

Michael Szogs¹
 Marvin Krüger¹
 Friederike Korneck¹

¹Goethe-Universität Frankfurt

Erhebung von Unterrichtsqualität mittels hoch-inferenter Videoratings - das Ratingmanual der Φ actio-Studie

Einleitung

Dieser Beitrag stellt das Ratingmanual der Φ actio-Studie vor, mit dem die Qualität von Physikunterrichtsminiaturen gemessen wird. Ziel der Studie ist die Untersuchung des Einflusses von Professionswissen und Lehrerüberzeugungen angehender Physiklehrkräfte auf ihre Unterrichtsqualität. Die Studie ist an eine Microteaching-Lehrveranstaltung gekoppelt, bei der Studierende und Lehrkräfte im Vorbereitungsdienst eine inhaltlich abgeschlossene Unterrichtsminiatur zu einem Freihandexperiment planen, vor einer fremden Schulklasse durchführen und anschließend reflektieren. Die Unterrichtsminiaturen zeichnen sich durch eine Komplexitätsreduktion dahingehend aus, dass nur eine Klassenhälfte unterrichtet wird und die Dauer des Unterrichts einer Lehrperson auf circa zwölf Minuten beschränkt ist. Die Lehrkräfte erhalten die Möglichkeit, ihre Unterrichtsminiatur nach einem kollegialen Feedback leicht verändert ein zweites Mal an demselben Unterrichtstag mit der anderen Klassenhälfte zu gestalten (vgl. Korneck et al. 2016).

Zur Erhebung der Unterrichtsqualität werden die Unterrichtsminiaturen videografiert, um sie mit Hilfe des nachfolgend dargestellten Ratingmanuals auszuwerten.

Zielsetzung und Struktur des Ratingmanuals

Ein zentrales Ziel in der Entwicklung des Manuals ist es, die Unterrichtsqualität auf Ebene von Subdimensionen reliabel erfassen zu können. Damit soll die Auflösung bei der Untersuchung von Zusammenhangsstrukturen zur professionellen Kompetenz erhöht werden, um spezifischere Ergebnisse erhalten zu können. Gleichzeitig soll das Instrument mit einem möglichst geringen Zeitaufwand und ohne eine aufwendige Einarbeitung der Rater verwendet werden können. Hierzu wird angestrebt, das Ratingmanual möglichst gut handhabbar zu gestalten, um dennoch zufriedenstellende Beobachterübereinstimmungen zu erzielen.

Inhaltlich orientiert sich das Manual an dem Modell der drei Basisdimensionen von Klieme, Schümer und Knoll (2001) in der von Kunter und Voss (2011) beschriebenen Variante und umfasst entsprechend die *kognitive Aktivierung*, die *konstruktive Unterstützung* und die *Klassenführung*. Dabei wird die konstruktive Unterstützung getrennt in einen strukturellen Bereich, der die gezielte Förderung des Lernprozesses der Schülerinnen und Schüler durch die Lehrperson untersucht und einen affektiven Bereich, der auf das Schaffen eines positiven Klassenklimas mit Motivation und Interesse fördernden Elementen fokussiert. Kognitive Aktivierung beschreibt, inwiefern die Lehrperson die Schülerinnen und Schüler zu einem vertieften Nachdenken über den Lerngegenstand anregt, während die Klassenführung umfasst, ob die Lehrperson einen reibungslosen Unterrichtsablauf gewährleistet. Weiterhin werden die drei Basisdimensionen durch zwei Skalen zur *fachlichen Korrektheit* und *fachlichen Transparenz* des Unterrichts ergänzt.

Eine Herausforderung bei der Entwicklung des Manuals war die physikunterrichtsspezifische Operationalisierung der allgemeinen Unterrichtsqualitätsmerkmale. Hierzu wurden Items aus der PERLE-Videostudie (Lotz, Lipowsky & Faust, 2011), der IPN-Videostudie (Seidel, Prenzel, Duit & Lehrke, 2003), der COACTIV-Studie (Baumert et al., 2009), Vogelsang (2015) sowie einem im hessischen Referendariat eingesetzten Reflexionsbogen zusammengetragen und an Unterrichtsfach sowie Setting

adaptiert. Anschließend wurden die inhaltlich sowie empirisch strukturierten Subdimensionen durch Eigenentwicklungen ergänzt. Ein Teil dieser neu entwickelten Items ließ sich dabei unmittelbar aus der Veranstaltung selbst gewinnen. Die von den Studierenden in den Feedbackrunden eingebrachten Verbesserungsvorschläge zu den Unterrichtsminiaturen haben sich als sehr geeignete Indikatoren für Unterrichtsqualität gezeigt, sodass sie mit geringer Modifikation als Items übernommen werden konnten. Das Ergebnis ist ein Ratingmanual, mit dem die Ausprägung von 20 Subskalen der Unterrichtsqualität auf Basis des Gesamteindrucks der Unterrichtsminiatur beurteilt werden kann. Die nachfolgende Tabelle 1 bietet eine Übersicht der Subskalen.

Basisdimension	Subskalen (Anzahl der Items)
Kognitive Aktivierung	Umgang mit Vorwissen und Schülerbeiträgen (8), Kognitive Selbstständigkeit (11), Diskursives Lernen (8), Potential zum Konzeptwechsel (8)
Konstruktive Unterstützung (affektiv)	Lehrer-Schüler-Beziehung (8), Anerkennung (5), Fehlerkultur (7), Interesse (6), Relevanz (6), Autonomie (9)
Konstruktive Unterstützung (strukturell)	Interaktionstempo (6), Adaptive Erleichterung (9), Diagnose von Verständnisschwierigkeiten (12), Inhaltliche Kohärenz (8)
Klassenführung	Allgegenwärtigkeit (6), Gruppenfokus (9), Übergangs- und Zeitmanagement (8), Störungsfreiheit (7)
	Fachliche Korrektheit (6), Fachliche Transparenz (6)

Tab. 1: Überblick über die Skalen des Ratingmanuals

Ratingprozess

Jeder Skala ist eine einführende Grundidee vorangestellt. Diese beschreibt einleitend den adressierten Unterrichtsqualitätsaspekt und unterstützt so die Beurteilung der Items, welche auf einer vierstufigen Likert-Skala eingeschätzt werden. Für das Rating wird zunächst die gesamte Unterrichtssequenz beobachtet, um dann die Items auf Basis sämtlicher Vorkommnisse und Interaktionen in der jeweiligen Miniatur einzuschätzen. Für die Durchführung eines vollständigen Ratings werden weniger als 30 Minuten benötigt: ungefähr 12 Minuten für das Betrachten der Unterrichtsminiatur und weitere 12-15 Minuten für die Einschätzung der Items.

Anstelle einer expliziten Schulung der Rater, welche einen zusätzlichen Arbeitsaufwand von bis zu 50 Stunden bedeuten würde (Praetorius, 2014), ist es aufgrund des geringen Zeitaufwandes der Ratings lohnenswerter gewesen, mit vier Ratern eine größere Anzahl an Ratern einzusetzen. Zudem wurden die Rater durch Einbeziehen in die Endphase der Entwicklung des Manuals implizit auf ihre Tätigkeit vorbereitet. Falls möglich wurden sie in die Pilotierungen und Auswahl von Items einbezogen, sodass sie erste Einblicke gewinnen konnten.

Reliabilitäten der Ratings

Mit Hilfe des Manuals wurde die Unterrichtsqualität von 186 Unterrichtsminiaturen bzw. 93 Lehrpersonen eingeschätzt. Tabelle 2 zeigt die erzielten Skalen- und Beobachterreliabilitäten der vier Rater sowie die relativen Generalisierbarkeitskoeffizienten für die drei Basisdimensionen und den Durchschnittswert für alle 20 Subdimensionen.

	Kognitive Aktivierung	Konstruktive Unterstützung	Klassenführung	Mittelwert der Subskalen
Cronbachs α	.91	.87	.67	.88
ICC (2)	.88	.87	.83	.79
Rel. G-Koeffizient	.85	.77	.64	.78

Tab. 2: Reliabilitäten der Basis- und Subdimensionen

Mit der kognitiven Aktivierung und der konstruktiven Unterstützung werden bei allen drei Reliabilitätskriterien gute Werte erreicht. Lediglich die Klassenführung zeigt sich mit einem Wert für Cronbachs α von .67 und einem G-Koeffizienten von .64 als heterogene Skala: Die einzelnen Subskalen haben unterschiedliche Schwierigkeiten und werden von den Lehrpersonen in unterschiedlichem Maß erfüllt. Die niedrige Skalenreliabilität der Klassenführung lässt sich auf die Spezifität des Settings zurückzuführen, bei dem Bereiche des Klassengeschehens im Rahmen der Lehrveranstaltung vorstrukturiert sind und somit eine äußere Störungsprävention die Varianz zwischen den Lehrpersonen herabsetzt.

Auch auf Ebene der Subdimensionen werden die üblichen Cut-Off-Kriterien erfüllt. Im Durchschnitt erreichen sie Werte von etwa .80 bzw. im Falle der internen Konsistenz von .90. Mit Ausnahme der *Diagnose von Verständnisschwierigkeiten* mit einer ICC(2) von .60 erreichen sämtliche Subdimensionen zufriedenstellende Reliabilitäten. Hier konnte mithilfe der Generalisierbarkeitstheorie aufgedeckt werden, dass die schlechte Beobachterübereinstimmung auf Strenge- und Milde-Effekte zurückzuführen ist, bei der einzelne Rater alle Lehrpersonen gleichermaßen besser bzw. schlechter bewerten.

Weiterentwicklung: Kurzfragebogen

Mit Hilfe der Generalisierbarkeitstheorie wurde weiterführend eine Kurzversion des Manuals entwickelt. Diese soll während der Lehrveranstaltung genutzt werden, um die Unterrichtsqualität durch die etwa neun hospitierenden Peers einschätzen zu lassen. Unter Berücksichtigung dieser erhöhten Rateranzahl konnte im Rahmen einer Entscheidungsstudie (vgl. Brennan, 2011) die für die einzelnen Subdimensionen notwendige Itemanzahl berechnet und im Zuge von Optimierungen stark gekürzt werden. Der so entwickelte Fragebogen wurde bereits pilotiert und verspricht eine reliable Erhebung bei einer benötigten Ausfüllzeit von lediglich vier Minuten. Es gilt zu überprüfen, inwiefern dieser neu entwickelte Peer-Fragebogen das Videorating überflüssig werden lässt.

Resümee

Das Manual kann die gesetzten Ziele in einem sehr zufriedenstellenden Maß erfüllen. So ist es möglich, in einer relativ kurzen Ratingzeit mit vier Ratern die Qualität von Physikunterrichtsminiaturen auf Subdimensionsebene, also mit einer hohen Auflösung, reliabel zu erfassen.

Die so gewonnenen Daten zur Unterrichtsqualität wurden bereits in Regressionsanalysen mit Daten der professionellen Kompetenz der Lehrpersonen zusammengeführt. Dabei konnten tatsächlich Effekte der Lehrerüberzeugungen und des Fachwissen auf die Unterrichtsqualität nachgewiesen werden, welche in Krüger, Szogs und Korneck (in diesem Band) näher diskutiert werden.

Literatur

- Baumert, J., Blum, W., Brunner, S., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M., & Tsai, Y.-M. (2009). COACTIV Dokumentation. Berlin: MPI für Bildungsforschung.
- Brennan, R. L. (2011). Generalizability Theory and Classical Test Theory. *Applied Measurement in Education*, 24, 1-21.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I. "Aufgabenkultur" und Unterrichtsgestaltung: In E. Klieme, J. Baumert, P. Baptist, W. Blum, W. Bos, J. Doll et al. (Hrsg.), TIMSS - Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen ; Praxisberichte und Video-Dokumente (BMBF publik, Stand: Sept. 2001, S. 43-57). Bonn.
- Korneck, F., Oettinghaus, L., Kunter, M. & Redinger, R. (2016). Überzeugungen und Handlungen von Lehrpersonen - Messung von Unterrichtsqualität in komplexitätsreduzierten Settings des Physikunterrichts. In Rauin, Herrle, & Engartner (Hrsg.), Videoanalysen in der Unterrichtsforschung - Methodische Vorgehensweisen und aktuelle Anwendungsbeispiele. Weinheim: Beltz Juventa
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), Professionelle Kompetenz von Lehrkräften - Ergebnisse des Forschungsprogramms COACTIV (S. 85–113). Münster: Waxmann.
- Lotz, M., Lipowsky, F., Faust, G. (Hrsg.) (2013). Dokumentation der Erhebungsinstrumente des Projekts "Persönlichkeits- und Lernentwicklung von Grundschulkindern" (PERLE). 3. Technischer Bericht zu den PERLE-Videostudien. Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Praetorius, A.-K. (2014). Messung von Unterrichtsqualität durch Ratings (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 90). Münster: Waxmann.
- Seidel, T., Prenzel, M., Duit, R., & Lehrke, M. (Hrsg.) (2003). Technischer Bericht zur Videostudie Lehr-Lern-Prozesse im Physikunterricht". Kiel: IPN
- Vogelsang, C. (2015). Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von Physiklehrkräften - Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz. Berlin: LOGOS-Verlag