

Berichte aus der Psychologie

**Michael Krämer, Siegfried Preiser,
Kerstin Brusdeylins (Hrsg.)**

Psychologiedidaktik und Evaluation XI

Materialien aus der Sektion Aus-, Fort- und Weiterbildung in Psychologie (AFW)
im Berufsverband Deutscher Psychologinnen und Psychologen e.V. (BDP)
Band 14

Dem Wunsch mehrerer Autorinnen und Autoren folgend sei an dieser Stelle darauf hingewiesen, wenn in den folgenden Beiträgen entweder nur die männliche oder nur die weibliche Sprachform genutzt wird, so sind jeweils beide Geschlechter gemeint. Es geschieht ausschließlich der besseren Lesbarkeit halber.

Um den Datenschutz zu wahren, wurde auf die Veröffentlichung der Autorenadressen verzichtet. Wenn Sie Kontakt zu einer Autorin oder einem Autor aufnehmen wollen, schreiben Sie bitte eine e-mail an folgende Adresse. Der Herausgeber leitet Ihren Wunsch gerne weiter: kraemer@fh-muenster.de

Inhalt

Aus-, Fort- und Weiterbildung in Psychologie

MICHAEL KRÄMER

Novellierung des Psychotherapeutengesetzes – Stand der Dinge und Auswirkungen auf das Studium der Psychologie und die Profession 3

SIEGFRIED PREISER UND MICHAEL GIEBEL

Weiterbildungsmotivation von Studierenden der Psychologie 2012 - 2014 - 2016 11

LARS BEHRMANN

„Je höher der Frauenanteil eines Fachs, desto weniger Frauen promovieren ...“ 17

MIRIAM THYE, KATHARINA MOSEN, ULRICH WEGER UND DIETHARD TAUSCHEL

Meditation und akademische Prokrastination – eine qualitative Studie 25

HANS-PETER LANGFELDT

Kulturelle Grenzen der (Psychologie)Didaktik – Erfahrungen an einer äthiopischen Universität 35

CARL P. A. KESSELER, STEFAN TROCHE UND MICHAELA ZUPANIC

Zur Konsistenz der Erwartungen an die ideale Persönlichkeit von Psychologie-Studierenden und die Auswirkungen auf Studienzufriedenheit und Studienleistung 43

CHARLOTTE VEHOFF, MICHAELA ZUPANIC, ROBIN JÖRN SIEGEL UND STEFAN TROCHE

Die Motivationsquellen studentischer GutachterInnen im Auswahlverfahren Psychologie: Neugier, Engagement oder doch das Bedürfnis nach Macht? 53

STEPHAN DUTKE, LENA KOEPCKE UND ELMAR SOUVIGNIER

Beiträge der Psychologie zum Praxissemester in der Lehramtsausbildung 61

Lehren und Lernen

HANS-PETER NOLTING

Einführung in die Psychologie: themenbezogen oder systembezogen? Das Konzept der Integrativen Didaktik 73

MARLENE WAGNER, STEPHANIE MOSER, INES DEIBL UND JÖRG ZUMBACH

Psychologiedidaktik trifft Philosophiedidaktik: Der Einsatz von Wikis im interdisziplinären Unterricht 81

SIEGFRIED PREISER UND TAMARA TURASHVILI Wissenschaftskommunikation und Experten-Laien-Kommunikation: Kompetenzerwerb durch Trainingsseminare an deutschen und georgischen Universitäten	93
JULIA MENDZHERITSKAYA UND CAROLINE SCHERER Herausforderungen, Maßnahmen und Verbesserungsindikatoren guter Methodenlehre	101
LARS BEHRMANN UND STEFANIE VAN OPHUYSEN „Forschendes Lernen“ lernen – Die Methodenausbildung für Lehramtstudierende an der WWU Münster	109
DAGMAR TREUTNER Optimierung eines Kommunikationsseminars mit Videofeedback anhand von Videoannotation	119
BASTIAN HODAPP Medienbasiertes Forschendes Lernen – ein Modellprojekt	127
INGO JUNGCLAUSSEN UND SILVIANA STUBIG „Fack ju Pädda!?“ – Neue Wege in der Didaktik der Pädagogischen Psychologie. Ergebnisse einer online-Umfrage zum Einsatz von Spielfilmen in der universitären Lehramtsausbildung am Beispiel der Schulkomödie „Fack ju Göhte“	135
INGO JUNGCLAUSSEN Die ‚Psychodynamik-Animation‘ – Ein mediengestützter Beitrag zur Didaktik der Psychoanalyse	145
MARKUS KNÖPFEL, FRANK MUSOLESI UND WILLI NEUTHINGER Konzeption eines PBL-Moduls im Rahmen des Psychologiestudiums	157
NICOLA BUCHHOLZ UND SUSANNE HILDEBRAND Selbstorganisiertes Lernen im Psychologieunterricht	167
NINA ZEUCH UND ELMAR SOUVIGNIER Wissenschaftliches Denken bei Lehramts- und Psychologiestudierenden	175
TOM ROSMAN, ANNE-KATHRIN MAYER UND GÜNTER KRAMPEN Die Förderung differenzierter epistemologischer Überzeugungen bei Studienanfängern der Psychologie: Empirische Befunde und fachdidaktische Implikationen	185

ANNE-KATHRIN MAYER, NIKOLAS LEICHNER UND GÜNTER KRAMPEN Förderung fachlicher Informationskompetenz von Psychologie- Studierenden durch ein curricular integriertes Blended Learning-Training	193
KATRIN B. KLINGSIECK, DANIEL AL-KABBANI, CARLA BOHDICK, JOHANNA HILKENMEIER, SEBASTIAN KÖNIG, HANNA S. MÜSCHE, SASKIA PRAETORIUS UND SABRINA SOMMER Gamebasiertes Lernen in der Lehrerbildung – spielend zur diagnostisch kompetenten Lehrkraft werden	203
NATHASHA BODONYI, VIKTORIA FALKENHORST UND ULRIKE STARKER, Planspiel – Papiersternmanufaktur	213
MIRIAM THYE, DÉsirÉE RITZKA, ROSE LINK UND DIETHARD TAUSCHEL Lernst du schon oder liest du noch? – Zu der Frage, wie man als Student das akademische Lernen lernen kann	219

Psychologie an Schulen

PAUL GEORG GEIß Kompetenzorientierter Psychologieunterricht in Österreich	229
DOMINIK MOMBELLI Kompetenzorientierter Pädagogik- und Psychologieunterricht aus der gymnasialen Oberstufe in der Schweiz	239
JÜRGEN MALACH UND MARGRET PETERS Von der Input- zur Outputorientierung – Intention, Struktur und Implementation des kompetenzorientierten Kernlehrplans Psychologie für die gymnasiale Oberstufe NRW	247

Praxisbezogene Anwendung psychologischer Erkenntnisse

KERSTIN BRUSDEYLINS UND JORINTHE HAGNER PENTApus Stressprävention für Oberstufenschüler und Studierende	257
TIMO BERSE Krank vor Sorgen – Ein Workshop zur klinischen Psychologie für Psychologielehrer/innen	263

SABINE FABRIZ, LUKAS SCHULZE-VORBERG UND HOLGER HORZ „Beratung und Betreuung von Studierenden im Studium“. Konzeption & Evaluation einer Schulungsreihe für schulische Betreuer/innen und Praktikumsbeauftragte im Praxissemester	271
--	-----

Evaluation

MICHAEL KRÄMER Zufrieden und glücklich?! Zum Zusammenhang zwischen Studienzufriedenheit und Glücksempfinden	281
---	-----

ARNOLD HINZ Wie nützlich ist Lehrevaluation durch Studierende mittels Ratingskalen?	291
--	-----

DANIELA FEISTAUER UND TOBIAS RICHTER Wie zuverlässig sind studentische Einschätzungen der Lehrqualität? Eine Analyse mit kreuzklassifizierten Mehrebenenmodellen	299
--	-----

ELISABETH DALLÜGE, MICHAELA ZUPANIC, CORNELIA HETFELD UND MARZELLUS HOFMANN Wie bildet sich das Curriculum des Studiums im Progress Test Psychologie (PTP) ab?	307
---	-----

MICHAELA ZUPANIC, JAN P. EHLERS, THOMAS OSTERMANN UND MARZELLUS HOFMANN Progress Test Psychologie (PTP) und Wissensentwicklung im Studienverlauf	315
---	-----

JONATHAN BARENBERG, EVA SEIFRIED, BIRGIT SPINATH UND STEPHAN DUTKE Die Bearbeitung schriftlicher Problemaufgaben erhöht den Lernerfolg in einer Psychologie-Vorlesung	323
--	-----

JULIANE SCHWIEREN, JONATHAN BARENBERG UND STEPHAN DUTKE Testeffekt in Psychologie-Lehrveranstaltungen? Eine metaanalytische Perspektive	331
---	-----

DOROTHEA KRAMPEN, KARL SCHWEIZER, SIEGBERT REIß UND ANDREAS GOLD Erprobung einer Kurzsкала zur Erfassung von Impulsivität	339
---	-----

NIKOLAI ZINKE, STEFAN STÜRMER UND LAURA FROEHLICH Validierung einer deutschsprachigen Skala zur multidimensionalen Erfassung von interkulturellen Kompetenzen in der universitären Ausbildung	349
---	-----

Wie zuverlässig sind studentische Einschätzungen der Lehrqualität?

Eine Analyse mit kreuzklassifizierten Mehrebenenmodellen

Daniela Feistauer und Tobias Richter

An vielen Hochschulen wird Lehrqualität mittels studentischer Evaluationsfragebögen erfasst. Diese Studie untersuchte die Reliabilität globaler Einschätzungen von Lehrkräften und Lehrveranstaltungen durch einen Vergleich der Varianzanteile, die durch Lehrkräfte, Lehrveranstaltungen und Studierenden erklärt wurden. Studierendenurteile können als umso reliabler angesehen werden, je mehr Varianz der Einschätzungen von Lehrkräften und Lehrveranstaltungen auf Unterschiede in den Lehrkräften und Lehrveranstaltungen selbst und je weniger Varianz auf Unterschiede zwischen Studierenden zurückgeht. Analysen mittels kreuzklassifizierter Mehrebenenmodelle – basierend auf 3403 Evaluationsfragebögen von 699 Studierenden – ergaben eine höhere Varianzaufklärung durch Lehrkräfte und Lehrveranstaltungen als durch Studierende. Zusätzlich konnte die unerklärte Varianz durch eine Interaktion zwischen Lehrkräften und Studierenden erheblich reduziert werden. Dies deutet darauf hin, dass die jeweilige Passung von Lehrkraft und Studierenden einen großen Einfluss auf die studentische Evaluation der Lehrveranstaltung hat.

An Hochschulen werden Studierende regelmäßig gebeten, anhand von Evaluationsfragebögen eine Bewertung zu Lehrveranstaltungen abzugeben, die sie besucht haben (Peiffer, Rach, Rosanowitsch, Wörl & Schneider, 2015). Damit wird meist das Ziel verfolgt, Lehrqualität aus Sicht der Lernenden zu erfassen (Helmke, 2007; Metz-Göckel, Kamphans & Scholkmann, 2012; Rindermann, 2003). Unter Lehrqualität wird in diesem Kontext die Messung der Güte einer Lehrveranstaltung verstanden (Helmke, 2007). Eine gute Lehrveranstaltung zeichnet sich z.B. durch Strukturiertheit, Klarheit und Beteiligung der Lernenden aus (Rindermann, 2003). Darüber hinaus orientiert sich gute Lehre an den Bedürfnissen der Lernenden (Stubbs, 1995) und stellt im besten Fall ein Angebot dar, das die Lernenden annehmen (vgl. Angebot-Nutzungs-Modell von Helmke, 2007). Eine solche Definition guter Lehre im Sinne des Angebot-Nutzungs-Modells impliziert, dass die Güte der Lehrveranstaltungen

zumindest teilweise auch von Studierenden beurteilt werden können. Vor diesem Hintergrund bietet sich der Einsatz von Evaluationsfragebögen an, welche von der Zielgruppe guter Lehre - den Lernenden - ausgefüllt werden.

Evaluationsfragebögen haben sich in vielen europäischen und außereuropäischen Ländern und Instituten als gebräuchliches Erhebungsinstrument etabliert (vgl. Marsh, 1984; Nasser & Hagtvet, 2006), da sie einerseits sehr zeitökonomisch gehandhabt werden können und andererseits die Anonymität der Lernenden wahren. Weiterhin können die Ergebnisse der Evaluation Lehrkräften dabei helfen, neue Lehrveranstaltungen zu planen.

An deutschen Hochschulen wird vermehrt angestrebt, studentische Evaluationsergebnisse in Struktur- und Personalentscheidungen einfließen zu lassen (z.B. §61 Abs. 3 HHG). Eine wichtige – aber meist ungeprüft unterstellte – Voraussetzung hierbei ist jedoch, dass Evaluationsfragebögen reliabel und valide Lehrqualität erfassen. Die Reliabilität zweier typischerweise eingesetzter globaler Einschätzungen, der Leistung von Lehrkräften und der Qualität von Lehrveranstaltungen, soll in dieser Studie überprüft werden.

In bisherigen Untersuchungen wurde die Reliabilität unterschiedlicher Evaluationsfragebögen z.B. durch korrelative Studien (Marsh & Roche, 1997; Rindermann & Schofield, 2001) oder Mehrebenenanalysen (Rantanen, 2013; Spooren, 2010, Staufenbiel et. al., 2016) untersucht. So wurden in den mehrebenenanalytischen Untersuchungen ungefähr gleich große Einflüsse der Lehrkräfte und Lehrveranstaltungen sowie ein sehr großer Anteil an unerklärter Varianz gefunden.

In der vorliegenden Untersuchung stand die Reliabilität in Form der Beurteilerübereinstimmung (Interrater-Reliabilität) der globalen Einschätzungen der Leistung der Lehrkraft und der Qualität der Lehrveranstaltungen im Vordergrund. Dazu wurde die Varianz dieser Einschätzungen wie bei Rantanen (2013) und Spooren (2010) mittels kreuzklassifizierter Mehrebenenanalysen (Baayen, Davidson & Bates, 2008) mit Zufallseffekten in die einzelnen Varianzbestandteile Lehrkräfte, Lehrveranstaltungen, Studierende sowie – als ein neuer Aspekt – in die Interaktion von Lehrkräften und Studierenden aufgesplittet.

Sollte die Lehrqualität reliabel durch den eingesetzten Evaluationsfragebogen erfasst worden sein, müsste ein möglichst hoher Varianzanteil durch Lehrkräfte und

Lehrveranstaltungen erklärt werden. Gleichzeitig sollte möglichst wenig Varianz durch Studierende erklärt werden und möglichst wenig unerklärte Varianz verbleiben.

Methode

Die Stichprobe setzte sich aus 3403 Evaluationsfragebögen (FEVOR - Staufenbiel, 2000; Staufenbiel, Seppelfricke & Rickers, 2016) von 699 Psychologie-Studierenden (20,7 % Männer) des Zeitraums zwischen WiSe 2014/2015 und WiSe 2015/2016 aus dem Institut für Psychologie der Universität Kassel zusammen. Im Durchschnitt bewertete jede/r Studierende vier Lehrveranstaltungen ($Mdn = 4$; $Spannbreite = 1 - 23$), 18 verschiedene Lehrkräfte hielten 49 Vorlesungen und 45 verschiedene Lehrkräfte hielten 89 Seminare. Der vorliegende Beitrag konzentriert sich auf die globalen Einschätzungen zur Lehrkraft und Lehrveranstaltung, die beide Bestandteil des Fragebogens sind. Diese Einschätzungen wurden von den an der Lehrveranstaltung teilnehmenden Studierenden auf einer Schulnotenskala von 1 (sehr gut) bis 5 (mangelhaft) abgegeben.

Ergebnisse

Um Aussagen zur Reliabilität der studentischen Einschätzungen gegenüber Lehrkräften und der Lehrveranstaltungen machen zu können, wurden diese jeweils als abhängige Variable in getrennten Modellen für die beiden Lehrveranstaltungstypen Vorlesung und Seminar eingesetzt. Somit ergaben sich vier kreuzklassifizierte Mehrebenenmodelle, in denen jeweils die Zufallseffekte (random intercepts) der Lehrkräfte, der Lehrveranstaltungen, der Studierenden sowie – ebenfalls als Zufallseffekt – der Interaktion zwischen Lehrkräften und Studierenden berücksichtigt wurden. Weitere Interaktionen (z.B. zwischen Lehrveranstaltungen und Studierenden oder Lehrkräften und Lehrveranstaltungen, Rantanen, 2013) sind theoretisch möglich, waren aber aufgrund zu geringer Fallzahlen in den unterschiedlichen Kombinationen nicht schätzbar.

Die vier Modelle wurden mit Hilfe der Statistiksoftware R Version 3.2.2 (R Core Team, 2015) und der R-Pakete lme4 (Bates, Maechler, Bolker & Walker, 2015) und RLRsim (Scheipl, Greven & Kuechenhoff, 2008) berechnet. Als Maß der Reliabilität wurde der relative Varianzanteil im Sinne des Intra-Klassen-Korrelationskoeffizient

verwendet (vgl. Marsh, 1984, 2007; Rantanen, 2013; Rindermann, 1995, 2009). In Abbildung 1 sind die Ergebnisse der vier Modelle dargestellt.

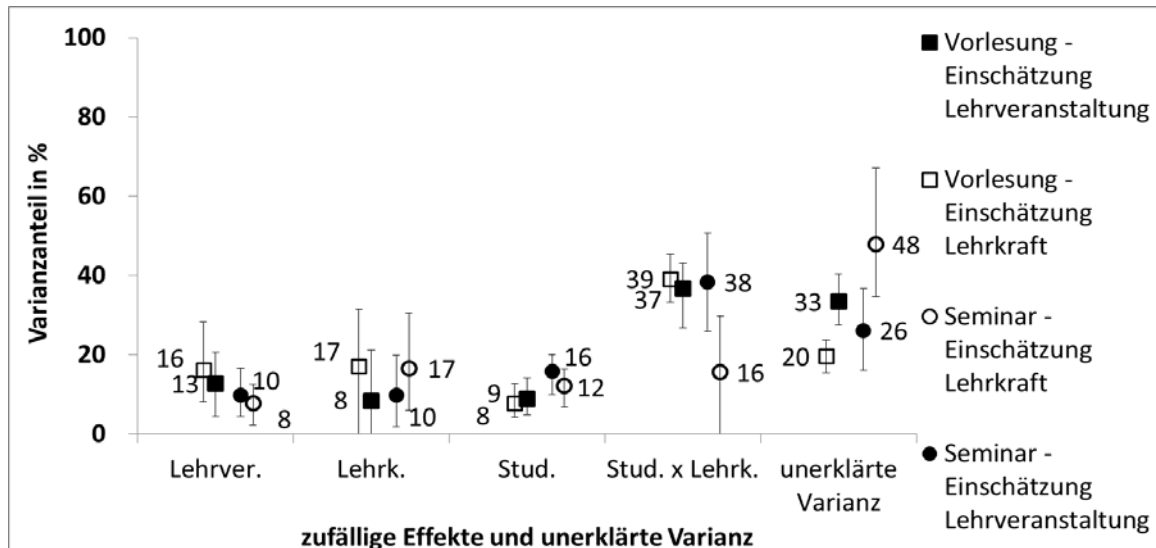


Abb. 1: Aufteilung der Varianz (%) der beiden globalen Einschätzungen Lehrkraft und Lehrveranstaltung auf die zufälligen Effekte von Lehrveranstaltung, Lehrkraft, Studierende und die Interaktion von Studierenden und Lehrkräften sowie die unerklärten Varianz in Seminaren und Vorlesungen. Als Fehlerbalken wurden 95% Konfidenzintervalle verwendet.

Einschätzung der Lehrkraft

Die Varianz der Einschätzung der Lehrkraft wurde in Vorlesungen zu 16% durch Lehrveranstaltungen, zu 17% durch die Lehrkräfte und zu 8% durch Studierende aufgeklärt. Zusätzlich wurden 39% der Varianz durch die Interaktion zwischen Lehrkräften und Studierenden aufgeklärt. Somit blieb 20% der Gesamtvarianz unerklärt. Im Vergleich dazu wurde in Seminaren die Varianz zu 8% durch Lehrveranstaltungen, zu 17% durch Lehrkräfte und zu 12% durch Studierende erklärt. Durch die Interaktion zwischen Lehrkräften und Studierenden wurde darüber hinaus noch einen Varianzanteil von 16% erklärt. Somit verblieb eine unerklärte Varianz von 48%.

Einschätzung der Lehrveranstaltung

Die Varianz der Einschätzung der Lehrveranstaltung wurde in Vorlesungen zu 13% durch Lehrveranstaltungen, zu 8% durch Lehrkräfte und zu 9% durch Studierende aufgeklärt. Zusätzlich wurden 37% der Varianz durch die Interaktion zwischen Lehrkräften und Studierenden aufgeklärt. Somit blieb 33% der Gesamtvarianz unerklärt. Im

Vergleich dazu wurde in Seminaren die Varianz zu 10% durch Lehrveranstaltungen, zu 10% durch Lehrkräfte und zu 16% durch Studierende erklärt. Durch die Interaktion zwischen Lehrkräften und Studierenden wurde darüber hinaus noch einen Varianzanteil von 38% erklärt. Somit blieb 26% der Gesamtvarianz unerklärt.

Es ist somit ersichtlich, dass durch die drei Zufallseffekte Lehrkräfte, Lehrveranstaltungen und Studierende ähnlich viel Varianz aufgeklärt wurde. Diese Varianzaufklärung lag jedoch meist unter der Varianzaufklärung der Interaktion von Lehrkräften und Studierenden, mit der der Anteil unerklärter Varianz deutlich reduziert werden konnte.

Diskussion

Die Summe der Varianzen, die durch Lehrkräfte und Lehrveranstaltungen erklärt wurden, war in drei der vier Modelle ca. zwei- bis dreimal so groß, wie die Varianz, die durch die Studierenden erklärt wurde. Dies ist mit den Ergebnissen von Rantanens (2013) Modell F vergleichbar, aber konträr zu Spoorens (2010) Modell 0, welches eine größere Varianzaufklärung durch die Studierenden aufzeigt. Außerdem liegt die durch Lehrkräfte und Lehrveranstaltung aufgeklärte Varianz über der von Marsh und Roche (1997) als Norm vorgeschlagenen Single-Rater-Reliabilität von 20%. Dies spricht für eine insgesamt reliable Erfassung der Einschätzung der Lehrkraft in Vorlesungen und Seminaren und der Einschätzung der Lehrveranstaltung in Vorlesungen.

Einzige Ausnahme war das Modell zur Einschätzung der Lehrveranstaltungen in Seminaren: In diesem wurden durch Lehrkräfte und Lehrveranstaltungen zusammen 20% Varianz und durch Studierende 16% Varianz aufgeklärt. Auch in diesem wird das von Marsh und Roche vorgeschlagene Kriterium für eine akzeptable Single-Rater-Reliabilität erreicht. Gleichzeitig ist die durch Studierende aufgeklärte Varianz aber ähnlich hoch.

Über diese Ergebnisse hinaus ging ein weit höherer Varianzanteil in drei Modellen auf die Interaktion aus Studierenden und Lehrkräfte zurück. Einzige Ausnahme war das Modell der Einschätzung der Lehrkraft in Seminaren. Hier lag die Varianzaufklärung der Interaktion auf einem ähnlichen Wert wie die der zufälligen Effekte. Diese insgesamt hohe Varianzaufklärung spricht dafür, dass das Zusammenspiel zwischen bestimmten Studierenden und Lehrkräfte nicht ignoriert werden sollte.

So wäre es zum Beispiel möglich, dass Studierende eine bestimmte Lehrkraft wegen ihrer Persönlichkeit (Helmke, 2007), ihres Geschlechts und ihrer Attraktivität (Böttcher & Grewe, 2010) anders einschätzen als andere Lehrkräfte. Ebenso könnte diese Interaktion auch durch die Expertise der Lehrkraft (Helmke, 2007) ausgelöst werden. So wäre es denkbar, dass bestimmte Lehrkräfte speziell auf Eigenschaften der Studierenden (z.B. Vorwissen) eingehen, sich die Lehrveranstaltung dadurch verändert und die Studierenden mit ihrer Einschätzung auf die veränderte Lehrveranstaltung reagieren. Um hierzu jedoch weitere Aussagen treffen zu können, ist eine weitergehende Analyse der Interaktion notwendig.

Die unerklärte Varianz konnte durch die Einführung der Interaktion zwischen Lehrkräften und Studierenden stark reduziert werden. Im Vergleich zu den Befunden anderer Forscher (Rantanen, 2013; Spooren, 2010) kann es als ein Fortschritt angesehen werden, den Wert der meist 50% unerklärten Varianz zu halbieren. Dabei ist es wichtig zu sehen, dass der Einbezug dieser Interaktion fast ausschließlich die unerklärte Varianz, jedoch nicht auch die anderen Zufallseffekte reduzierte. Zusammenfassend legen die Befunde nahe, dass mit den verwendeten globalen Variablen – gemessen an dem 20%-Kriterium von Marsh und Roche (1997) – eine akzeptabel reliable Einschätzung der Lehrveranstaltung und der Lehrkraft erzielt werden kann. In jedem Fall sollte nur auf aggregierte Werte zurückgegriffen werden, um Aussagen über Lehrveranstaltungen bzw. Lehrkräfte zu treffen, da einzelne Evaluationen hohen Messfehlern unterliegen und eine starke interindividuelle Varianz aufweisen. Dies kann zum Beispiel durch die Studierenden oder die Kombination von Studierenden und Lehrkräften bedingt sein.

Literatur

- Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.

- Böttcher, W. & Grewe, C. M. (2010). Eine Untersuchung zur Wirksamkeit der studentischen Lehrveranstaltungskritik am Beispiel der Westfälischen Wilhelms-Universität Münster. In P. Pohlenz (Hrsg.), *Wie viel Wissenschaft braucht die Evaluation? Eine Einführung* (Qualität - Evaluation - Akkreditierung, Bd. 4, S. 73-82). Bielefeld: UVW Universitäts Verlag.
- Helmke, A. (2007). *Unterrichtsqualität erfassen, bewerten, verbessern* (6. Aufl.). Seelze: Klett Kallmeyer.
- Hessisches Hochschulgesetz (HHG) vom 14.12.2009 (GVBl. I S. 666) zuletzt geändert durch Artikel 1 des Gesetzes vom 30. November 2015 (GVBl. S. 510).
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H. W. & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Hrsg.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (S. 319-383). Dordrecht: Springer Netherlands.
- Metz-Göckel, S., Kamphans, M. & Scholkmann, A. (2012). Hochschuldidaktische Forschung zur Lehrqualität und Lernwirksamkeit. *Zeitschrift für Erziehungswissenschaft*, 15, 213-232.
- Nasser, F. & Hagtvet, K. A. (2006). Multilevel analysis of the effects of student and instructor/course characteristics on student ratings. *Research in Higher Education*, 47, 559-590.
- Peiffer, H., Rach, H., Rosanowitsch, S., Wörl, J. & Schneider, M. (2015). Lehr-evaluation. In M. Schneider & M. Mustafić (Hrsg.), *Gute Hochschullehre: Eine evidenzbasierte Orientierungshilfe. Wie man Vorlesungen, Seminare und Projekte effektiv gestaltet* (S. 153-184). Berlin: Springer.
- R Core Team (2015). *R: A language and environment for statistical computing* [Computer program]. R Foundation for Statistical Computing, Vienna, Austria. Verfügbar unter: <https://www.R-project.org/>.

- Rantanen, P. (2013). The number of feedbacks needed for reliable evaluation. A multilevel analysis of the reliability, stability and generalisability of students' evaluation of teaching. *Assessment & Evaluation in Higher Education*, 38, 224-239.
- Rindermann, H. (1995). *Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen*. Dissertation. *Psychologie: Vol. 6*. Landau: Empirische Pädagogik.
- Rindermann, H. (2003). Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. *Zeitschrift für Evaluation*, 2, 233-256.
- Rindermann, H. (2009). Qualitätsentwicklung in der Hochschullehre. *Beiträge zur Lehrerbildung - Qualifizierung von Dozierenden der Hochschulstufe*, 27, 64-73.
- Scheipl, F., Greven, S. & Kuechenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis*, 52, 3283-3299.
- Spooren, P. (2010). On the credibility of the judge: A cross-classified multilevel analysis on students' evaluation of teaching. *Studies in Educational Evaluation*, 36, 121-131.
- Staufenbiel, T. (2000). Fragebogen zur Evaluation von universitären Lehrveranstaltungen durch Studierende und Lehrende. *Diagnostica*, 46, 169-181.
- Staufenbiel, T., Seppelfricke, T. & Rickers, J. (2016). Prädiktoren studentischer Lehrveranstaltungsevaluationen. *Diagnostica*, 62, 44-59.
- Stubbs, W. H. (1995). Quality in higher education: A funding council perspective. In D. Green (Hrsg.), *What is Quality in Higher Education?* (S. 21-28). Buckingham, Bristol, PA: Society for Research into Higher Education & Open University Press.