-Studie zur Generierung von Kompetenzprofilen. In F. Oser, T. Bauder, P. Salzmann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzprofilen bei Lehrpersonen und Berufsbildungsverantwortlichen* (S. 66–95). Bad Heilbrunn: Klinkhardt.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- König, J., & Tachtsoglou, S. (2012). Pädagogisches Professionswissen und selbsteingeschätzte Kompetenz. In J. König & A. Seifert (Hrsg.), *Lehr- amtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung* (S. 284–297). Münster: Waxmann.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
- Kunter, M., & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen. *Unterrichtswissenschaft*, 38(1), 68–86.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften – Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lee, A., Keh, N., & Magill, R. (1993). Instructional effects of teacher feedback in physical education. *Journal of Teaching in Physical Education*, 12(3), 228–243.
- Magill, R. A. (1993). Augmented feedback in skill acquisition. In R. N. Singer, M. Murphey & L. K. Tennant (Hrsg.), *Handbook on Research in Sport Psychology* (S. 193–212). New York: Macmillan.
- Marschall, F., & Dausgs, R. (2003). Feedback. In H. Mechling & J. Munzert (Hrsg.), *Handbuch Bewegungswissenschaft – Bewegungslehre* (S. 281–294). Schorndorf: Hofmann.
- McClelland, D. (1973). Testing for competence rather than for 'intelligence'. *American Psychologist*, 28(1), 1–14.
- Meier, S. (2015). *Kompetenzen von Lehrkräften. Eine empirische Studie zur Entwicklung fachübergreifender Kompetenzeinschätzungen*. Münster: Waxmann.
- Moser, P., & Hascher, T. (2000). *Lernen im Praktikum (Projektbericht)*. Bern: Universität Bern.
- Orton, P. Z. (1993). *Cliffs law school admission test preparation guide*. Lincoln: Cliff Notes Incorporated.
- Oser, F. (2001). Standards: Kompetenzen von Lehrpersonen. In F. Oser & J. Oelkers (Hrsg.), *Die Wirksamkeit der Lehrerbildungssysteme. Von der Allrounderbildung zur Ausbildung professioneller Standards* (S. 215–342). Zürich: Rüegger.
- Oser, F., Heinzer, S., Bauder, T., Grüter, A., Salzmann, P., Joho, C., Critelli, N., Baumgartner, M., & Wittwer, V. (2010). *Leading House Qualität der beruflichen Bildung (Projektbericht)*. Fribourg: Universität Fribourg.
- Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford: Elsevier.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
- Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, 2(1), 41–63.
- Shavelson, R. J. (2013). On an approach to testing and modeling competence. *Educational Psychologist*, 48(2), 73–86.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- Silverman, S., Tyson, L. A., & Krampitz, J. (1992). Teacher feedback and achievement in physical education: interaction with student practice. *Teaching and Teacher Education*, 8(4), 333–344.
- Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz*. Münster: Universität Münster.
- Terhart, E. (2007). Erfassung und Beurteilung der beruflichen Kompetenz von Lehrkräften. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 37–62). Münster: Waxmann.
- Voerman, L., Meijer, P. C., Korthagen, F. A., & Simons, P. R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and Teacher Education*, 28(8), 1107–1115.
- Vonken, M. (2005). *Handlung und Kompetenz. Theoretische Perspektiven für die Erwachsenen- und Berufspädagogik*. Wiesbaden: VS.
- Widorski, D., Salzmann, P., Bauder, T., Heinzer, S., & Oser, F. (2012). *Lernenden fördernd Feedback geben. Ein interaktives Arbeitsinstrument für Berufsbildungsverantwortliche und Lehrpersonen*. Bern: Hep.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.