

Marvin Krüger<sup>1</sup>  
 Friederike Korneck<sup>1</sup>  
 Lars Oettinghaus<sup>1</sup>  
 Mareike Kunter<sup>2</sup>

Goethe-Universität Frankfurt am Main  
<sup>1</sup>Institut für Didaktik der Physik  
<sup>2</sup>Institut für Psychologie

## Perspektiven auf Unterrichtsqualität in Unterrichtsminiaturen

Motivationale wie kognitive Lernergebnisse von Schüler(innen) werden durch das Handeln ihrer Lehrkraft geprägt. In der bildungswissenschaftlichen wie didaktischen Forschung liegt der Fokus daher häufig neben der theoretischen Betrachtung dieses Handelns auf dessen empirischer Messung. Die Studie  $\Phi$ actio untersucht die Lehrerüberzeugungen als (vermutlich) wirksame Determinanten qualitativollen Unterrichtshandelns und verfolgt bei der Erhebung der Unterrichtsqualität einen neuen Weg. Statt der Betrachtung von Regelunterricht werden sog. Unterrichtsminiaturen als Erhebungsgrundlage gewählt. Dabei handelt es sich um kurze, inhaltlich abgeschlossene Unterrichtssequenzen von 12 Minuten Länge, in deren Mittelpunkt ein Freihandexperiment aus der Mechanik steht.

Der vorliegende Beitrag diskutiert die Verwendung unterschiedlicher Zugänge zu Unterrichtsqualität in diesem Setting, stellt die in  $\Phi$ actio eingesetzten Instrumente dar und beleuchtet ihre Eignung in Hinblick auf Reliabilität und Validität.

### Theoretischer Hintergrund

Bei der Betrachtung unterrichtlicher Prozesse lassen sich prinzipiell unterschiedliche Strukturebenen unterscheiden, von deren Wahl der Grad an Inferenz bei der Erhebung entsprechender Merkmale abhängt. Im Bereich der Sichtstruktur finden sich Oberflächenmerkmale, die sich mit keiner oder wenig Inferenz beurteilen lassen, z. B. Unterrichtsmethoden oder Sozialformen. Merkmale der Tiefenstruktur sind dagegen solche, die sich einem Beobachter nicht unmittelbar erschließen, jedoch mit einem höheren Grad an Inferenz dennoch beurteilbar sind, z. B. die Beziehungsqualität von Lehrkräften und Schüler(innen). Seidel und Shavelson (2007) sowie Hattie (2009) zeigten in ihren Metaanalysen, dass es insbesondere die Merkmale der Tiefenstrukturen sind, die erfolgreiches Lernen prädiktieren. Ein empirisch und theoretisch fundiertes Modell der wesentlichen Merkmale von Unterrichtsqualität im Bereich der Tiefenstrukturen entwickelten Klieme et al. (2006). Sie beschreiben die drei Basisdimensionen „kognitive Aktivierung“, „Klassenführung“ und „schülerorientiertes Sozialklima“, wobei letzteres hier im Sinne von Kunter und Trautwein (2013) als „konstruktive Unterstützung“ verstanden wird. Kognitive Aktivierung entspricht dem „intellektuellen Anforderungsgehalt“ (Kunter & Trautwein, 2003, S. 86) des Unterrichts mit der dahinterstehenden Vorstellung einer aktiven Wissenskonstruktion seitens der Schüler(innen). Die Klassenführung umfasst die klare, strukturierte und insbesondere störungspräventive Unterrichts-führung durch eine entsprechende Vorbereitung und Ausgestaltung des Unterrichts. Konstruktive Unterstützung bezieht sich wesentlich auf die Selbstbestimmung der Schüler(innen) und wird z. B. durch ein adäquates Interaktionstempo und eine positive Fehlerkultur ermöglicht.

Zur Erhebung von Unterrichtsqualität werden in der Regel Beurteilungen von Schüler(innen), Lehrkräften und/oder externen (Video-)Beobachtern herangezogen. Alle drei Zugänge weisen spezifische Vor- und Nachteile auf (vgl. Clausen, 2002): Aufgrund der Möglichkeit zur Aggregation vieler Urteile zu einer Lehrkraft, weisen die Beurteilungen durch Schüler(innen) in der Regel eine hohe Reliabilität auf, zudem ist ihr Einsatz sehr ökonomisch. Demgegenüber sind aufgrund mangelnder didaktisch-methodischer Expertise jedoch Halo-Effekte zu erwarten, zudem sind die Urteile möglicherweise durch andere Faktoren, wie die Lehrerbeliebtheit, beeinflusst. Schüler(innen) fällt es schwer, Urteile zu eige-

nen Lehrkräften bezüglich eines bestimmten Beurteilungszeitraumes abzugeben. Die unterrichtenden Lehrkräfte lassen sich im Sinne einer Selbstauskunft ebenfalls befragen, mögliche selbstdienliche Verzerrungen überwiegen jedoch den Vorteil des ökonomischen Einsatzes, weshalb in dieser Studie auf die Erhebung der Lehrerperspektive verzichtet wurde. Der Einsatz externer (Video-)Beobachter scheint der beste Weg für die Erhebung zu sein, da sie eine hohe Objektivität, eine hohe Validität (aufgrund ihrer Expertise) und Reliabilität (aufgrund ihrer Schulung) aufweisen. Nachteil der Beobachterperspektive ist der hohe Aufwand und die damit verbundenen Kosten. Szogs et al. (in diesem Band) entwickelten daher ein sehr schlankes, aber dennoch differenziertes Videomanual zur ökonomischen Erhebung der Unterrichtsqualität im nachfolgend beschriebenen Setting. Im Fokus dieses Beitrags stehen die beiden anderen, in der Studie *Phactio* eingesetzten, Perspektiven: Neben den Schüler(innen) sind dies die (hospitierenden) Peers, andere angehende Physiklehrkräfte, die als entsprechend ungeschulte, aber didaktisch versierte Beobachter beschrieben werden können.

### **Design und Stichprobe**

Die vorliegende Studie erfasst Unterrichtsqualität angehender Physiklehrkräfte durch eine Erhebung im Rahmen 12-minütiger Unterrichtsminiaturen, welche in einer studiengang- und phasenübergreifenden Lehrveranstaltung von den Teilnehmer(innen) eigenständig entwickelt werden und in deren Mittelpunkt jeweils ein Freihandexperiment steht (Korneck et al., 2015). Der Unterricht findet an kooperierenden Gymnasien und Gesamtschulen mit Klassen der Sekundarstufe I statt. Die Lehrkräfte sind Studierende des Haupt-/Realschul- sowie des gymnasialen Lehramts im Hauptstudium und gymnasiale Lehrkräfte im ersten Modul des Vorbereitungsdienstes.

Die Stichprobe umfasst die Teilnehmer(innen) der Seminarveranstaltungen zwischen WiSe 2012/13 und SoSe 2014 und damit insgesamt 84 angehende Physiklehrkräfte. Da alle Teilnehmer(innen) doppelt unterrichten, stehen für die Auswertungen insgesamt 168 Unterrichtssequenzen sowie die dazugehörigen Beurteilungen der Unterrichtsqualität von durchschnittlich 10 hospitierenden Peers und 12 unterrichteten Schüler(innen) zur Verfügung.

### **Instrument**

Für die Erhebung der Einschätzungen von unterrichteten Schüler(innen) und hospitierenden Peers bedurfte es aufgrund der organisatorischen Gegebenheiten im Rahmen der Lehrveranstaltung eines kurzen Fragebogens, der innerhalb von zwei Minuten beantwortbar ist und dennoch die drei Basisdimensionen möglichst umfassend abdeckt. Die Operationalisierung fand in direkter Anlehnung an die COACTIV-Studie (Baumert et al., 2009) statt, deren Items für das vorliegende Setting adaptiert wurden. Es wurden dabei die Items des entsprechenden Schülerfragebogens verwendet, allerdings auf den Bereich der sozialen konstruktiven Unterstützung verzichtet, da diese im vorliegenden Falle fremder Lehrkräfte bzw. Schüler(innen) nicht zum Tragen kommen kann. Daraus entstand ein Itempool von insgesamt 22 Items (7 bis 8 pro Basisdimension), die in zwei verschiedene Testhefte aufgeteilt wurden. Die 10 repräsentativsten Items für die jeweilige Skala wurden als Ankeritems eingesetzt und 6 weitere jeweils auf die Testhefte verteilt.

### **Analysen und Ergebnisse**

Bisherige Analysen auf Ebene der 84 Lehrkräfte, in die die Urteile von Schüler(innen) und Peers durch die Bildung von Durchschnittsscores Eingang fanden, zeigten durch angemessene bis sehr gute Skalenreliabilitäten und Intraklassenkorrelationen eine ausreichende Reliabilität des eingesetzten Instrumentes an. Die Einschätzungen der Peers weisen gegenüber den Schüler(innen) zum Teil sehr gute Intraklassenkorrelationen ( $ICC(2) > .85$ ) auf und erlauben es sogar, eine Analyse auf Ebene der einzelnen Unterrichtssequenzen durchzuführen. Die

Schüler(innen) weisen im Allgemeinen akzeptable Beobachterübereinstimmungen auf, liegen jedoch bezüglich aller Merkmale unter denen der Peers und knapp unterhalb von .70. Neben der Reliabilität sind insbesondere Rückschlüsse auf die Validität des eingesetzten Tests möglich. Ein Vergleich der Schüler(innen)- und Peer-Perspektive zeigt substantielle, signifikante Übereinstimmungen beider Zugänge und spricht für eine konvergente Konstruktvalidität, die in anderen Studien zur Unterrichtsqualität bislang nicht erreicht werden konnte (z. B. Fauth et al., 2014). Dies ist vermutlich auf das spezielle Setting zurückzuführen, das durch die Beurteilung der Unterrichtsminiaturen fremder Lehrkräfte zwei wesentliche Nachteile der Schüler(innen)perspektive abschwächt: Die Urteile der Schüler(innen) beziehen sich nur auf den untersuchten Beobachtungszeitraum und können hier als objektiv gelten. Die Modellierung des Zusammenhangs zwischen Unterrichtsqualität und Lehrerüberzeugungen liefert Ergebnisse, die zur empirischen Befundlage konform sind (siehe Oettinghaus et al., in diesem Band). Dies kann als positiver Befund zur Kriteriumsvalidität gewertet werden.

### Diskussion und Ausblick

Im Rahmen der bisherigen Analysen in der Studie  $\Phi$ actio konnte nachgewiesen werden, dass das Schülerinstrument der COACTIV-Studie auch für den Einsatz bei miniaturisiertem Unterricht geeignet erscheint und eine hohe Güte aufweisen kann. Aufgrund des berichteten Zusammenhangs von Schüler- und Peer-Perspektive und der gleichermaßen schlechteren Beobachterübereinstimmung der Schüler(innen), wird im Folgeprojekt auf die Erhebung der drei Basisdimensionen durch die Schüler(innen) verzichtet. Stattdessen erscheint es nun sinnvoll, den Fokus auf andere Aspekte zu legen, über die nur die Schüler(innen) Auskunft geben können. Dazu gehören insbesondere motivationale und kognitive Merkmale ihres Lernprozesses, wie ihre soziale Eingebundenheit. Derartige Merkmale liegen näher am tatsächlichen Outcome des Unterrichts und können damit weitere interessante Einblicke liefern. Aktuell werden die Unterrichtssequenzen mit Hilfe des neu entwickelten Videomanuals beurteilt. Durch einen Vergleich der Urteile von Schüler-, Peer- und Videobeobachtern wird dadurch eine weitere Möglichkeit zur Überprüfung der Konstruktvalidität gegeben. In Anlehnung an das Videomanual, das gegenüber dem Instrument der COACTIV-Studie eine bessere Passung an das Setting aufweist, wird zudem der Peerfragebogen überarbeitet.

### 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.
- Clausen, M. (2002). Unterrichtsqualität: Eine Frage der Perspektive? Münster: Waxmann.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28 (3), S. 127–137.
- Klieme, E., Lipowsky, R., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts Pythagoras. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.
- Korneck, F., Oettinghaus, L., Kunter, M., & Redinger, R. (2015). Überzeugungen und Handlungen von Lehrpersonen - Messung von Unterrichtsqualität in komplexitätsreduzierten Settings des Physikunterrichts. In U. Rauin, M. Herrle, & T. Engartner (Hrsg.), *Videoanalysen in der Unterrichtsforschung - Methodische Vorgehensweisen und aktuelle Anwendungsbeispiele*. Weinheim: Beltz Juventa.
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Paderborn: Schöningh.
- Oettinghaus, L., Krüger, M., Korneck, F., Kunter, M. (in diesem Band). Überzeugungen und Unterrichtshandeln.
- Szogs, M., Krüger, M., Korneck, F., Oettinghaus, L. & Kunter, M. (in diesem Band). Kognitive Aktivierung in standardisierten Unterrichtsminiaturen.