

Stabilität der metakognitiv-diskursiven Unterrichtsqualität zwischen den Unterrichtsstunden in einer Klasse

Bei Untersuchungen zur Lernwirksamkeit von Unterricht wird das Interesse zunehmend auf die Qualität und Wirkmechanismen der Unterrichtskommunikation gerichtet (vgl. Cohors-Fresenborg, Nowińska, Rott, in diesem Band). Dies betrifft insbesondere das Praktizieren von Metakognition. Um Mechanismen der Lernwirksamkeit von Metakognition im Klassengesprächen herausarbeiten zu können, wurde in der Arbeitsgruppe „Kognitive Mathematik“ der Universität Osnabrück ein Ratingsystem zur Analyse und Beurteilung metakognitiver und diskursiver Aktivitäten im Klassengespräch entwickelt (Nowińska, 2016). Die Beurteilung soll eine *zuverlässige* Einschätzung der metakognitiv-diskursiven Unterrichtsqualität in einer Klasse ermöglichen. Damit ist sowohl die *Reliabilität* der Beurteilungen gemeint wie auch die Erfassung der für eine Klasse charakteristischen Ausprägung von Metakognition und Diskursivität und – damit verbunden – die *Generalisierbarkeit* der Urteile. Zur Gewährleistung der Reliabilität (im Sinne der Interrater-Übereinstimmung) ist eine Rater-Qualifizierung erforderlich. Der zweiten Anforderung kann nur dann Rechnung getragen werden, wenn die Variabilität der zu beurteilenden Merkmale zwischen den Unterrichtsstunden in einer Klasse angemessen beachtet wird.

Im Folgenden werden zunächst die mit dem Ratingsystem zu beurteilenden Merkmale der Ausprägung der metakognitiv-diskursiven Unterrichtsqualität vorgestellt. Anschließend werden Ergebnisse einer Untersuchung zur Variabilität der Ausprägung dieser Merkmale präsentiert. Eine ausführliche Diskussion der Ergebnisse findet sich in Nowińska und Praetorius (2017).

Das Ratingsystem zur Beurteilung metakognitiv-diskursiver Unterrichtsqualität

Das entwickelte Ratingverfahren ist zweischrittig: Im *ersten Schritt* werden metakognitive, diskursive und negativ diskursive Aktivitäten mithilfe eines Kategoriensystems (dargelegt u.a. in Nowińska, 2016, S. 51) klassifiziert. Im *zweiten Schritt* werden diese Aktivitäten bezüglich ihrer Qualität beurteilt. Die Beurteilung fokussiert Merkmale der Ausprägung von Metakognition im Klassengespräch, die aufgrund theoretischer Überlegungen und empirischer Befunde das Lernverhalten von Schülerinnen und Schülern langfristig beeinflussen können. Diese wurden zu sieben Beurteilungsdimensionen ausgebaut und sind jeweils auf einer mehrstufigen Skala einzuschätzen. Im Kern jeder Beurteilungsdimension steht eine *Leitfrage*, die die

Aufmerksamkeit des Raters auf die ausgewählten Qualitätsmerkmale richtet, und aus eine Reihe von Antwortmöglichkeiten, die eine Skala zur Beurteilung dieser Merkmale bildet.

Leitfrage 1 fokussiert das Ausmaß, in dem von den Lernenden und von der Lehrkraft metakognitive Aktivitäten mit erkennbarem Bemühen um eine elaborierte Auseinandersetzung mit den anstehenden „Inhalten“ praktiziert werden. Die Abstufung der Antwortmöglichkeiten orientiert sich u.a. daran, inwieweit die Lernenden selbstständig (ohne direkte Aufforderung seitens der Lehrkraft) solche Aktivitäten praktizieren bzw. dazu angeleitet werden.

Leitfrage 2 dient der Einschätzung des Bemühens um eine argumentative Auseinandersetzung im Gespräch. Die Abstufung der Antwortmöglichkeiten orientiert sich daran, inwieweit es erkennbar ist, dass Begründungen in Kombination mit metakognitiven Aktivitäten zu der Unterrichtskultur gehören und von den Lernenden auch selbstständig formuliert werden.

Mit der 3. Leitfrage wird der Stellenwert der metakognitiven Schüler- und Lehreraktivitäten für Erkenntnisprozesse und Einsichten bezüglich eines verständigen Umgangs mit den anstehenden fachspezifischen „Inhalten“ fokussiert. Die Abstufung der Antwortmöglichkeiten orientiert sich danach, in welchem Ausmaß solche Erkenntnisse bei Einzelnen oder im Diskurs mit mehreren Beteiligten erkennbar sind.

Mit der 4. Leitfrage wird beurteilt, inwieweit eine diskursive, an den Denkprozessen der Lernenden orientierte Auseinandersetzung angestrebt wird. Dazu gehört auch ein Urteil über Konsequenzen aus dem Fehlen diskursiver Aktivitäten, die notwendig wären, um Vorstellungen der Lernenden oder Unterschiede zwischen Dargestelltem und Gemeintem zu elaborieren.

Komplementär dazu wird in der 5. Leitfrage beurteilt, inwieweit Negative Diskursivität (d.h. Verstöße gegen die Diskursivität) das Unterrichtsgespräch, das Verstehen der anstehenden „Inhalte“ und die adäquate Benutzung der Fachsprache erschwert.

Mithilfe der 6. Leitfrage wird beurteilt, ob im Unterrichtsgespräch diskursive, argumentativ stringent geführte „Debatten“ angestrebt werden, die in der Abstufung von sehr kurzen Gesprächen unter den Lernenden (ohne Einbringung seitens der Lehrkraft) bis hin zu längerer Auseinandersetzung unter ihnen variieren können.

Die 7. Leitfrage dient der Beurteilung des Ausmaßes, in dem metakognitive und diskursive Aktivitäten für eine gedankliche oder sprachliche Präzisierung bei der Auseinandersetzung mit einer anspruchsvollen Fragestellung praktiziert werden.

Stabilität der metakognitiv-diskursiven Unterrichtsqualität

Zur Untersuchung der Variabilität zwischen Unterrichtsstunden einer Klasse in Bezug auf die mit den Leitfragen einzuschätzenden Merkmale wurde die Generalisierungstheorie herangezogen (Shavelson & Webb, 1991). Konkret wurden Entscheidungsstudien durchgeführt, die eine Schätzung der benötigten Anzahl an Unterrichtsstunden pro Klasse für eine zuverlässige Beurteilung der jeweiligen Ratingdimension ermöglichen. Im Unterschied zur klassischen Testtheorie erlaubt die Generalisierbarkeitstheorie, die Bedeutung verschiedener Varianzquellen, die für Unterschiede in Ratings verantwortlich sein können, differenziert zu erfassen. Im vorliegenden Fall wurden folgende Varianzquellen betrachtet: Lehrkraft/ Klasse (l), Unterrichtsstunde genestet in Lehrkraft/ Klasse (u:l), Rater (r), Interaktion zwischen Rater und Lehrkraft/ Klasse (r*l) und Residualvarianz (r*(u:l), e).

Der Datensatz bestand aus 24 videographierten Unterrichtsstunden (6 Klassen à 4 Unterrichtsstunden). Aus jeder wurde ein 10-minütiges Klassengespräch zur Kernfragestellung des Unterrichts ausgewählt und von drei geschulten Ratern mithilfe der sieben Leitfragen beurteilt.

Leitfragen		LF 1	LF 2	LF 3	LF 4	LF 5	LF 6	LF 7
Lehrkraft /Klasse	l	89	45	60	52	71	50	6
Unterricht: Lehrkraft	u:l	0	46	21	39	22	30	14
Rater Bias	r	0	1	0	0	0	0	5
Rater*Lehrkraft	r*l	3	0	0	2	1	3	51
Fehler	r*(u:l); e	8	9	19	8	7	16	25
Relative G-Koeffizienten		0,98	0,78	0,90	0,83	0,92	0,83	0,38

Tab. 1: Relative G-Koeffizienten und Varianzzerlegung (in %) (4 Unterrichtsstunden pro Lehrkraft, 3 voneinander unabhängig beurteilende Rater)

Tabelle 1 zeigt die Ergebnisse der Generalisierbarkeitsstudien zur Identifikation der Größenordnung der jeweiligen prozentualen Varianzanteile. Die relativen Generalisierbarkeitskoeffizienten geben Auskunft über die Zuverlässigkeit der Urteile und können analog zu Reliabilitätskoeffizienten der klassischen Testtheorie interpretiert werden (Daumenregel $>.70$). Bis auf Leitfrage 7 weisen die Koeffizienten sehr gute Ausprägungen auf. Entsprechend niedrig sind die Varianzkomponenten, die auf eine Verzerrung durch Rater Bias hindeuten (r; r:l). Aus dem Verhältnis der Werte l (stabiler Anteil über Unterrichtsstunden einer Lehrkraft/ Klasse hinweg) zu u:l (spezifischer Anteil einzelner Unterrichtsstunden) kann Rückschluss auf die Stabilität der jeweiligen Merkmale zwischen den Unterrichtsstunden in einer Klasse gezogen werden: Insbesondere die mit Leitfrage 1, aber auch Leitfragen 3, 5 und 6 beurteilten Merkmale variieren zwischen den Lehrkräften/ Klassen mehr als zwischen den Unterrichtsstunden einer Lehrkraft in

einer Klasse. Es handelt sich somit um vergleichsweise stabile Unterrichtsmerkmale.

Die auf der Basis der Generalisierbarkeitsstudien durchgeführte Entscheidungsstudien (vgl. Abb. 1) zeigen, dass über die Leitfragen 1 bis 6 hinweg bei 3 Ratern maximal 3 Unterrichtsstunden pro Klasse hinreichend wären, um zuverlässige Urteile (relativer G-Koeffizient $\geq .70$) der metakognitiv-diskursiven Unterrichtsqualität zu erhalten.

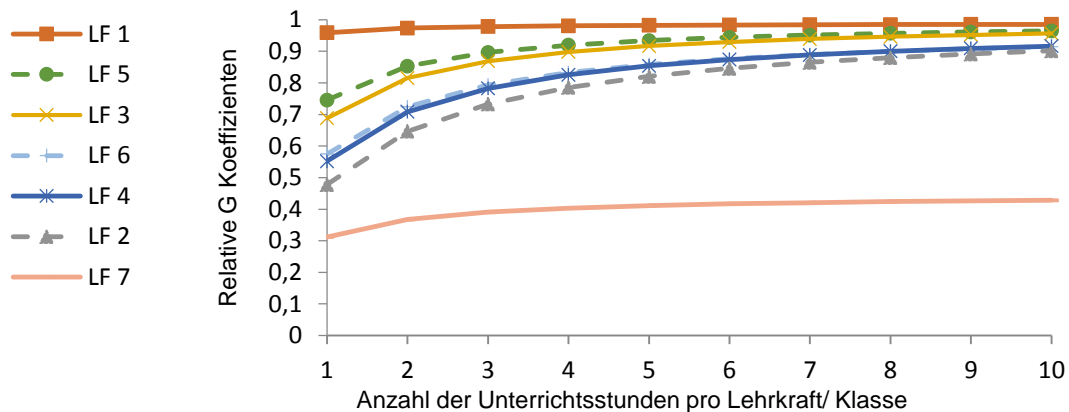


Abb. 1: Relative G-Koeffizienten für Entscheidungsstudien

Insgesamt lässt sich daher festhalten, dass mit dem Ratinginstrument ein praktikables Verfahren zur Beurteilung von Unterrichtsqualität entwickelt wurde. Um die Wirkmechanismen von Metakognition herauszuarbeiten, sind Zusammenhänge zwischen der mit dem Ratingsystem zu beurteilenden Unterrichtsqualität und der Leistungsentwicklung der Lernenden zu untersuchen. Eine solche Untersuchung steht noch aus.

Literatur

- Cohors-Fresenborg, E.; Nowińska, E. & Rott, B. (in diesem Band). Methodische und konzeptionelle Herausforderungen einer zuverlässigen Beschreibung und Beurteilung von Unterrichtsqualität.
- Nowińska, E. (2016). *Leitfragen zur Analyse und Beurteilung metakognitiv-diskursiver Unterrichtsqualität*. Osnabrück: Forschungsinstitut für Mathematikdidaktik.
- Nowińska, E. & Praetorius, A.-K. (2017). Evaluation of a Rating System for the Assessment of Metacognitive-Discursive Instructional Quality. Dooley, T. & Gueudet, G. (Eds.): *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education*, S. 3121-3128. Dublin, Ireland: DCU Institute of Education & ERME.
- Praetorius, A.-K., Lenske, G. & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, 22(6), 387-400.
- Shavelson, R. & Webb, N. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: Sage.