

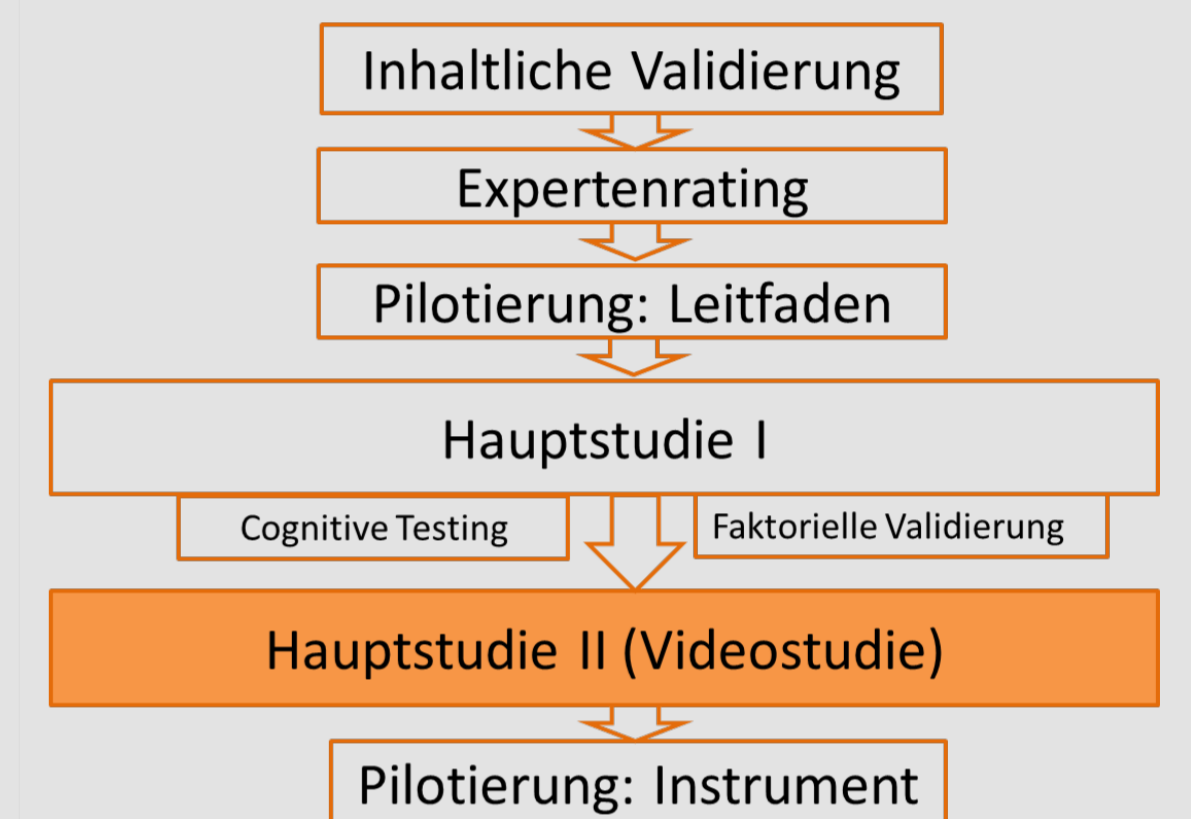
PROJEKTZIELE UND THEORETISCHER HINTERGRUND

Wissenschaft und Politik sind sich einig: Lehrkräfte benötigen Feedback, um das eigene Handeln zu professionalisieren und den eigenen Unterricht weiterzuentwickeln (Helmke, 2012). Eine ökonomische Variante, um sich Rückmeldung zum eigenen Unterricht einzuholen, stellt die Schülerbefragung dar. Auf Basis des aktuellen Forschungsstandes gelten Sekundarstufenschüler als kompetente Beurteiler von Unterrichtsqualität (Ditton, 2002; Gruehn, 2000; Hattie, 2009; Hofer, 1981). Das auf Klassenebene aggregierte Urteil erweist sich als reliabel, faktoriell valide und prädiktiv valide in Bezug auf die Leistungsentwicklung (Lüdtke, Trautwein, Schnyder & Niggli, 2007; Marsh, 2007; Clausen, 2002). Wie gut Grundschulerratings zur Messung von Unterrichtsqualität geeignet sind, wurde bislang nur in wenigen Arbeiten gezielt untersucht. Dennoch wird sowohl innerhalb von Forschungsprojekten als auch im Rahmen der internen und externen Schulevaluation auf Grundschulerratings rekurriert. Die vorliegende Arbeit befasst sich unter Einbezug aktueller Kenntnisse aus der Entwicklungspsychologie und der kognitiv fundierten Surveyforschung mit der grundlegenden Frage nach der Validität von Grundschulerratings über Unterrichtsqualität. Auf Basis der Erkenntnisse soll ein altersadäquates Feedbackinstrumentarium für die interne Evaluation an Grundschulen entwickelt werden.

METHODE(N)

Zentrale Forschungsfragen:

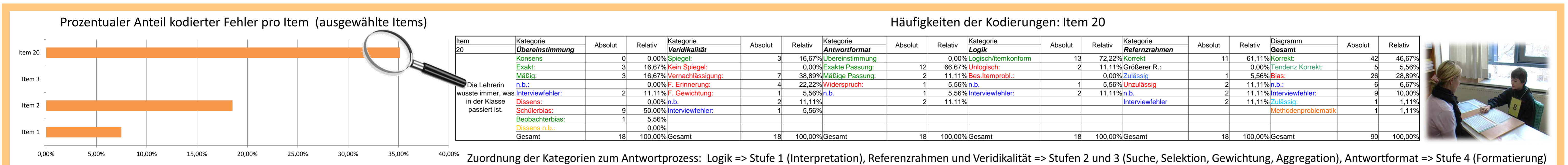
- (1) Verstehen Grundschüler Items aus gängigen Feedbackinstrumenten im intendierten Sinn?
- (2) Sagt eine inhaltliche Validierung durch Experten die empirische inhaltliche Itemvalidität voraus?
- (3) Sind Grundschulerratings zur Unterrichtsqualität struktural valide?
- (4) Genügen faktorielle Validierungen, um Items zu identifizieren, die Schüler nicht im intendierten Sinn beurteilen können?
- (5) Lassen sich bei der Einschätzung von Unterrichtsqualität Probleme im Urteils- bzw. Antwortprozess identifizieren?
- (6) Sind Grundschulerratings über eine konkrete Unterrichtsstunde veridikal?
- (7) Können Grundschüler ausgewählte Merkmale von Unterrichtsqualität per Fragebogen hinreichend valide einschätzen?



Um diese Fragen zu beantworten wurde ein multi-method Design gewählt, welches 3 Vor- und 2 Hauptstudien umfasst. Das aus der Arbeit resultierende Feedbackinstrument wurde darüber hinaus abschließend pilotiert. Die Gesamtstichprobe umfasst ca. 800 Dritt- und Viertklässler sowie 11 Experten.

Das Poster ergänzt den heutigen Vortrag und fokussiert auf **Hauptstudie II**: Eine Unterrichtsstunde einer vierten sowie einer dritten Klasse wurden videografiert. Im Anschluss an die Stunde füllten die Schüler (20 pro Klasse) einen Fragebogen zur Unterrichtsqualität aus und wurden nach jedem Kreuz zu ihrer Begründung gefragt (Kognitive Interviews, Technik Probing). Unter Einbezug der Videos wurde die Güte der Begründungen kodiert (Perspektive, Referenzrahmen, Veridikalität, Beurteilerübereinstimmung, Logik). Die Intercoder-Reliabilität erwies sich pro Kategorie als gut bis ausgezeichnet (Cohens Kappa: .75-.96).

AUSGEWÄHLTE ERGEBNISSE



Auf allen Stufen des Antwortprozesses nach Tourangeau und Kollegen (2000) wurden Fehler identifiziert, wobei das Fehlerausmaß pro Item stark variierte (5). Der geringste Fehleranteil fiel auf die Kategorie Perspektive (1%), der größte Fehleranteil auf die Kategorie Veridikalität (16%). Die Veridikalität der Grundschülerurteile hing vom zu beurteilenden Aspekt ab. Besondere Probleme bereiteten Items, welche die permanente Beobachtung der Lehrkraft erfordern (z. B. Item 20), sie konnten nicht veridikal beurteilt werden. Nach Abschluss von Hauptstudie II lässt sich auf Basis der gesamten Studien sagen, dass Grundschüler der dritten und vierten Klassenstufe durchaus in der Lage sind, zu verschiedenen Aspekten der Qualitätsbereiche *Klassenführung, lernförderliches Klima, Motivierung, Klarheit und Strukturiertheit und Aktivierung* ein valides Urteil abzugeben (7). Eine umfassende Beurteilung ist ihnen jedoch noch nicht möglich.

DISKUSSION UND AUSBLICK

Hauptstudie II verdeutlichte, dass Items, welche verständlich für die Schüler sind, dennoch zu einem invaliden Urteil führen können. Somit ist die Verständlichkeit von Items zwar eine notwendige Voraussetzung für ein valides Urteil, aber keine hinreichende Bedingung. Bei Validierungen durch Cognitive Testing sollte deshalb nicht nur die Verständlichkeit der Items, sondern der gesamte Antwortprozess überprüft werden. Des Weiteren sollte angesichts des vergleichsweise hohen Fehleranteils in der Kategorie Veridikalität unbedingt die Realitätstreue der Antworten untersucht werden. Das Endprodukt der Arbeit bildet ein umfangreich validiertes Instrumentarium, welches Lehrkräfte zur evidenzbasierten Unterrichtsdiagnostik im Rahmen der internen Evaluation einsetzen können. Das Instrument ist online verfügbar unter: www.unterrichtsdiagnostik.info.

REFERENZEN

Clausen, M. (2002). Unterrichtsqualität: eine Frage der Perspektive? Münster: Waxmann.
 Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. Zeitschrift für Pädagogik, 48 (2), 262–286.
 Gärtner, H. (2010). Wie Schülerinnen und Schüler ihre Lernumwelt wahrnehmen. Zeitschrift für Pädagogische Psychologie, 24 (2), 111–122.
 Gruehn, S. (2000). Unterricht und schulisches Lernen. Schüler als Quelle der Unterrichtsbeobachtung. Münster: Waxmann.
 Hattie, J. (2009). Visible learning. A synthesis of over 800 meta-analyses relating to achievement. London, UK: Routledge.
 Helmke, A. (2012). Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts (4. Aufl.). Seelze: Klett-Kallmeyer.
 Hofer, M. (1981). Lehrerverhalten aus der Sicht der Schüler. Pädagogische Welt, 35 (1), 49–56.
 Lüdtke, O., Trautwein, U., Schnyder, I. & Niggli, A. (2007). Simultane Analysen auf Schüler- und Klassenebene. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 39 (1), 1–11.
 Marsh, H. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. Perry & J. S. Smart (Hrsg.), The scholarship of teaching and learning in higher education: An evidence-based perspective (S. 319–383). Netherlands: Springer.

STAND DER ARBEIT

