



Leibniz - Netzwerk
Unterrichtsforschung



LEUPHANA
UNIVERSITÄT LÜNEBURG

Wirkung von Unterrichtsqualität: Bisherige Forschungsbefunde und methodische Herausforderungen

- Diskussion -

Marcus Pietsch



Vortrag 1: Klieme et al. - Zusammenfassung

Zusammenschau von empirischen Arbeiten zum Modell der drei Basisdimension (aka Klieme-Modell).

Befunde zeigen,

- a) Seit 2001 im Einsatz
- b) Vermehrt Studien seit 2013
- c) Fast immer im Fach Mathematik
- d) Überwiegend im Sekundabereich
- e) Perspektiven ausgewogen, jedoch nur selten alle 3 simultan
- f) Formulierungen von (Sub-)Dimensionen und Items teils sehr heterogen
- g) Nicht alle erwarteten Effekte werden immer und konsistent ermittelt
- h) Studiendesigns variieren (teils) sehr stark



Vortrag 1: Klieme et al. - Auffälligkeiten

- Mehrebenenmodelle sind State-of-the-Art.
- Faktorielle Struktur nur selten untersucht (insb. Mehrebenen).
- Zeitliche Stabilität erst in ganz aktuellen Studien untersucht.
- Keine Befunde zur Stabilität über längeren Zeitraum; insbesondere Veränderung bei Schülerratings diesbezüglich nicht vorhanden.
- Theoretischer Rahmen eher nicht vorhanden (bzw. post-hoc).
- ICC2 (teilweise) niedrig = Unterrichtsklima?
- Messmodelle und Methoden variieren (teils) stark und könnten Erklärung für heterogene Ergebnisse (auch) sein.



Vortrag 1: Klieme et al. – Offene Fragen

“We excluded the measure of classroom management in the study of Taut and Rakoczy (2016), as the G coefficient of 0.05 indicates a severe problem with raters, scales, or the analyses”

- Raten ‘normale’ Rater ggf. anders als trainierte Videorater`?
- Video vs. in Vivo?
- Reporting für die Praxis auf Ebene von Subdimensionen sinnvoll (z.B. PMRSE)?
- Wie steht es um Wechselwirkungen von Unterricht und Leistung?
- Wie sieht es mit Cross-Level-Effekten aus?

Methodische Anmerkung: Auch bei einer Übersichtsarbeit wäre eine standardisierte Vorgehensweise wünschenswert, z.B. Cochrane!



Vortrag 2: Fischer et al. - Zusammenfassung

Untersuchung zur Messinvarianz der PISA-Unterrichtsskalen in und zwischen Sprachgruppen.

Befunde zeigen,

- a) Alle Skalen sind zwischen Sprachgruppen nicht messinvariant (Cut-Off: invariante Parameter 25%, Unterstützung=24%)
- b) Innerhalb der chinesischen Sprachgruppe sind alle Skalen nicht messinvariant
- c) Besonderes problematisch „kognitive Aktivierung“



Vortrag 2: Fischer et al. - Auffälligkeiten

- Einsatz einer aktuellen (und angemessenen) Methode zur Untersuchung von MI (Alignment, intr. ~ 2014).
- Durch die gewählte Methode wird deutlich, welche Gruppe zur non-invariance beiträgt.
- Bisläng jedoch kaum Studien, die Ansätze zur Überprüfung von MI (z.B. Cluster Bias, Jak et al. 2013) vergleichen.



Vortrag 2: Fischer et al. – Offene Fragen

- Unklar: Welche Items tragen besonders zur non-invariance bei? Wurde eine Monte Carlo Studie durchgeführt (Muthen 2014)?
- Was bedeuten die Befunde für internationale LSA? Valide Vergleiche?
- Insbesondere auch für Zusammenhänge mit AV (Leistung)?
- Äußerst wünschenswert zu erfahren, warum Unterschiede vorhanden sind:
 - Conceptual meaning or understanding of the construct
 - Extent of social desirability or social norms,
 - Different reference points (when making statements about themselves)
 - Responses to extreme items are different,
 - Particular items are more applicable for one group than another
 - Translation of one or more item is improper (Chen, 2008)



Vortrag 3: Köhler et al. - Zusammenfassung

Veranschaulichung von Modellierungsmöglichkeiten, deren Chancen und Grenzen, bei 2 Pre-Post-Test-Desings

Befunde zeigen:

- a) Befunde zu Effekten von Unterrichtsmerkmalen variieren mit Fragestellung bzw. Modell (latent Growth/Change vs. varianzanalytisch)
- b) Effekt auf Veränderung größer als auf AV zu t1



Vortrag 3: Köhler et al. - Auffälligkeiten

- Klarer Bezug zu Marsh/Lüdtke Context-Culture-Diskussion
- Modelle stets als doubly-latent modelliert, daher Kontrolle für Mess- und Stichprobenfehler
- Konstrukt „Unterstützung“ als Kontextkonstrukt



Vortrag 3: Köhler et al. – offene Fragen

- Was passiert bei Modellierung als
 - manifest-latent und
 - manifest-manifest?
- Was passiert bei Nicht-Modellierung von Ebene 1 bzw. 2?
- Welchen Effekt hat die geringe ICC2 der Unterrichtsvariablen ($<.70$), „*a low ICC together with small samples of L1 individuals from each L2 group will affect the bias of contextual effect considerably*“ (Marsh et al. 2008). Hat sie überhaupt einen Effekt?
- Interpretation Level 1 bei Kontextkonstrukten einfach. Aber was ist mit Kulturkonstrukten (Abweichung in der Wahrnehmung der Person vom Schul/Klassenmittel.)?



Zusammenfassung

- Theoretische Fundierung unklar.
- Analysen zunehmend methodisch elaboriert.
- Insgesamt sehr heterogene Forschungslage – sowohl beim Erhebungsdesign als auch bei Analyse usw.
- Anscheinend kulturelle Unterschiede.

Daher:

- Theoretische Grundlagen klären (woher kommen TIMSS `95/BIJU-Skalen, warum gibt es die?)
- Vorgehen bei eigenen Studien stets transparent machen
- Ggf. auch mit unterschiedlichen Maßen je Ebene arbeiten:
 - L1: „Wenn ich Hilfe brauche, bekomme ich sie von meinem Englischlehrer/ meiner Englischlehrerin.“
 - L2: „Unser/e Englischlehrer/in hilft, wenn jemand in der Klasse Hilfe braucht.“
- Ideal: Set von Items entwerfen, die jeweiligen Anlässen gerecht werden. -> **Vom Framework zum Instrumentenpool?**