

**Gemeinsame Tagung der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und der Sektion Methoden der empirischen Sozialforschung am Robert-Koch-Institut (RKI) in Berlin am 10./11. November 2017 zum Thema „Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente“.**

Natalja Menold, GESIS

Tobias Wolbring, Friedrich-Alexander-Universität Erlangen-Nürnberg

Die Qualität sozialwissenschaftlicher Erhebungsinstrumente hat wesentlichen Einfluss auf die Belastbarkeit empirischer Schlussfolgerungen, die auf Grundlage von Umfragedaten gezogen werden. Die ASI und die DGS-Sektion „Methoden der empirischen Sozialforschung“ stellten daher die Qualität von Erhebungsinstrumenten, Verfahren zur Bestimmung ihrer Güte und Methoden der Qualitätssicherung in den Mittelpunkt der gemeinsamen Herbsttagung. Die Tagung umfasste einen breiten Komplex an Themen und traf mit ca. 80 Teilnehmenden auf breites Interesse. Die insgesamt 13 Vorträge waren in fünf inhaltliche Sessions gegliedert, die sich den Themen „Kognitive & Mixed-Method-Ansätze“, „Messäquivalenz“, „Effekte der Erhebungssituation“, „Effekte des Erhebungsinstruments“ und „innovative Ansätze & Perspektiven“ widmeten.

Nach Grußworten durch Bärbel-Maria Kurth (RKI), Frank Faulbaum (Universität Duisburg-Essen & Vorstandsvorsitzender der ASI) und Tobias Wolbring (FAU Erlangen-Nürnberg & Sprecher der Methodensektion) begann die Tagung mit dem Vortrag von Cornelia Neuert (GESIS). Anhand zweier empirischer Studien konnte Neuert zeigen, dass sich durch Eye-Tracking zusätzliche Probleme in kognitiven Pretests aufdecken lassen und wie sich die Technik bei kognitiven Pretests sinnvoll eingeben lässt. Der zweite Vortrag von Arne Bethmann und Christina Buschle (DJI) stellte ebenfalls kognitive Pretests in den Mittelpunkt und unterbreitete den programmatischen Vorschlag, qualitative Forschungsmethoden, wie problemzentrierte und diskursive Interviewtechniken, stärker bei Pretests einzusetzen. Katharina Meitinger (GESIS) stellte in ihrer Präsentation zwei Ansätze zur Untersuchung von Messinvarianz gegenüber und demonstrierte deren Stärken und Schwächen anhand einer Anwendung zum Thema Nationalismus. Die konfirmatorische Faktorenanalyse erscheint dafür geeignet, Messäquivalenz für eine größere Zahl von Gruppen (z.B. Länder) zu untersuchen, während Web-Probing ein besseres Verständnis von Äquivalenzproblemen liefert. Die erste Session wurde mit dem Vortrag von Folke Brodersen (DJI) abgeschlossen, der Spezifika der Qualitätssicherung bei der Befragung von Jugendlichen mit geistiger Behinderung diskutierte und Herausforderungen „inklusive Methoden“ herausarbeitete.

Ein Highlight der Tagung fand anschließend mit der Keynote von Andreas Diekmann (ETH Zürich & Wissenschaftskolleg Berlin) statt. Ausgehend von Problemen sozialer Erwünschtheit stellte Diekmann methodische Verfahren zur Steigerung der Anonymität der Befragungssituation dar. Diese sollten – so die Hoffnung – die Validität von Antworten im Vergleich zu direkten Fragen deutlich erhöhen. Die Randomized-Response-Technik (RRT) erwies sich empirisch – entgegen der ursprünglichen Erwartungen – nicht als „Wundermittel“ gegen Probleme sozialer Erwünschtheit. Wie in mehreren Validierungsstudien gezeigt werden konnte, besteht nämlich nicht nur ein Trade-Off zwischen der Effizienz der Schätzung und der Antwortvalidität. Vielmehr kann RRT wie auch deren Weiterentwicklung in Form des Crosswise-Modells tatsächliche Prävalenzen überschätzen. Allerdings gibt es neuere

Ansätze, welche vermutlich nicht unter diesen Problemen leiden und daher, so das Fazit, hoffnungsfroh stimmen lassen. Im Anschluss daran gab Frank Faulbaum die Preisträger des erstmals verliehenen ASI-Nachwuchspreises bekannt. Prämiert wurde zum einen die Soziologin Stefanie Jähnen (WZB) für einen in der *KZfSS* erschienen Artikel zu den Effekten schulrechtlicher Reformen auf die Bildungsungleichheit, zum anderen der Politikwissenschaftler Robert A. Huber (ETH Zürich) für einen in der *ZfVP* publizierten Beitrag zu den Konsequenzen der Ausbreitung rechtspopulistischer Parteien für die Demokratiequalität in Europa.

Danach folgte eine Session, welche die Thematik Messäquivalenz weiter vertiefte. Jessica Fischer (DIPF) setzte sich mit der internationalen Vergleichbarkeit von Unterrichtsqualitätsmessungen in PISA auseinander. So konnte empirisch belegt werden, dass Messäquivalenz für viele Länder nicht gegeben ist. Gruppiert man die Länder nach Sprachgruppen und legt ein schwächeres Kriterium an, sieht das Bild jedoch positiver aus. Während dieser Vortrag damit nach Messinvarianz auf Länderebene fragte, stand im Beitrag von Antje Rosebrock (Universität Mannheim) Messinvarianz für Personen mit und ohne Migrationshintergrund innerhalb eines Landes im Mittelpunkt. Auf Basis des niederländischen LISS-Panels dokumentierte die Autorin, dass ein naiver Ansatz, der Messfehler ignoriert, bei Mittelwertenvergleichen zu falschen Schlüssen kommt, weshalb die Daten vor der Analyse zu korrigieren sind.

Den Abschluss des ersten Tages bildeten zwei Vorträge zu Effekten der Erhebungssituation. Hawal Shamon (Forschungszentrum Jülich) thematisierte Probleme bei Online-Umfragen, die sich aus einer mangelnden intrinsischen Motivation der Respondenten ergeben. Auf Grundlage zweier Studien wurde daher empfohlen, Kontrollinstrumente zur Aufdeckung mangelnder Aufmerksamkeit einzusetzen. Verena Ortmanns (GESIS) wies schließlich auf Inkonsistenzen bei der ISCED-basierten Bildungsmessung in unterschiedlichen internationalen Umfragen hin. Als erklärungsrelevant erwiesen sich für diese Unterschiede Umfragecharakteristika, wie das Vorgehen bei der Codierung und die Umfrageorganisation.

Am Samstag begann die vierte Session „Effekte des Erhebungsinstruments“ mit der Präsentation von Natalja Menold (GESIS). Ihr Thema war die Fragestellung, inwieweit doppelte Stimuli die Messqualität beeinflussen. Die vorgestellten Experimente zeigten, dass Personen ähnliche Stimuli unterschiedlich bewerten. Anstatt jedoch Items ad hoc zu verändern, sollte stattdessen bereits beim Entwurf der Items auf deren Eindimensionalität geachtet werden. Im Anschluss rückten Jan Karem Höhne und Stephan Schlosser (Universität Göttingen) die Verbalisierung von Antwortskalen in den Mittelpunkt. Während ausgewogen beschriftete und endpunktbenannte Skalen in einer Studierendenbefragung zu sehr ähnlichen Antwortverteilungen führten, lieferten unausgewogen benannte Skalen differierende Resultate. Letzterer Ansatz erscheint daher nicht empfehlenswert.

Die letzte Session „Innovative Ansätze und Perspektiven“ wurde von Knut Petzold (Universität Bochum) eröffnet. In der vorgestellten Studie verglichen die Autoren faktorielle Surveys zu diskriminierendem Verhalten mit Resultaten verdeckter Feldexperimente. Vignetten lieferten zwar keine validen Schätzungen der Verteilung realen Verhaltens, jedoch zeigten sich in beiden Studien ähnliche Verhaltensdeterminanten. Der vorletzte Vortrag von Hans Dietrich (IAB) behandelte

Effekte sozialer Erwünschtheit bei der Befragung junger syrischer Fluchtmigranten. Die aus Registerdaten gezogene Stichprobe IAB-WELLCOME erlaubte es dabei, die Abhängigkeit des Antwortverhaltens auf heikle Fragen von Umfragemodus und Interviewergeschlecht nachzuweisen. Abgeschlossen wurde die Tagung mit einem Vortrag von Gina Schöne (RKI), die Anforderungen an ein effektives Qualitätsmanagement epidemiologischer Studien veranschaulichte. Als zentrale Herausforderungen wurden u. a. Abstimmungsprozesse zwischen Stakeholdern, aber auch eine zielorientierte Führung des Qualitätsmanagements identifiziert.