

## PROJEKTZIELE UND THEORETISCHER HINTERGRUND

Ratings externer Beobachter werden oft als „Königsweg“ zur Erfassung von Unterrichtsqualität beschrieben (Clare et al. 2001; Helmke, 2010; Pianta & Hamre, 2009). In der Unterrichtsforschung existieren bislang allerdings nur wenige Arbeiten, die sich dezidiert mit der Eignung solcher Beobachterratings zur Messung von Unterrichtsqualität beschäftigen (Pietsch & Tosana, 2008; Praetorius et al., 2012). In der Dissertation wurde mittels fünf empirischer Studien die Reliabilität und Validität von Beobachterratings thematisiert.

### Forschungsfragen:

- (1) In welchem Ausmaß sind hoch-inferente Beobachterratings zu Unterrichtsqualität durch Raterfehler verzerrt? (Studie 1)
- (2) **Sind Unterrichtseinschätzungen geschulter Beobachter reliabler und valider als solche von ungeschulten Ratern? (Studie 2)**
- (3) In welchem Ausmaß sind Unterrichtsratings geschulter Beobachter durch situationale Merkmale beeinflusst? (Studie 3)
- (4) Lassen sich bei der Einschätzung von Unterrichtsqualität Probleme im Urteilsprozess der Rater identifizieren? (Studie 4)
- (5) **Wie stabil sind Merkmale von Unterrichtsqualität über Unterrichtsstunden hinweg? (Studie 5)**

## STUDIE 2

### Hypothesen (Auswahl):

H<sub>2,3</sub>: Die Anzahl an Ratern, die für eine hinreichend reliable Messung der Unterrichtsdimensionen Klassenführung und Schülerorientierung notwendig ist, ist bei ungeschulten Ratern höher als bei geschulten Ratern.

H<sub>2,5</sub>: Die Begründungen von Ratings zu den Dimensionen Klassenführung und Schülerorientierung weisen bei geschulten Ratern einen höheren Differenzierungsgrad auf als bei ungeschulten Ratern.

### Stichprobe:

- 12 trainierte Rater (11 weiblich; Lehramtsstudierende; Ratertraining: 8 h, vgl. Hoyt & Kerns, 1999)
- 11 Lehrkräfte (55% weiblich, Berufserfahrung  $M = 17$ ,  $SD = 11$ )
- 197 Lehramtsstudierende 1.Sem. (75% weiblich, Alter  $M = 21$ ,  $SD = 3$ )
- 79 Lehramtsstudierende 4.Sem. (85% weiblich, Alter  $M = 23$ ,  $SD = 5$ )

**Stimuli:** 3 Videosequenzen (10 min.), Deutschunterricht, 4. Jahrgangsstufe

**Instrument:** hoch-inferentes Ratinginstrument EMU (Helmke et al., 2011), 4-stufig; Auswahl: Klassenführung (5 Items), Schülerorientierung (7 Items)

**Analysen:** Generalisierbarkeitstheorie (Brennan, 2001); Qualitative Inhaltsanalyse (Mayring, 2008)

### Ergebnisse (Auswahl):

- Reliabilität der Ratings bei geschulten Ratern nicht höher ausgeprägt als bei ungeschulten Ratern
- höhere Validität der Einschätzungen der geschulten Rater (höherer Differenzierungsgrad und engerer Bezug zum Ratingmanual der Begründungen)

**Diskussion:** unerwartet hohe Reliabilität ungeschulter Rater

- Vermutung: einheitliche Sichtweise aufgrund von ähnlichen Schulerfahrungen
- aber: reliabel  $\neq$  valide (vgl. auch Strong et al., 2011)
- bisherige Praxis zur Bestimmung der Effektivität von Trainings unzureichend

Entwicklung von Effektivitätskriterien in Bezug auf Ratertrainings nötig

## STUDIE 5

### Fragestellungen (Auswahl):

F<sub>5,2</sub>: Wie viele Unterrichtsstunden pro Lehrkraft sind nötig, um die drei Basisdimensionen von Unterrichtsqualität hinreichend zuverlässig zu erfassen?

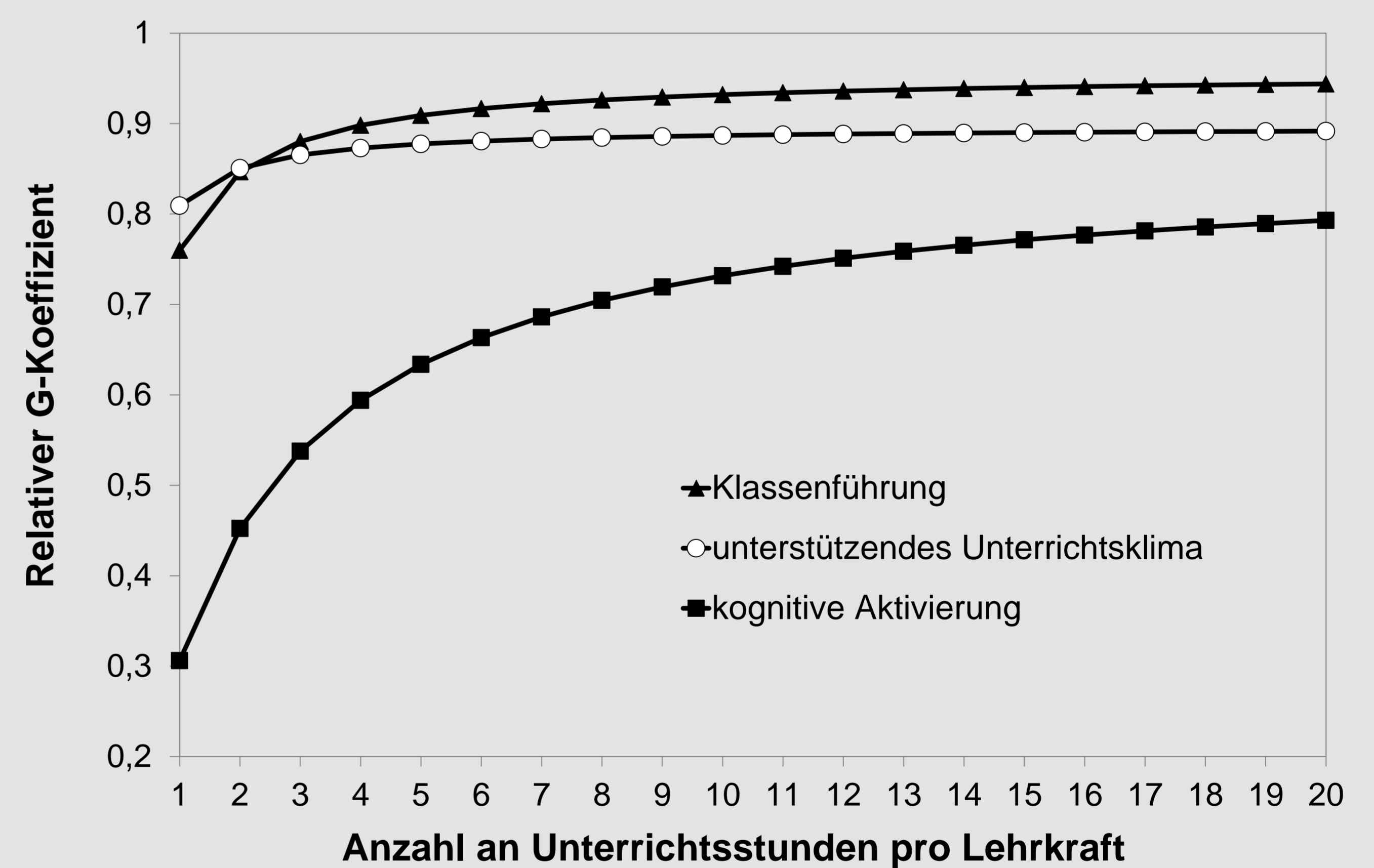
### Stichprobe:

- Pythagoras-Videostudie (s. Hugener et al., 2006)
- 38 Lehrkräfte à 5 Unterrichtsstunden
- mdst. 2 Rater

**Instrument:** 3 Basisdimensionen von Unterrichtsqualität (s. Rakoczy & Pauli, 2006): Klassenführung (2 Items), Schülerorientierung (3 Items), kognitive Aktivierung (3 Items); Items jeweils 4-stufig

**Analysen:** Generalisierbarkeitstheorie

### Ergebnisse (Auswahl):



**Diskussion:** hohe Instabilität kognitiver Aktivierung

- mögliche Ursachen: unklare Definition des Konstrukts; nur bestimmte Aspekte kognitiver Aktivierung beobachtbar

hohe Anzahl an Unterrichtsstunden oder alternative Operationalisierung notwendig

## REFERENZEN

Clare, L., Valdés, R., Pascal, J. & Steinberg, J. (2001). *Teachers' assignments as indicators of instructional quality in elementary schools* (CSE Technical Report Nr. 545). Los Angeles: National Center for Research on Evaluation; Brennan, R. L. (2001a). *Generalizability theory*. New York, NY: Springer; Helmke, A. (2009). Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts. Seelze: Klett-Kallmeyer; Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W. et al. (2011). *EMU - Unterrichtsdiagnostik*. Landau: Universität Koblenz-Landau; Hoyt, W. T. & Kerns, M.-D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, 4, 403-424; Hugener, I., Klieme, E., Pauli, C. & Reusser, K. (Hrsg.). (2006). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie "Unterrichtsqualität, Lehrerverhalten und mathematisches Verständnis"*. 3. Videoanalysen. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung (DIPF); Mayring, P. (2008). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (10. Aufl.). Weinheim: Beltz; Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38, 109-119; Pietsch, M. & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität: Das Multifacetten-Rasch-Modell und die Generalisierbarkeitstheorie als Methoden der Qualitätssicherung in der externen Evaluation von Schulen. *Zeitschrift für Erziehungswissenschaft*, 11, 430-452; Praetorius, A.-K., Lenske, G. & Helmke, A. (2012). Observer ratings of instructional quality - Do they fulfill what they promise? *Learning and Instruction*, 6, 387-400; Rakoczy, K. & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In I. Hugener, E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie "Unterrichtsqualität, Lehrerverhalten und mathematisches Verständnis"*. 3. Videoanalysen (S. 206-233). Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung (DIPF); Strong, M., Gargani, J. & Hacifazlioglu, O. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, 62, 367-382.

## STAND DER ARBEIT

2009	2010	2011	2012
Theoretische Auseinandersetzung	Durchführung + Auswertung von Vorstudien	Datenerhebung + Auswertung Studien 1-4	Auswertung Studien 1-5; Verfassen der Monografie

September 2009  
Beginn der Promotion

August 2012  
Abgabe der Dissertation

