Marvin Krüger
Friederike Korneck
Lars Oettinghaus
Mareike Kunter

Goethe-Universität Frankfurt am Main

Bewertung von Unterrichtsqualität
Schulart- und lehrkraftspezifische Unterschiede

In diesem Beitrag sollen die folgenden Fragestellungen beantwortet werden: Welchen Effekt hat die Gruppe der Lehrkräfte (Haupt- und Realschullehramtsstudierende, Gymnasiallehramtsstudierende, Gymnasialreferendare) auf die Unterrichtsqualität? Welchen Effekt hat die Schulform der Schüler (Gymnasium, Gesamtschule) auf die Unterrichtsqualität?

Theoretischer Hintergrund

Unterrichtliche Prozesse sowie deren Tiefenstrukturen werden weder alleine durch die Lehrkraft noch alleine durch die Schüler gestaltet. Die Einflüsse beider Gruppen sind grundsätzlich konfundiert und lassen sich nicht ohne Weiteres trennen. Bei der Untersuchung des Einflusses der Schulart der Schüler bzw. der Lehrkraft auf die Qualität des Unterrichts muss daher berücksichtigt werden, dass Lehrkräfte schulartspezifisch ausgebildet und eingesetzt werden. Während dieser Umstand in den bisherigen Untersuchungen zum Einfluss der Schulform auf die Unterrichtsqualität (z. B. Kunter et al., 2011) nicht berücksichtigt werden kann, ermöglicht das spezielle Setting der Studie Φactio durch die zufällige Zuordnung angehender Lehrkräfte unterschiedlicher Studiengänge zu den verschiedenen Schulformen, diese Effekte einzeln zu untersuchen.

Design

Die Unterrichtsqualität angehender Physiklehrkräfte wird in 12- bis 15-minütigen komplexitätsreduzierten Unterrichtsminiaturen, welche in einer studiengang- und phasenübergreifenden Lehrveranstaltung entwickelt werden und in deren Mittelpunkt jeweils ein einzelnes physikalisches Freihandexperiment steht, erhoben. Das Lehr- und Erhebungsdesign wird in Korneck, Oettinghaus & Kunter (in Arbeit) ausführlicher beschrieben. Die Erhebung erfolgt mit Hilfe eines Fragebogens, in dem sowohl die unterrichteten Schüler als auch die hospitierenden Peers ihre Einschätzungen abgeben. Die Teilnehmer (Haupt- und Realschullehramtsstudierende, Gymnasiallehramtsstudierende, Gymnasialreferendare) werden randomisiert zu den verschiedenen Schulformen (Gymnasium, Gesamtschule) zugeteilt.

Instrument

Die Erfassung der Unterrichtsqualität erfolgt mit Hilfe vierstufig Likert-skalierter Fragebögen, deren Items von der COACTIV-Studie (Baumert et al., 2009) stammen. Durch eine entsprechende Adaption an das Setting ist ein Fragebogen mit insgesamt 22 Items entstanden, von denen aufgrund der kurzen Erhebungszeit von lediglich zwei Minuten in zwei unterschiedlichen Testheften jeweils 18 Items verwendet werden. 14 Items sind als Ankeritems geführt, die anderen acht rotieren. Das Instrument basiert auf der Einteilung in die drei Basisdimensionen von Unterrichtsqualität: Kognitive Aktivierung, Klassenführung und konstruktive Unterstützung. Jede Skala ist mit sieben bis acht Items sowie drei Subskalen vertreten. Gegenüber dem Instrument aus der COACTIV-Studie wurde dabei aufgrund des speziellen Settings auf die soziale Komponente der konstruktiven Unterstützung verzichtet.

Stichprobe

Die Stichprobe umfasst aktuell 172 Unterrichtsminiaturen, die von insgesamt 86 Lehrkräften (zwei pro Lehrkraft) gestaltet wurden. Es handelt sich um eine Vollerhebung der Physiklehr-

Erschienen in: S. Bernholt (Hrsg.) (2015). Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Bremen 2014. Kiel: IPN.

amtsstudierenden am Studienort Frankfurt a. M. seit Wintersemester 2012/13. Jede Sequenz wird von durchschnittlich zehn Peers sowie den durchschnittlich zwölf Schülern einer Klassenhälfte bewertet. Daraus resultiert eine Gesamtzahl von 1719 Peer- und 2060 Schülerfragebögen, die an insgesamt 15 Unterrichtstagen bzw. in 15 Klassen aus 5 Schulen erfasst wurden. Bei den Schulen handelt es sich um Gymnasien (3) oder Gesamtschulen (2). In diesem Beitrag werden exemplarisch Ergebnisse, die auf den Unterrichtsbewertungen der hospitierenden Peers basieren, vorgestellt und diskutiert.

Auswertung

Aufgrund der geschachtelten Datenstruktur (jede Lehrkraft/Sequenz wird von etwa 10/20 Peers und 12/24 Schülern bewertet), werden zur Auswertung Mehrebenenstrukturgleichungsmodelle in M*plus* (Muthén & Muthén, 2010) verwendet. Für alle drei Basisdimensionen wurde ein eigenes Modell geschätzt, in dem jeweils sowohl auf der Rating- als auch auf der Lehrerebene drei Indikatoren den latenten Faktor der entsprechenden Dimension bilden. Als Indikatoren wurden auf beiden Ebenen entsprechende WLE-Schätzer verwendet, die mit Hilfe von ConQuest (Wu et al., 2007) einem jeweils dreidimensionalen Partial Credit Model entnommen wurden. Anschließend wurden die entsprechenden Regressoren auf Lehrerebene eingeführt. Die resultierenden Modelle zeigen entsprechend allgemeiner Cut-off-Kriterien angemessene Modellfits (Geiser, 2011). Die Intraklassenkorrelationen, die in Mehrebenenalysen als Maß der Reliabilität verwendet werden können, zeigen mit ICC(2) = .70 bis .83 außerdem an, dass die Einschätzungen der Peers reliabel sind (Lüdtke et al., 2006; Wagner, 2011).

Ergebnisse

Bezüglich des Einflusses der Probandengruppe der Lehrkräfte lassen sich die folgenden Effekte identifizieren (siehe Abb. 1):

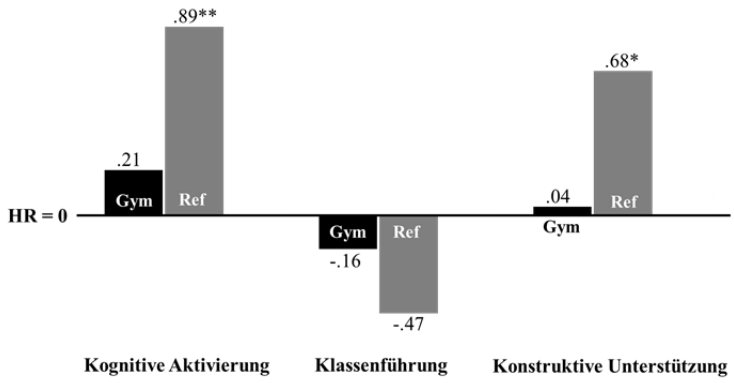

*Abb. 1: Einfluss der Probandengruppe der Lehrkräfte auf die Unterrichtsqualität, teilweise standardisierte Regressionskoeffizienten, **p<.01, *p<.05 (zweiseitig), kleine Effekte > .20, mittlere Effekte > .50, große Effekte > .75*

Die Referendare (Ref.) zeigen gegenüber der Referenzgruppe der Haupt- und Realschullehramtsstudierenden (HR-Stud.) einen signifikanten großen Effekt auf die kognitive Aktivierung. Die Gymnasiallehramtsstudierenden (Gym-Stud.) zeigen lediglich einen kleinen Effekt, der nicht abgesichert werden kann. Der im Bereich der Klassenführung gefundene kleine negative Effekt der Ref. kann ebenfalls nicht abgesichert werden. Die Effekte der Gym.-Stud. gegenüber den HR-Stud. in Klassenführung und konstruktiver Unterstützung sind nicht substanziell. Im Bereich der konstruktiven Unterstützung kann jedoch ein mittlerer Effekt der Ref. gegenüber den HR-Stud. abgesichert werden. Ebenso lassen sich die Effekte der Ref. in der kognitiven Aktivierung wie auch in der konstruktiven Unterstützung

auch gegenüber den Gym-Stud. absichern, sie liegen beide im mittleren Bereich. Bezüglich des Einflusses der Schulform der Schüler lassen sich die folgenden Effekte identifizieren (siehe Abb. 2):

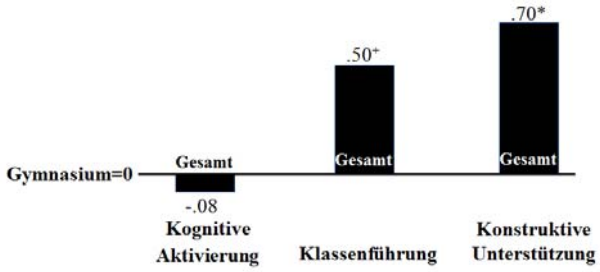

*Abb. 2: Einfluss der Schulform der Schüler auf die Unterrichtsqualität, teilweise standardisierte Regressionskoeffizienten, *p<.05, +p<.10 (zweiseitig), kleine Effekte > .20, mittlere Effekte > .50, große Effekte > .75*

Im Bereich der kognitiven Aktivierung ist kein substanzieller Effekt zu finden, der anfänglich mittlere Effekt im Bereich der Klassenführung zugunsten der Gesamtschule lässt sich nur trendweise absichern. Signifikant ist lediglich der mittlere Effekt in der konstruktiven Unterstützung, die in der Gesamtschule entsprechend ausgeprägter zu sein scheint.

Diskussion und Ausblick

Durch das spezielle Setting von Φactio konnten die Einflüsse von Lehrkräften sowie Schülern getrennt und jeweils erfolgreich untersucht werden. Der Einfluss der Gruppe der Lehrkraft hat sich als dominant gegenüber dem Einfluss der Schulform der Schüler herausgestellt. Ersterer konnte im mittleren bis großen Effektbereich sowohl für die kognitive Aktivierung als auch für die konstruktive Unterstützung gefunden werden, für die Schulform konnte lediglich ein mittlerer Effekt in der konstruktiven Unterstützung nachgewiesen werden. Im Bereich der Klassenführung, die durch den Rahmen der Unterrichtsminiaturen sehr starr vorgegeben ist, konnte erwartungskonform kein Effekt abgesichert werden.Die Frage danach, welche Effekte gefunden werden können, wenn noch unterschiedlichere Schulformen wie Haupt- und Realschule untersucht werden, steht aus.

Literatur

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U. et al. (2009). Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente (Materialien aus der Bildungsforschung Nr. 83). Berlin: Max-Planck-Institut für Bildungsforschung.

Geiser, C. (2010). Datenanalyse mit Mplus. Wiesbaden: VS-Verlag für Sozialwissenschaften.

Korneck, F., Oettinghaus, L. & Kunter, M. (in Arbeit). Φactio: Messung von Unterrichtsqualität in komplexitätsreduzierten Unterrichtssequenzen. In Rauin, U., Herrle, M., Engartner, T., Videoanalysen in der Unterrichtsforschung Methodische Vorgehensweisen und Anwendungsbeispiele. Weinheim: Beltz.

Kunter, M., Baumert, J., Blum,W., Klusmann, U., Krauss, S. & Neubrand, M. (2011). Professionelle Kompetenz von Lehrkräften - Ergebnisse des Forschungsprogramms COACTIV. Münster: Waxmann.

Lüdtke, O., Trautwein, U., Kunter, M. & Baumert, J. (2006). Analyse von Lernumwelten. Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. Zeitschrift für Pädagogische Psychologie, 20(1/2), 85–96.

Muthén, L. K., & Muthén, B. O. (2010). Mplus user's guide. Los Angeles, CA: Muthén & Muthén.

Wagner, W. (2008). Methodenprobleme bei der Analyse der Unterrichtswahrnehmung und -wirksamkeit - am Beispiel der Studie DESI (Deutsch Englisch Schülerleistungen International) der Kultusministerkonferenz. Dissertation. Universität Koblenz-Landau, Campus Landau, Fachbereich Psychologie.

Wu, M. L., Adams, R. J., Wilson, M. R., Haldane, S. A. (2007). Acer ConQuest Version 2.0: generealised item response modelling software. Melbourne: ACER Press.