

Data Sharing in der Long-Tail der empirischen Bildungsforschung

Nutzungspotenziale und Limitationen von Sekundäranalysen: Eine Fallstudie am Beispiel Epistemologischer Überzeugungen

Martin Kerwer, Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)
5. Tagung der Gesellschaft für empirische Bildungsforschung, 12.-15. März, Heidelberg

Data Sharing

- Fächerübergreifend nimmt die Forderung nach der Bereitstellung von Forschungsdaten zu. **Gründe** dafür sind...
- Qualitätssicherung. In Folge der Replikationskrise soll die Nachprüfbarkeit von Forschungsergebnissen durch die Bereitstellung der Daten gesteigert werden.
 - Forschungsökonomie. Effizienterer Einsatz von Ressourcen, die in Forschung investiert werden.
-> Sekundärforschung auf Basis bereitgestellter Forschungsdaten

Fallstudie - Epistemologische Überzeugungen

Schritt 1: Inhaltliche Fragestellung

- Differenzielle Ausprägung epistemologischer Überzeugungen während des Hochschulstudiums in verschiedenen Fachdisziplinen
- Epistemologische Überzeugungen = erlebte Strukturiertheit des Wissens einer Disziplin
- Wichtiger Prädiktor des Wissenserwerbs im Studienverlauf

➤ **Fragestellung 1:** Welche Unterschiede epistemologischer Überzeugungen gibt es zwischen Disziplinen?

➤ **Fragestellung 2:** Unterscheiden sich psychometrische Eigenschaften von Verfahren zur Erfassung epistemologischer Überzeugungen zwischen Disziplinen?

Schritt 2: Identifikation eines geeigneten Messinstruments

CAEB (Stahl & Bromme, 2007): Etabliertes Verfahren zur Erfassung epistemologischer Überzeugungen auf den Dimensionen *Variability* und *Texture*.

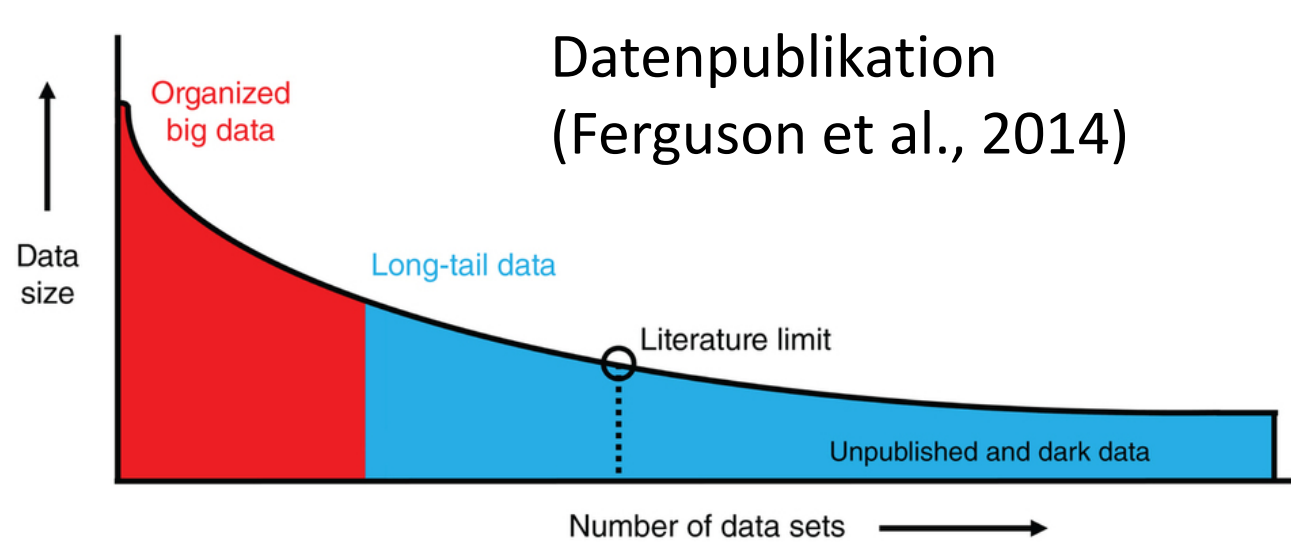
Schritt 3: Identifikation geeigneter Datensätze

Zwei öffentlich zugängliche Datensätze im FDZ PsychData:

(1) Merk und Bohl (2016): Querschnitt von Lehramtsstudierenden zur Bewertung der Pädagogik und ihrer Fachwissenschaft (n = 198),

(2) Mayer, Rosman, Birke, Gorges und Krampen (in press): Längsschnitt von Informatik- (n = 89) und Psychologiestudierenden (n = 137)

... in der empirischen Bildungsforschung

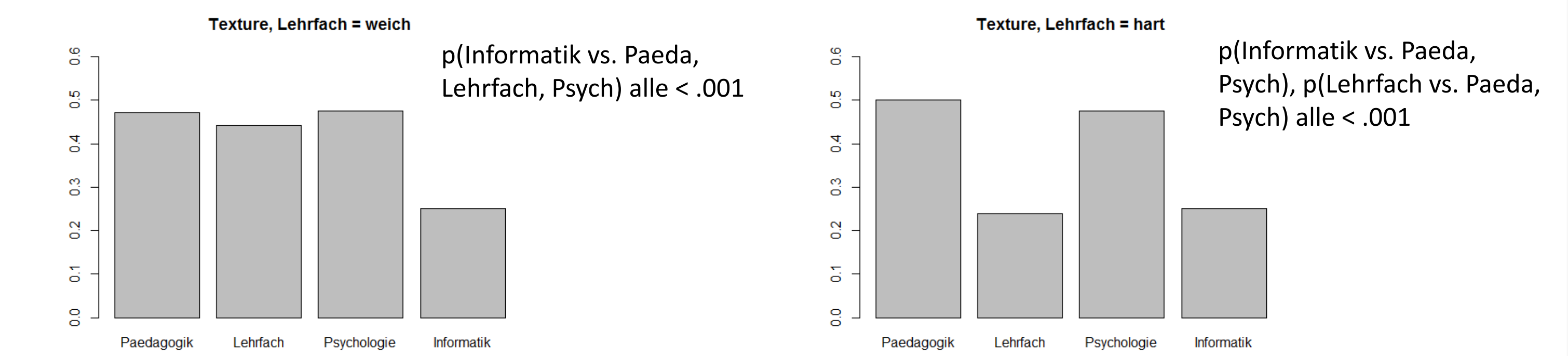
- 
- Nachnutzung von Daten aus großen Panelstudien weitverbreitet (NEPS, PISA,...)
 - Nachnutzung von Daten aus kleineren Projekten selten
 - Potenzial von Long-Tail-Daten jenseits der Qualitätssicherung zur Beantwortung neuer Fragestellungen durch Sekundäranalysen wird bisher kaum betrachtet
 - **Fallstudie zur Identifikation von Potenzialen und Problemstellungen**

Methode

- Analyse der Mittelwertsunterschiede (Fragestellung 1) über Regressionsanalysen und t-tests
- Analyse der psychometrischen Eigenschaften über Mehrgruppen-Strukturgleichungsmodelle (Fragestellung 2)

Ergebnisse (Fragestellung 1)

Bestätigung bestehender Befunde zwischen *weichen* und *harten* Disziplinen (Muis, Bendixen, & Haerle, 2006)



Ergebnisse (Fragestellung 2)

Item: *Das in den ... behandelte Wissen ist: ausgehandelt – entdeckt*

Lehrfach gesamt (0.378, $p < 0.01$), Lehrfach weich (0.263, $p = 0.136$), Lehrfach hart (0.618, $p < 0.05$), Pädagogik (-0.051, $p = 0.807$), Psychologie (0.369, $p < .05$), Informatik (-0.414, $p = 0.182$)

-> Items scheinen je nach Disziplinen unterschiedlich zu funktionieren

Lessons Learned...

- Leicht lösbare Probleme – durch den Datennehmer selbst

➤ uneinheitliche Skalen (5 vs. 7-stufige Skala) -> Vereinheitlichung
- Mittelschwer lösbare Probleme – durch Kontakt zu Datengeber

➤ im Rahmen der Anonymisierung zur Beantwortung der Fragestellung relevante Informationen entfernt -> Anfrage an Datengeber

➤ *Was ist wenn der Datengeber nicht mehr verfügbar ist?*
- Unlösbare Probleme – fächerübergreifender Diskurs notwendig

➤ Weitere Entfernung vom Ideal des Experimentellen Designs

➤ *Welche Datensätze können zur Beantwortung welcher Forschungsfragen sinnvoll zusammengeführt werden?*

Ausblick

- Potenziale

- Fortlaufenden Normierung von Testverfahren im Feld
 - Vergleichsdaten zu wiederholtem Einsatz von Testverfahren
 - Datengetriebene Forschung
- Probleme

- Unterstützung der Forscher bei hochwertigem Datenmanagement
 - Datenstandards. Auffindbarkeit der Daten, Dokumentation und Interpretierbarkeit
 - Datenschutz. Nutzbarkeit für Sekundäranalysen sicherstellen

Referenzen
Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E., & Martone, M. E. (2014). Big data from small data: data-sharing in the 'long tail' of neuroscience. *Nat Neurosci*, 17(11), 1442-1447. doi: 10.1038/nn.3838
Mayer, A.-K., Rosman, T., Birke, P., Gorges, J., & Krampen, G. (in press). Entwicklung professioneller Wissensnetze bei Novizen im Kontext von Präsenzlehre und Informationsrecherchen im Internet. Forschungsdaten einer Längsschnittstudie von 2013 bis 2015. Psychologisches Datenarchiv PsychData des ZPID. DOI: 10.5160/psychdata.mrae15ent24
Merk, S., & Bohl, T. (2016). Prädiktive Effekte domänenspezifischer epistemologischer Überzeugungen angehender Lehrerinnen und Lehrer auf deren Bedeutsamkeitseinschätzung allgemeinen pädagogischen sowie fachdidaktischen Wissens. Forschungsdaten einer experimentellen Untersuchung. Psychologisches Datenarchiv PsychData des ZPID. <https://doi.org/10.5160/psychdata.mksl12pr27>
Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-generalty and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review*, 18(1), 3-54. doi: 10.1007/s10648-006-9003-6.
Stahl, E., & Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction*, 17(6), 773–785. doi: 10.1016/j.learninstruc.2007.09.016.