

GESIS Survey Guidelines

Messinstrumente in internationalen Studien

Dorothee Behr, Michael Braun & Brita Dorer

Zusammenfassung

Dieses Kapitel beschäftigt sich mit Messinstrumenten, d.h. Fragebögen und einzelnen Fragen, in internationaler Umfrageforschung. Unter internationalen Umfragen verstehen wir dabei Studien, bei denen verschiedene Länder miteinander verglichen werden und bei denen die Forderung nach Vergleichbarkeit die Konzeption, Durchführung und Organisation der Studien prägt.

Wir konzentrieren uns zunächst auf allgemeine Fragen der Vergleichbarkeit internationaler Umfragen, wobei wir hinsichtlich der Probleme, die für alle Umfragen allgemein gelten, sowie der Probleme der Stichprobenziehung und Non-response in internationalen Umfragen auf die entsprechenden thematischen Kapitel sowie weiterführende Literatur verweisen. Wir unterscheiden sodann zwischen der ex-ante- und ex-post-Sicherstellung vergleichbarer Messinstrumente, d.h. bevor (z.B. in der Form von Pretests) und nachdem die eigentliche Umfrage durchgeführt wird (im Rahmen der Datenanalyse oder zusätzlicher Studien). Besonderes Gewicht legen wir auf die Übersetzung von Fragen.

Zitierung

Behr, Dorothee, Braun, Michael & Dorer, Brita (2015): Messinstrumente in internationalen Studien. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_006



1. Allgemeines zu Messinstrumenten in internationaler Umfrageforschung

Die folgenden Kapitel befassen sich mit internationaler Umfrageforschung. Hierunter werden Umfragen in zwei oder mehr Ländern verstanden, bei denen das Konzept der Vergleichbarkeit bzw. Äquivalenz maßgebend für deren Konzeption, Durchführung und Organisation ist.

Auch wenn der Fokus auf internationalen Umfragen liegt, gilt jedoch vieles des hier Gesagten auch für interkulturelle Umfragen innerhalb eines Landes, so z.B. wenn unterschiedliche ethnische und sprachliche Gruppen innerhalb eines Landes miteinander verglichen werden sollen.

Die Gliederung der folgenden Kapitel basiert auf der Prämisse, dass für die Herstellung von vergleichbaren Messinstrumenten in internationaler Umfrageforschung sowohl ex-ante als auch ex-post-Prozeduren notwendig sind. Vergleichbarkeit muss zunächst im Rahmen der Fragebogenentwicklung und Übersetzung durch verschiedene Prozesse und Maßnahmen ex-ante hergestellt werden (Kapitel 2 und 3). Fragebogenentwicklung und Übersetzung sind dabei eng miteinander verbunden, d.h., die Qualität des einen ist unabdingbare Voraussetzung für die Qualität des anderen. Nach Erhebung der endgültigen Daten einer Studie ist ex-post die Vergleichbarkeit der Daten zu testen (Kapitel 4). Kapitel 1 stellt eine kurze allgemeine Einführung in Messinstrumente im internationalen Vergleich dar.

Folgende Themen werden in Kapitel 1 angesprochen:

- 1.1 Probleme der Vergleichbarkeit der Projektkomponenten
- 1.2 Messfehler
- 1.3 Vergleichbarkeit des Messinstruments („Äquivalenz“)
- 1.4 Fragebogentypen in internationaler Umfrageforschung

1.1 Probleme der Vergleichbarkeit der Projektkomponenten

Für internationale Umfragen sind nicht nur die Probleme, die für alle Umfragen allgemein gelten, zu berücksichtigen. Vielmehr bemisst sich die Qualität eines internationalen Projekts an der Qualität der einzelnen nationalen Umfragen und insbesondere auch daran, dass diese „vergleichbar“ sind. Sind einige der nationalen Umfragen mit Fehlern behaftet, dann entsprechen beobachtete Ähnlichkeiten und Unterschiede möglicherweise lediglich methodischen Artefakten. Dabei kann auch die Verwendung identischer Prozeduren in den einzelnen Ländern nicht notwendigerweise Vergleichbarkeit garantieren. Teilweise ist die Verwendung unterschiedlicher, aber den jeweiligen Kontexten angepasster Untersuchungsstrategien vorzuziehen (z.B. Stichprobenziehung auf der Grundlage unterschiedlicher Listen oder länderspezifische Formulierung von Antwortkategorien). Allgemein kann man unterschiedliche Fehlertypen unterscheiden:

- Repräsentieren die in den unterschiedlichen Ländern gezogenen und realisierten Stichproben die jeweiligen Grundgesamtheiten in vergleichbarer Weise (coverage/sampling/non-response error)?
- Funktionieren die Messinstrumente, d.h. die Fragebögen und die einzelnen Fragen nach der Übersetzung, überall gleich (Messfehler)?

Der coverage/sampling/non-response error wird insbesondere im GESIS Survey Guidelines Artikel „Nonresponse Bias“ (Koch & Blohm, 2015) behandelt. Der Messfehler wird im Folgenden erläutert.

1.2 Messfehler

Als Messfehler („measurement error“) wird die Abweichung zwischen dem beobachteten Wert und dem „wahren“ Wert bezeichnet. Der Messfehler hat eine Reihe von miteinander interagierenden Ursachen, die auf das Messinstrument, die Interviewer, die Befragten oder den Umfragemodus zurückgeführt werden können (Biemer et al., 1991).

- Messinstrumente können durch ihr kompliziertes Design, die Formulierung einzelner Fragen (und natürlich ihre unzureichende Übersetzung), die Fragenreihenfolge und die Antwortalternativen zum Messfehler beitragen.
- Interviewer können zum Messfehler beitragen, indem sie z.B. vom Fragetext abweichen oder die Angaben der Befragten nicht korrekt aufzeichnen. Länder unterscheiden sich in ihrer Umfrage-tradition und damit auch in den Durchführungsstandards, wie z.B. dem Training der Interviewer.
- Verhaltenstendenzen von Befragten wie soziale Erwünschtheit und Akquieszenz sind weitere Quellen für Messfehler. Sie können sowohl Persönlichkeitseigenschaften darstellen als auch durch situationale Charakteristika aktiviert werden. Vorhandene Evidenz belegt, dass auch die Tendenz zur Abgabe sozial erwünschter Antworten interkulturell variiert (Johnson & van de Vijver, 2003).
- Eine interkulturell unterschiedliche Tendenz, sozial erwünscht zu antworten, kann sich darüber hinaus je nach Umfragemodus anders auswirken.

Um Vergleichbarkeit zwischen den einzelnen Umfragen eines internationalen Projektes zu gewährleisten, müssen die Messfehler in allen beteiligten Ländern möglichst gleichmäßig minimiert werden. Ein Vergleich von Umfragen, die nur unter geringen Messfehlern leiden, mit solchen, die durch große Messfehler gekennzeichnet sind, ist dabei eher ungünstig. Weniger problematisch, zumindest aus Sicht des Vergleichs, wäre es demgegenüber, über alle Länder den gleichen Messfehler beizubehalten. Dies könnte z.B. bedeuten, dass suboptimale Fragen in allen Ländern gestellt werden und Länder nicht eigenmächtig Fragen verbessern.

Um die oben genannten unterschiedlichen Probleme in ihren Auswirkungen abzuschwächen, gibt es eine Vielzahl möglicher Maßnahmen.

Wie Vergleichbarkeit beim Einsatz eines Erhebungsinstrumentes (z.B. hinsichtlich von Interviewertraining und Monitoring) erreicht werden kann, wird ausführlich in den Cross-cultural survey guidelines (<http://ccsg.isr.umich.edu/index.php/chapters/interviewer-recruitment-selection-and-training-chapter>) beschrieben.

Wie die Vergleichbarkeit der Messinstrumente gewährleistet bzw. getestet werden kann, wird in den nun folgenden Kapiteln beschrieben.

Für eine vertiefte Auseinandersetzung mit der Problematik internationaler Studien insgesamt bieten sich darüber hinaus Braun (2006), Davidov, Schmidt, & Billiet (2011), Harkness et al. (2010), Harkness, van de Vijver, & Mohler (2003), van Deth (2013[1998]) und van de Vijver & Leung (1997) an.

1.3 Vergleichbarkeit des Messinstruments („Äquivalenz“)

Die Vergleichbarkeit bzw. „Äquivalenz“ der Messinstrumente ist die Voraussetzung eines jeden sinnvollen Vergleichs zwischen einzelnen Ländern. In der Regel werden mehrere Typen von Äquivalenz unterschieden (van de Vijver und Leung, 1997):

- Konstruktäquivalenz („construct equivalence“)

- Äquivalenz der Maßeinheit („measurement unit equivalence“) und
- skalare Äquivalenz („scalar equivalence“).

Das Vorliegen von Konstruktäquivalenz, d.h. dass in den unterschiedlichen Ländern dieselbe zugrunde liegende Dimension erfasst wird, ist Grundvoraussetzung für jeden inhaltlichen Vergleich zwischen Ländern. Sie ist gleichfalls Voraussetzung für die beiden anderen Äquivalenzarten, die Äquivalenz der Maßeinheit und die skalare Äquivalenz. Die Sicherstellung von Konstruktäquivalenz kann einen Einsatz verschiedener Messinstrumente in unterschiedlichen Ländern erfordern. Dann sind allerdings die beiden anderen Äquivalenzarten in der Regel nicht mehr gegeben. Äquivalenz der Maßeinheit bedeutet, dass die Maßeinheit der Skala in unterschiedlichen Ländern gleich ist. Für den Ursprung der Skala braucht dies jedoch nicht zu gelten. Skalare Äquivalenz bedeutet demgegenüber, dass sowohl Maßeinheit als auch Ursprung der Skala gleich sind. Dies ist Voraussetzung z.B. für die Interpretierbarkeit von Mittelwertsunterschieden zwischen Ländern.

1.4 Fragebogentypen in internationaler Umfrageforschung

Fragebogenentwicklung im internationalen Kontext kann zum einen auf Basis eines „ask-the-same-question“-Ansatzes (ASQ) geschehen, zum anderen auf Basis eines „ask-a-different question“-Ansatzes (ADQ).

Der ASQ-Ansatz beruht auf der Grundidee, dass in den teilnehmenden Ländern und Kulturen einer Studie dieselben Fragen gestellt werden. Dies wird typischerweise dadurch erreicht, dass zunächst ein Ausgangsfragebogen in einer Projektsprache, zumeist Englisch, erstellt wird, der dann in die Sprachen der teilnehmende Länder oder Kulturen einer Studie übersetzt wird. ASQ ist somit im Wesentlichen mit der Übersetzung eines Fragebogens gleichzusetzen.

Der ADQ-Ansatz beruht auf der Annahme, dass zwar Konzepte gemeinsam definiert werden, dass aber die Umsetzung durch Fragen/Antwortkategorien und/oder Indikatoren in den teilnehmenden Ländern und Kulturen einer Studie unterschiedlich aussieht. Ein Einsatz länderspezifischer Messinstrumente ist erforderlich, wenn identische Fragen in verschiedenen Ländern nicht oder unterschiedlich verstanden würden, und zwar nicht nur aufgrund einer schlechten Übersetzung, sondern aufgrund unterschiedlicher Realitäten in den Ländern. Ein typisches Beispiel für länderspezifische Frage ist die Frage nach der Bildung. Selbst wenn die eigentliche Frageformulierung über Länder hinweg gleich ausfallen mag, so sind doch die Antwortkategorien von den jeweiligen Bildungssystemen der Länder anhängig. Darüber hinaus kann im Rahmen eines ASQ-Ansatzes das von Przeworski und Teune (1966) vorgeschlagene „identity-equivalence“-Verfahren eingesetzt werden (van Deth, 2013): Zunächst werden die für alle Länder identischen Indikatoren bestimmt und dann zusätzlich die jeweils länderspezifischen Indikatoren berücksichtigt. Daraus resultieren unterschiedliche Skalen für die einzelnen Länder. Mithilfe der identischen Items lässt sich dann die funktionale Äquivalenz der länderspezifischen Items überprüfen. Häufig ist allerdings nicht einfach bestimmbar, wie Varianten einer Frage auszusehen haben, mit denen in allen Ländern das Gleiche erfasst werden kann. ADQ kann im weitesten Sinne auch durch den Begriff der Adaptation abgedeckt werden.

Selbstverständlich gibt es hier auch Mischformen im Kontext der Fragebogenentwicklung, z. B. basieren die meisten Items im European Social Survey (ESS) auf dem ASQ-Ansatz. Die Bildung ist jedoch eine der Variablen, die länderspezifisch abgefragt wird (ADQ) und im Anschluss nach vorgegebenen Regeln harmonisiert wird.

Die folgenden Kapitel befassen sich mit Fragebogenentwicklung und -übersetzung im Rahmen des ASQ-Ansatzes.

2. Fragebogenentwicklung in internationaler Umfrageforschung

Im Folgenden wird es um die Fragebogenentwicklung im Kontext des ASQ-Ansatzes gehen. Neben den inhaltlichen und methodischen Grundprinzipien allgemeiner Fragebogenentwicklung gilt es im internationalen Kontext einen Fragebogen zu entwickeln, der für die verschiedenen Kulturen relevant ist und zudem gut übersetzt werden kann.

Neben den Ausführungen dieses Kapitels bieten z.B. Harkness, van de Vijver, & Johnson (2003), Harkness, Edwards, et al. (2010), Harkness, Bilgen, et al. (2011) und Smith (2003) Einblicke in die Entwicklung von Fragebögen in internationalen Studien.

Folgende Themen werden in Kapitel 2 angesprochen:

2.1 Internationale Zusammenarbeit

2.2 Pretesting

2.3 Problemorientierte Codierung, *translatability assessment* und *advance translation*

2.1 Internationale Zusammenarbeit

Best Practice für internationale Studien verlangt, dass die verschiedenen Länder einer Studie an der Fragebogenentwicklung zu beteiligen sind. Nur auf diese Art und Weise kann sichergestellt werden, dass Fragen entwickelt werden, die international und interkulturell eingesetzt werden können. Unterschiedliche Modelle der Zusammenarbeit sind möglich.

Handelt es sich z.B. um eine zwei Länder umfassende Studie, ist es leicht, die Zusammenarbeit zu koordinieren. Handelt es sich demgegenüber um Mehrländerstudien mit z.B. 20 bis 30 teilnehmenden Ländern, gestaltet sich die Koordination und Zusammenarbeit schwieriger. Praktikable Lösungen für große Mehrländerstudien finden sich im European Social Survey (ESS) oder dem International Social Survey Programme (ISSP). In beiden Studien gibt es eine mit der Fragebogenentwicklung beauftragte Gruppe, die bereits verschiedene Kulturen und Länder umfasst. In regelmäßigen Abständen kommen Vertreter aller Länder zusammen, um Zwischenfassungen eines Fragebogens gemeinsam zu diskutieren.

2.2 Pretesting

Der Fragebogen ist im Rahmen der Fragebogenentwicklung an der Zielgruppe zu testen. Bei *quantitativen* Pretests wird der für die Hauptstudie vorgesehene Fragebogen unter – mehr oder weniger – realen Bedingungen getestet. Quantitatives Pretesting kann im Idealfall dazu dienen, die Äquivalenz der Messinstrumente über die beteiligten Länder hinweg mit Hilfe komplexer statistischer Verfahren zu überprüfen (s. Kapitel 4). Dies setzt allerdings voraus, dass diese Pretests in möglichst allen Ländern mit einer hinreichend hohen Fallzahl realisiert werden. Dies ist allerdings in der Praxis, insbesondere auch bei den meisten internationalen Großprojekten, nicht der Fall. Vielfach dienen quantitative Pretests lediglich der Untersuchung von Verständnisproblemen bei einzelnen Fragen, der Handhabbarkeit des gesamten Instruments oder der Schätzung der für die Durchführung einer Befragung benötigten Zeit. Die Konsequenz ist, dass die Qualität der Messinstrumente in der Regel vom Sekundärforscher im Zuge der Datenanalyse überprüft werden muss.

Bei den *qualitativen* Verfahren dominieren kognitive Interviews, bei denen Probleme bei der Beantwortung der Fragen herausgearbeitet werden sollen (Beatty & Willis, 2007; Willis, 2005). In den letzten Jahren wurde die Methoden-Toolbox dahingehend erweitert, dass nun auch interkulturell und international vergleichende kognitive Interviews durchgeführt werden. Wenn diese während der

Entwicklungsphase des Fragebogens eingesetzt werden, können Probleme der Non-Äquivalenz frühzeitig aufgedeckt und umgangen werden.

International vergleichende kognitive Interview-Studien sind allerdings im Vergleich zu *interkulturell* vergleichenden kognitiven Interview-Studien in einem Land sehr selten (als Ausnahme: Fitzgerald et al., 2011, Miller et al., 2011).

Mehr Informationen, einschließlich lessons learned und Empfehlungen, wie vergleichende kognitive Interviews zu planen und durchzuführen sind, finden sich in Farrall et al. (2012), Fitzgerald et al. (2011), Goerman & Caspar, Lee (2012), Miller et al. (2011), and Thrasher et al. (2011).

In der Realität tritt die Situation, in der in allen beteiligten Ländern große quantitative Pretests oder qualitative Vorstudien durchgeführt werden, wegen des hohen Aufwandes praktisch nicht auf. Selbst im European Social Survey (ESS) werden qualitative Pretests nur in einer begrenzten Zahl teilnehmender Länder und ein quantitativer Pretest in der Regel nur in zwei der beteiligten Länder durchgeführt. Zudem werden diese Pretests – wie auch beim ISSP – vor allem unter dem Gesichtspunkt der Auswahl von inhaltlich interessanten und methodisch adäquaten Fragen durchgeführt und nicht im Sinne eines Tests des endgültigen, auch in der Hauptstudie verwendeten Messinstruments. Daher ist es erforderlich, zusätzlich eine Überprüfung der Äquivalenz ex-post, d.h. auf der Grundlage der in der Hauptstudie erhobenen Daten, vorzunehmen (s. Kapitel 4).

Sowohl die Ergebnisse von quantitativen als auch qualitativen Pretests führen in der Regel zu einer – teilweise beträchtlichen – Veränderung der Messinstrumente. Im Prinzip sollten auch die veränderten Versionen einer erneuten Überprüfung unterzogen werden. Auch dieses wird jedoch in der Praxis – meist aus Zeitgründen – unterlassen.

Eine besondere Form des Testens, die auch post-hoc durchgeführt werden kann, wird im Folgenden dargestellt. Es handelt sich hierbei um die Anwendung von Probing-Techniken in internationalen Web-Umfragen: Behr et al. (2014) nutzen zur Analyse eines ISSP-Items zu zivilem Ungehorsam sowohl „comprehension“ („Was verbinden Sie mit dem Begriff ‚ziviler Ungehorsam‘, nennen Sie Beispiele“) als auch „category-selection probes“ („Bitte begründen sie, warum Sie sich für ... entschieden haben“), die in internationalen Web-Umfragen nach dem betreffenden ISSP-Item implementiert werden. Mithilfe der von den Befragten eingetragenen Antworten können sie zeigen, dass die geringe Befürwortung von zivilem Ungehorsam in Ländern wie den Vereinigten Staaten und Kanada im Vergleich zu Ländern wie Deutschland und Spanien im Wesentlichen zwei Gründe hat. Zum einen wird in der ersten Gruppe von Ländern ziviler Ungehorsam stärker mit Gewalt assoziiert als in der zweiten. Dies ist ein Methodenartefakt, d.h. die Befragten beantworten faktisch unterschiedliche Fragen. Zum anderen ist das Vertrauen in die Politiker in der zweiten Gruppe (noch) geringer als in der ersten. Dies ist ein inhaltliches Ergebnis. In den Daten sind faktisch Methodenartefakt und reale Unterschiede konfundiert.

Braun, Behr & Kaczmirek (2013) können des Weiteren zeigen, dass Befragte aus unterschiedlichen Ländern bei Fragen zu Migranten zwar weitgehend an vergleichbare, auf der tatsächlichen Migrationsrealität basierende Gruppen denken, dass es dabei aber dennoch zu charakteristischen Verzerrungen kommt. Sie verwenden dazu „specific probes“ („An welche Gruppen haben Sie bei der Beantwortung der Frage gedacht?“), die im Anschluss an ISSP-Items zu Einstellungen zu Zuwanderern implementiert werden. An welche Migrantengruppen die Befragten denken, hätte mit den herkömmlichen statistischen Verfahren alleine nicht untersucht werden können.

2.3 Problemorientierte Codierung, translatability assessment und advance *translation*

Kategorienschemata zur systematischen Bewertung von Items, im englischen Sprachraum auch *question appraisal systems* genannt, können zusätzlich den Prozess der Fragebogenentwicklung

begleiten. Ziel dieser Kategorienschemata ist es, neue Items systematisch auf bestimmte Qualitätsdimensionen hin zu überprüfen. Ein Kategorienschema, das besonders auf die Bedürfnisse interkultureller Studien für den amerikanischen Raum ausgerichtet wurde, bieten Dean et al. (2007).

Zudem haben Aquadro et al. (2012) für medizinische Fragebögen eine Form des *translatability assessment* entwickelt. Auch hier wird eine Ausgangsfrage systematisch auf ihre interkulturelle Anwendbarkeit und Übersetzbarkeit hin getestet.

Ähnliche Ziele und Zwecke verfolgt die *advance translation*. Diese Methode besteht darin, innerhalb der Fragebogenentwicklung eine Vorab-Übersetzung durchzuführen. Diese soll dazu beitragen, spätere Übersetzungsprobleme und -fehler zu minimieren, den Ausgangsfragebogen also besser übersetzbar zu machen. Ziel ist es, den – meist englischsprachigen – Ausgangsfragebogen bereits so zu formulieren, dass er (a) zum einen möglichst wenige Stellen enthält, die später bei der Übersetzung in verschiedene Sprachen Probleme bereiten, und dass (b) auch der spätere Einsatz des Fragebogens in den Zielsprachen und -kulturen bereits während der Fragebogenentwicklung berücksichtigt wird. Die *advance translation* kann dazu beitragen, Stellen innerhalb des Ausgangsfragebogens zu ermitteln, die in Bezug auf diese beiden Punkte problematisch sind. Bei der *advance translation* empfiehlt es sich, dass Teams aus erfahrenen FragebogenübersetzerInnen und Umfragemethodikern gemeinsam eine noch nicht endgültige Version des Ausgangsfragebogens übersetzen und gezielt kommentieren, an welchen Stellen sie bei diesem Übersetzungsprozess auf Probleme stoßen. Des Weiteren ist es ratsam, wenn in diesem Stadium jegliche Arten von Kommentaren abgegeben werden, die diesen Personen in Bezug auf die spätere Verwendung des Fragebogens in den verschiedenen Zielkulturen in den Sinn kommen. Die *advance translation* ist beispielsweise im European Social Survey (ESS) seit 2009 fester Bestandteil der Entwicklung des englischsprachigen Ausgangsfragebogens (Dorer, 2011).

Den Fokus auf die Überprüfung formaler Kriterien der Fragebogenentwicklung legt SQP – das Survey Quality Prediction Programme. Hierzu bietet die Webseite <http://sqp.gesis.org> weitergehende Informationen.

3. Fragebogenübersetzung

Die Qualität von Fragebogenübersetzungen ist von entscheidender Bedeutung für die Qualität von Forschungsergebnissen.

Folgende Themen werden hierzu in Kapitel 3 angesprochen:

- 3.1 Worauf ist bei der Fragebogenübersetzung zu achten?
- 3.2 Wie sollte der Übersetzungsprozess organisiert werden?
- 3.3 Muss eine Fragebogenübersetzung gesondert getestet werden?
- 3.4 Wo finde ich Übersetzer?
- 3.5 Welche Informationen sollten Übersetzern zur Verfügung gestellt werden?
- 3.6 Ich brauche Übersetzungen nicht nur für ein anderes Land, sondern für eine besondere Migrantengruppe in diesem Land. Was ist hierbei zu berücksichtigen?
- 3.7 Ist Fragebogenübersetzung gleich Fragebogenübersetzung? Welche Typen der Fragebogenübersetzung gibt es?
- 3.8 Welche Formen der Frage(bogen)-Adaptation gibt es?
- 3.9 Was muss ich bei Projektanträgen berücksichtigen, in deren Rahmen eine internationale bzw. interkulturelle Studie durchgeführt werden soll?

3.1 Worauf ist bei der Fragebogenübersetzung zu achten?

Für eine gelungene Fragebogenübersetzung gelten zunächst dieselben Qualitätsanforderungen wie für jede andere Übersetzung auch: Sie muss inhaltlich und sprachlich korrekt und vollständig sein. Des Weiteren muss ein übersetzter Fragebogen als Umfrageinstrument in der Zielsprache auf die beabsichtigte Weise *funktionieren*. Es ist also besonders wichtig, dass die *Bedeutung* in der Zielsprache allgemein verständlich zum Ausdruck gebracht wird; der Stimulus des Ausgangsfragebogens sollte also dem Stimulus im Zielfragebogen entsprechen.

Weil in sozialwissenschaftlichen Umfragen meist ein Querschnitt der Gesamtbevölkerung befragt wird, ist es besonders wichtig, eine möglichst einfache und leicht verständliche Sprache zu verwenden; komplizierte Satzkonstruktionen sollten ebenso vermieden werden wie seltene Wörter oder gar Fremdwörter. Fragen und vorgegebene Antwortkategorien sollten auch in der Zielsprache klar und eindeutig formuliert sein.

Im Zweifelsfall sollten immer die im Zielland üblichen Umfragepraktiken als Vorgaben dienen, beispielsweise in Bezug auf die Ansprache der Befragten (z. B. Höflichkeit) oder den Umgang mit den verschiedenen Geschlechtern der befragten Personen. So sollte man sich beispielsweise im Deutschen entscheiden, ob ausschließlich die männliche Form verwendet werden soll, die Personen männlichen und weiblichen Geschlechts gleichermaßen einschließt, oder ob die männliche und die weibliche Form stets explizit nebeneinander genannt werden (z. B. „Wie stolz sind Sie, ein/e Deutsche/r zu sein?“).

Auch sollte auf Konsistenz innerhalb des übersetzten Fragebogens geachtet werden: Konsistenz einerseits in Bezug auf die einheitliche Verwendung einzelner Schlüsselwörter und Begriffe innerhalb eines Fragebogens, Konsistenz andererseits aber auch in Bezug auf die einheitliche Verwendung von Übersetzungen innerhalb einer Studie über verschiedene Runden oder Wellen hinweg. Dies ist insbesondere maßgeblich bei Panel-Befragungen oder wiederholten Querschnittbefragungen, bei denen es zu Zwecken der Vergleichbarkeit über die Jahre hinweg wichtig ist, dass gewisse Begrifflichkeiten und Wortlaute im Laufe der Zeit einheitlich verwendet werden.

Eine detaillierte Auflistung und Diskussion verschiedener bei der Fragebogenübersetzung zu beachtender Punkte findet sich im zweiten Teil der Translation Guidelines des European Social Survey (European Social Survey, 2014).

3.2 Wie sollte der Übersetzungsprozess organisiert werden?

Als best practice in der sozialwissenschaftlichen Forschung hat sich der sogenannte „Team-Ansatz“ (team approach oder committee approach) etabliert (siehe beispielsweise European Social Survey, 2014; Harkness 2003; Harkness, 2010). Dieser besteht zum einen darin, dass mehrere Personen an der Übersetzung beteiligt sind, die unterschiedliche Qualifikationen in die Übersetzung einbringen, und zum anderen darin, dass ein mehrstufiger Prozess durchlaufen wird. Dieses Verfahren bietet die folgenden Vorteile: Dadurch, dass man sich nicht auf eine einzelne Person verlässt, wird das Risiko minimiert, dass das Endprodukt zu stark von dem persönlichen Stil dieser einen Person geprägt wird. Des Weiteren unterlaufen auch den erfahrensten ÜbersetzerInnen Fehler, und es kommt immer wieder vor, dass auch sehr gut ausgebildete ÜbersetzerInnen ein und dasselbe Wort unterschiedlich verstehen. Durch die Beteiligung verschiedener Personen am Prozess werden diese Fehlerquellen reduziert. Ein zusätzlicher Sicherungsmechanismus ist die Tatsache, dass die Übersetzung mehrere Schritte durchläuft, sodass die Wahrscheinlichkeit fehlerhafter Übersetzungen oder eines zu persönlichen Übersetzungstils minimiert werden kann, bis die endgültige Übersetzung fertiggestellt ist.

Der Team-Ansatz ist durch verschiedene Pretest-Verfahren und sorgfältige Prozess-Dokumentation zu ergänzen. Das Gesamtverfahren wird auch „TRAPD“ genannt. Dieses ursprünglich von Janet Harkness

entwickelte Verfahren ist beispielsweise beschrieben in Harkness (2003) oder den ESS Translation Guidelines (2014). ‚TRAPD‘ ist ein Akronym, das aus den verschiedenen Schritten dieses Verfahrens gebildet wird:

- ‚T‘ (Translation) steht für den ersten Schritt, nämlich zwei unabhängig voneinander durchgeführte Übersetzungen. Bei diesem Schritt wird empfohlen, dass zumindest eine(r) der ÜbersetzerInnen ein ausgebildeter und/oder professioneller Übersetzer ist oder zumindest ein sprachwissenschaftliches Studium absolviert hat.
- ‚R‘ (Review) steht für die anschließende Sitzung, auf der die beiden unabhängigen Übersetzungen unter Moderation eines ‚Reviewers‘ diskutiert werden und für jedes Fragebogen-Item möglichst bereits eine Lösung erarbeitet wird. In der Phase der ...
- ‚A‘ (Adjudication) werden Unklarheiten oder offene Fragen, die unter Umständen nach dem Review-Schritt noch bestehen, mit einem erfahrenen Umfrageexperten, dem ‚Adjudicator‘, geklärt. Die Personen, die für die Funktionen des ‚Reviewer‘ sowie des ‚Adjudicator‘ ausgewählt werden, sollten über einschlägige Kenntnisse in Sozialwissenschaften und/oder der Umfragemethodik verfügen (zur Auswahl geeigneter Personen s. Kapitel 3.4).
- Im Pretest (‚P‘) werden dann sämtliche Fragebogen-Items in ihrer Übersetzung an einer kleinen Stichprobe der Zielbevölkerung getestet, um abschätzen zu können, inwieweit die Übersetzungen in der beabsichtigten Art und Weise verstanden werden (s. Kapitel 3.3).
- Der letzte Schritt, ‚D‘, besteht in der Dokumentation des gesamten Prozesses. Hierdurch sollen Informationen rund um die Entstehung der finalen Übersetzungen für Interessierte bereitgestellt werden, wenn möglich sowohl innerhalb als auch außerhalb des betreffenden Projekts.

Es gibt selbstverständlich andere Wege, eine Fragebogenübersetzung anzufertigen, von denen hier insbesondere auf die Rückübersetzung/Backtranslation eingegangen werden soll. Rückübersetzung/Backtranslation wird innerhalb der Fragebogenübersetzung noch immer eingesetzt, obwohl die Schwachstellen dieser Methode vielfach benannt werden (siehe beispielsweise Harkness, 2003; Schoua-Glusberg & Villar 2013). Diese Methode besteht darin, zunächst eine Übersetzung aus der Ausgangssprache A in die Zielsprache B von einer Person anfertigen zu lassen, woraufhin diese durch eine andere Person aus Zielsprache B wieder in Ausgangssprache A rückübersetzt wird. Der Ausgangstext in Sprache A wird sodann mit der Rückübersetzung in Sprache A verglichen; weichen diese beiden Versionen voneinander ab, wird die betreffende Übersetzung nochmals überprüft und gegebenenfalls überarbeitet. Aus übersetzungswissenschaftlicher Sicht ist diese Methode jedoch fragwürdig. Gute Übersetzungen zeichnen sich häufig gerade durch einen gewissen Freiheitsgrad aus. Die Wahrscheinlichkeit, dass eine gute, das heißt relativ freie, Übersetzung in ihrer Rückübersetzung wieder denselben Ausgangstext ergibt, ist sehr gering; dies bedeutet, dass unter Umständen gelungene Übersetzungen fälschlicherweise als problematisch dargestellt werden könnten. Andererseits können schlechte Übersetzungen leicht unbemerkt bleiben, wenn beispielsweise beide an dem Prozess beteiligten ÜbersetzerInnen (A-B und B-A) nicht gut übersetzen. Janet Harkness (Harkness, 2003) hat ein treffendes Beispiel gewählt, um diese Problematik zu veranschaulichen: Wenn der deutsche Ausdruck „das Leben in vollen Zügen genießen“ eine unidiomatische, also schlechte Übersetzung wie „to enjoy life in full trains“ ergibt und dies in einer wörtlichen Rückübersetzung wieder den ursprünglichen Wortlaut ergibt, kann es bei Anwendung der Rückübersetzung/Backtranslation also geschehen, dass dieser Übersetzungsfehler unbemerkt bleibt. Bei Anwendung des TRAPD-Ansatzes hingegen ist es sehr unwahrscheinlich, dass eine derart ungenügende Übersetzung von keiner der beteiligten Personen in keinem der durchlaufenen Schritte erkannt und aufgezeigt würde. Dies setzt natürlich auch voraus, dass die Personen, die an einem TRAPD-Verfahren beteiligt werden, qualifiziert für die Aufgabe sind. Selbst wenn eine Rückübersetzung/Backtranslation durch Studienvorgaben vorgesehen sein sollte, ist davon abzuraten, sie als alleinige/ausreichende Qualitätsmaßnahme zu betrachten.

3.3 Muss eine Fragebogenübersetzung gesondert getestet werden?

Ein übersetzter Fragebogen ist genauso sorgfältig an der Zielgruppe zu testen wie ein Fragebogen, der in der Zielsprache entwickelt wurde. Zur Testung können zum einen qualitative Verfahren zum Einsatz kommen – der GESIS Survey Guidelines Artikel „Kognitives Pretesting“ (Lenzner, Neuert & Otto, 2015) gibt hier Hilfestellung zur konkreten Durchführung von kognitiven Interviews. Zum anderen können quantitative Verfahren eingesetzt werden, die die Durchführung statistischer Analysen erlauben. Zu den quantitativen Verfahren gehört auch die split-ballot-Methode, die die Verwendung verschiedener Frageversionen in einem randomisierten Split und die anschließende Analyse vorsieht (Fowler, Jr., 2004; Krosnick, 2011). Die quantitative Test-Erhebung wiederum kann durch das so genannte Debriefing ergänzt werden, d. h. durch die Einholung zusätzlicher Informationen zum Befragungsverlauf im Anschluss an die eigentliche Befragung (Martin, 2004).

Für das Testen von Fragebogenübersetzungen gelten jedoch einige Besonderheiten, wenn das Ziel der Fragebogenübersetzung die Vergleichbarkeit mit dem Ausgangstext ist:

1. Deckt das Testen Probleme in der spezifischen Formulierung der Übersetzung auf, kann durch eine Umformulierung der Übersetzung das Problem häufig behoben werden. Beispielsweise kann ein Begriff, der nicht oder falsch verstanden wurde, durch einen passenderen Begriff oder eine Paraphrasierung ersetzt werden.
2. Deckt das Testen Probleme bei der Fragebogenübersetzung auf, die nicht die Formulierung der Übersetzung selber betreffen, sondern allgemeinerer Art sind (z.B. zu langer Erinnerungszeitraum), können diese Probleme nicht behoben werden, wenn der Ausgangstext bereits finalisiert und nicht mehr modifizierbar ist. Idealerweise werden daher schon während der Entwicklung des Ausgangstexts einer internationalen Umfrage in allen bzw. mehreren Sprachen Pretests durchgeführt, um Übersetzungsprobleme, kulturelle Probleme und allgemeine Probleme frühzeitig zu erkennen und anschließend beheben zu können. (s. Kapitel 2.2)

Statistische Überprüfung von Fragebogenübersetzungen – Besonderheiten in den verschiedenen Disziplinen

In der Psychologie oder Gesundheitsforschung gehört zum Gesamtkonzept der Erstellung einer Fragebogenübersetzung die psychometrische Überprüfung der Fragebogenübersetzung im Hinblick auf Reliabilität (siehe auch GESIS Survey Guidelines Artikel „Reliabilität“ (Danner, 2015)), Validität, Sensitivität/Spezifität (in der Gesundheitsforschung) und/oder Äquivalenz mit einer Ausgangsfassung. Diese Eckdaten werden häufig in Forschungsartikeln veröffentlicht.

In den Sozialwissenschaften werden Fragebogenübersetzungen zwar auch im Rahmen von quantitativen Vorstudien statistisch überprüft, insbesondere obliegt es dann aber Sekundärforschern, die Daten, die sie jeweils aus einer internationalen Studie verwenden, zum Beispiel dem ESS oder dem ISSP, vor ihrer Nutzung auf ihre Güte (z.B. Äquivalenz) hin zu testen. Statistische Qualitätsüberprüfung im Rahmen der Fragebogenentwicklung (-übersetzung) wurde jüngst jedoch auch für die sozialwissenschaftliche Umfrageforschung gefordert. Rammstedt et al. (2014) bieten in dieser Hinsicht einen Qualitätsstandard als allgemeine Richtschnur.

Die einschlägige Literatur in den verschiedenen (Sub-)Disziplinen ist im Hinblick auf empfohlene statistische Verfahren zu konsultieren. Schlagwörter für die Literaturrecherche stellen hierbei Äquivalenz/equivalence, Validierung/validation, psychometrische Überprüfung/psychometric testing oder Adaptation/adaptation dar. Darüber hinaus gibt das Kapitel 4 einige Einblicke.

3.4 Wo finde ich Übersetzer?

Geeignete Übersetzer zu finden, stellt durchaus eine Herausforderung dar, da der Bereich der Fragebogenübersetzung ein Nischenbereich ist, auf den sich nicht viele Übersetzer spezialisiert haben. Empfehlungen aus anderen Studien, nationale Übersetzerverbände – in Deutschland zum Beispiel der Bundesverband der Dolmetscher und Übersetzer e. V. (BDÜ) oder der Fachverband der Berufsübersetzer und Berufsdolmetscher e. V. (ATICOM) – oder Übersetzungshochschulen können erste Anlaufstellen für die Suche nach geeignetem Personal sein.

Insgesamt ist bei der Auswahl der Übersetzer darauf zu achten, dass sie

1. in ihre Muttersprache übersetzen;
2. sehr gute Kenntnisse der Ausgangssprache besitzen;
3. Übersetzungskompetenz besitzen (nicht notwendigerweise nachgewiesen durch ein Übersetzungsstudium), jedoch umfasst Übersetzungskompetenz wesentlich mehr als eine Fremdsprache mehr oder weniger zu verstehen (siehe EMT, 2009);
4. Sensibilität für verschiedene Sprachregister sowie insgesamt ein gutes Sprachgefühl mitbringen, um zielgruppenadäquat übersetzen zu können;
5. bereit sind, an Übersetzungsdiskussionen aktiv teilzunehmen, sofern der oben beschriebene TRAPD-Ansatz Anwendung findet, einschließlich der Bereitschaft, die eigene sowie Übersetzungen von Dritten konstruktiv-kritisch zu bewerten;
6. idealerweise über die Besonderheiten von Fragebogenübersetzungen Kenntnis haben;
7. idealerweise Kenntnisse zur Fragebogenentwicklung und zur Operationalisierung von Konzepten mitbringen;
8. idealerweise fachliche Kenntnisse im betreffenden Themengebiet mitbringen.

Gerade die letzten drei Punkte können durch entsprechendes Briefing und/oder eine Schulung, aufgefangen werden, wenn Übersetzer dieses Wissen nicht mitbringen. Darüber hinaus gewährleistet der oben beschriebene Team-Ansatz, dass die notwendige fachliche Kompetenz im Rahmen des Übersetzungsprozesses durch andere Personen, z.B. Sozialwissenschaftler oder Psychologen in dem betreffenden Fachgebiet, eingebracht werden kann.

An dieser Stelle soll des Weiteren auf die Cross-Cultural Survey Guidelines verwiesen werden und dabei insbesondere auf Vorschläge zur Rekrutierung von Übersetzern: <http://ccsg.isr.umich.edu/index.php/chapters/translation-chapter/finding-selecting-and-briefing-translation-team-members>

3.5 Welche Informationen sollten Übersetzern zur Verfügung gestellt werden?

Wie viele und welche Informationen den ÜbersetzerInnen zur Verfügung gestellt werden, hängt von verschiedenen Faktoren ab. Stellt beispielsweise ein bestimmtes Projekt oder eine bestimmte Umfrage bestimmte Anforderungen an das Fachvokabular, oder gibt es ein spezielles Zielpublikum, das in der Wahl des Sprachregisters berücksichtigt werden muss? Besitzen die ÜbersetzerInnen bereits Erfahrung in der Fragebogenübersetzung, oder müssen ihnen bestimmte Erfordernisse dieser Textgattung überhaupt erst vermittelt werden?

Folgende Angaben können als Richtlinie in Bezug auf die Informationen fungieren, die den ÜbersetzerInnen zur Verfügung gestellt werden sollten:

1. Hintergrundinformationen zu der betreffenden Umfrage / Studie, beispielsweise in Bezug auf Ziele und Inhalt der Umfrage oder Entwicklung des Fragebogens

2. Zielpublikum, in Bezug auf z. B. Alter, Bildung, Nationalitäten, spezielle Bevölkerungs- oder Berufsgruppen
3. Befragungsmodus: Erfolgen die Befragungen mündlich, als papierbasierter Selbstausfüller oder webbasiert?
4. Wird dieser Fragebogen nur in eine Sprache oder noch in andere Sprachen übersetzt?
5. Sind Übersetzungen in bestehende Übersetzungen einzufügen, das heißt, müssen bestehende Übersetzungen übernommen werden? Dies ist insbesondere bei wiederholten Querschnittsbefragungen und Panelbefragungen zu beachten, bei denen in der Regel nur ein Teil der Fragen neu zu übersetzen ist. Diese neu zu übersetzenden Fragen müssen sich aber nahtlos in die wiederholten Fragen einfügen, z.B. was die Art der Frageformulierung und die Formulierung der Antwortkategorien betrifft.
6. Ist der Ausgangstext final oder können am Ausgangstext noch Änderungen erfolgen, beispielsweise aufgrund von Fragen, die bei der Übersetzung auftauchen?
7. An wen können die ÜbersetzerInnen sich bei Fragen wenden?
8. Welche Methode wird zur Überprüfung bzw. Gewährleistung der Übersetzungsqualität angewandt?

Zudem sind den ÜbersetzerInnen die unter Kapitel 3.1 genannten Punkte, die generelle Richtlinien von Fragebogenübersetzungen betreffen (z.B. sprachliche Korrektheit, Vollständigkeit, Verständlichkeit, Berücksichtigung von Höflichkeitsnormen, Konsistenz), in jedem Fall mitzuteilen.

3.6 Ich brauche Übersetzungen nicht nur für ein anderes Land, sondern auch für eine besondere Migrantengruppe in diesem Land. Was ist hierbei zu berücksichtigen?

Prinzipiell gelten bei der Fragebogenübersetzung für eine Migrantengruppe in einem Land dieselben Regeln und Anforderungen, die hier in den Guidelines erörtert werden. Was Übersetzungen in Migrantsprachen von Übersetzungen in Sprachen anderer Länder unterscheidet, ist bislang kaum erforscht, wenn überhaupt im US-amerikanischen Kontext. Einige Aspekte, die durch (amerikanische) Praxiserfahrungen angezeigt wurden, können jedoch stichpunktartig genannt werden.

1. Aus welchem Land bzw. aus welchen Ländern kommt die Zielgruppe? Dies ist bei der Übersetzung zu berücksichtigen. Wird ein Fragebogen z.B. im Spanischen angeboten, ist das Herkunftsland der Zielpopulation(en) (z.B. verschiedene Spanisch-sprachige Länder mit etwaigen Unterschieden in der Sprache) bei der Formulierung von Fragebogenitems zu berücksichtigen (Martinez, Marín, & Schoua-Glusberg, 2006; Alegria et al., 2004).
2. Zuweilen macht es Sinn, einige Namen, z.B. von bestimmten staatlichen Einrichtungen oder Programmen, in der Ausgangssprache des Fragebogens beizubehalten, da diese den Migranten eher in der Ausgangssprache als in der Muttersprache bekannt sind. Stadler (2010) nennt für Österreich den Begriff „Arbeitsmarktservice“ (als österreichische Bezeichnung für Konjunkturprogramme), der – in die jeweiligen Migrantsprachen der Umfrage übersetzt – von den Migranten nicht verstanden wurde.
3. Wenn Fragen z.B. auf Deutschland bezogen werden, sollte die Übersetzung derart formuliert werden, dass der Bezugsrahmen Deutschland deutlich wird und nicht durch die Übersetzung der Bezugsrahmen des Ursprungslandes suggeriert wird. Hierbei kann auch die ausgangssprachliche Nennung eines Konzepts, wie oben aufgeführt, hilfreich sein.
4. Die unterschiedliche Perspektive, die Migranten und Einheimische auf Deutschland haben, ist bei Übersetzungen zu beachten. Was für den einen als „ausländisch“, „anderes/fremdes Land“, „Ausland“ gilt, ist nicht notwendigerweise das, was der andere darunter versteht. Daher sind bei

Übersetzungen die jeweiligen Bezugsrahmen zu hinterfragen. Als Beispiel kann hier die chinesische und koreanische Übersetzung (für die USA) der Frage „Where was this person born?“ angeführt werden. Eine der Antwortkategorien lautete: „Outside the United States - Print name of foreign country, or Puerto Rico, Guam, etc.“. Im Chinesischen und Koreanischen wurde „foreign country“ jeweils nur durch „country“ übersetzt, um Befragte, die hier ihr Heimatland eintragen sollen, nicht zu verwirren (Sha, Park, & Pan, 2012).

5. Fragen werden möglicherweise vor einem anderen kulturellen Hintergrund anders verstanden. Die Frage „In what year did this person¹ last get married?“ verursachte in einer ersten Fassung der koreanischen Übersetzung für die USA einige Probleme. Hochzeit wurde von vielen koreanischen Befragten in einem kognitiven Pretest als Life-time-event angesehen – man heiratet nur einmal. Vor diesem Hintergrund wurde die Übersetzung dahingehend interpretiert, wann innerhalb der letzten Zeit geheiratet wurde. Seit langem Verheiratete bezogen die Frage somit nicht auf sich. Eine Lösung des Problems erfolgte durch das Weglassen des Wortes „last“ in der Frage und eine entsprechende zusätzliche Befragtenanweisung (Sha et al., 2012).
6. Fragen, die in nationalen Umfragen gestellt werden, gelten möglicherweise als Tabu-Themen und können somit nicht gefragt werden (Sauer, 2008).
7. Eine weitere Schwierigkeit ergibt sich, wenn Migrantengruppen die Sprache ihres Ursprungslandes nicht (mehr) so sprechen, wie dies dort der Fall ist. Die kann z.B. daran liegen, dass vornehmlich eine besondere Gruppe aus diesem Land ausgewandert ist oder dass Elemente der Sprache des Ziellandes assimiliert worden sind. Dies ist bei der Suche nach geeignetem Übersetzungspersonal zu berücksichtigen.

3.7 Ist Fragebogenübersetzung gleich Fragebogenübersetzung? Welche Typen der Fragebogenübersetzung gibt es?

Die obigen Ausführungen haben sich mit der Fragebogenübersetzung im Rahmen intentionaler angelegter internationaler Studien befasst. Es gibt jedoch noch weitere Typen der Fragebogenübersetzung, die mehr oder weniger stark mit Adaptationen, d.h., kulturellen Adaptationen, im Zusammenhang stehen. Diese sollen nun vorgestellt werden.

Wenn die Übersetzung eines Fragebogens ansteht, können folgende Leitfragen helfen, sich zunächst einen Überblick über die allgemeinen Erfordernisse und Einschränkungen einer Fragebogenübersetzung zu verschaffen. Denn Fragebogenübersetzung ist nicht gleich Fragebogenübersetzung, wie im weiteren Verlauf deutlich wird.

Die Leitfragen sind:

1. Wurde der Ausgangsfragebogen, also der Fragebogen, der übersetzt werden soll, ursprünglich für eine bestimmte Kultur oder für ein bestimmtes Land entwickelt?
ODER
Wurde der Ausgangsfragebogen bereits mit Blick auf seine internationale Verwendung hin entwickelt?
2. Ist es das Ziel der Fragebogenübersetzung, Vergleichbarkeit mit dem Ausgangsfragebogen herzustellen?
ODER

¹ Gemeint ist der oder die Befragte.

Dient der Ausgangsfragebogen lediglich als Ideenquelle für einen neuen Fragebogen in einer anderen Sprache?

Aus diesen Fragen ergibt sich folgende Tabelle:

Tabelle 1: Unterschiedliche Formen der Fragebogenübersetzung

	Vergleichbarkeit mit Ausgangsfragebogen geplant	Vergleichbarkeit mit Ausgangsfragebogen <u>nicht</u> geplant
Entwicklung des Ausgangsfragebogens für eine bestimmte Kultur/bestimmtes Land	I: Adaptationen wahrscheinlich notwendig, jedoch eingeschränkt durch Forderung nach Vergleichbarkeit	II: Adaptationen wahrscheinlich notwendig, keine Einschränkungen
Entwicklung des Ausgangsfragebogens unter Berücksichtigung seiner internationalen Verwendung	III: Adaptationen möglicherweise notwendig, häufig bereits gesondert gekennzeichnet; weitere Adaptationen bedürfen i. d. R. einer Genehmigung in internationalen Studien	IV: Adaptationen möglicherweise notwendig, keine Einschränkung

Die vier Tabellenfelder werden nun im Hinblick darauf erläutert, wo sich eine Fragebogenübersetzung im Spannungsfeld zwischen Übersetzung und möglicher Adaptation befinden könnte (Behr, 2013). *Adaptation* steht in diesem Zusammenhang für Änderungen, die über die „typischen“ sprachlichen Änderungen im Rahmen einer Übersetzung hinausgehen und die notwendig sind, um einen Fragebogen an die Bedingungen eines neuen Kulturraums anzupassen. Siehe zu konkreten Adaptationen das hierauf folgende Kapitel 3.8.

Situation I:

Wird ein Fragebogen, der z.B. ursprünglich für den US-amerikanischen Kontext entwickelt wurde, nun als Grundlage für einen deutschen Fragebogen verwendet und soll eine möglichst große Vergleichbarkeit der Daten angestrebt werden, so gilt Folgendes: Adaptationen sind zwar einerseits erforderlich, um den Fragebogen aus seiner amerikanischen Verankerung zu holen und an deutsche Lebensbedingungen anzupassen, andererseits sind diese Änderungen jedoch auch auf das Notwendigste zu beschränken – andernfalls kann die Vergleichbarkeit der Daten nicht gewährleistet werden.

Eine Adaptation wäre z.B. für folgendes Item notwendig: „Do you have difficulty walking several blocks?“. Das „block“-System bildet das amerikanische Städtebausystem ab, im Deutschen könnte hier je nach Messintention eine Meter-Angabe stehen oder eine äquivalente Umschreibung stehen (Harkness, Edwards, & Villar, 2010).

Situation II:

Wird ein Fragebogen, der z.B. ursprünglich für den amerikanischen Kontext entwickelt wurde, nun als Grundlage für einen deutschen Fragebogen verwendet und soll in diesem Zusammenhang keine Vergleichbarkeit der Daten angestrebt werden, so ist man bei der Fragebogenübersetzung grundsätzlich einmal nicht gebunden. Änderungen können kulturell bedingt sein, aber auch schlichtweg nur das Ziel verfolgen, das Messinstrument für den eigenen Forschungsbedarf zu optimieren, z.B. durch die

Änderung einer 7er-Skala in eine 5er-Skala oder durch das Hinzufügen oder Streichen von Themeneinleitungen oder ganzer Fragen usw.

Situation III:

Wird ein Ausgangsfragebogen neu für eine international-vergleichende Studie entwickelt, so wird dieser idealerweise so formuliert, dass entweder eine Übersetzung in den teilnehmenden Ländern der Studie möglich ist oder dass erlaubte Adaptationen (z.B. das Hinzufügen länderspezifischer Beispiele oder Antwortkategorien) bereits im Vorfeld gesondert gekennzeichnet sind. Darüber hinausgehende Adaptationen bedürfen in der Regel der Genehmigung einer zentralen Koordinierungsinstanz und sind auf das Notwendigste beschränkt. Situation III entspricht z.B. der Situation im European Social Survey (ESS) oder dem International Social Survey Programme (ISSP). Diese Situation bildete die Grundlage für die Beschreibung der vorangegangenen Kapitel.

Situation IV:

Soll ein Fragebogen nur als Ideenquelle für einen neu zu entwickelnden Fragebogen dienen, so hat man Freiheiten bei der Fragebogenübersetzung in Bezug auf Änderungen, die man am Fragebogen vornehmen möchte. Da der Ausgangsfragebogen bereits mit Blick auf seine internationale Verwendung entwickelt wurde, ist möglicherweise eine Reihe von Fragen übersetzbar und muss nicht adaptiert werden.

Was im konkreten Fall noch übersetzt werden kann bzw. wo schon adaptiert werden muss oder was erlaubt ist, ist über die genannten Rahmenbedingungen hinaus auch von folgenden Faktoren abhängig:

Thema

Was ist das Thema des Fragebogens? Allgemeine Fragen zur Zufriedenheit in verschiedenen Lebensbereichen können typischerweise eher übersetzt werden als Fragen, die sich mit spezifischen Strukturen des Gesundheitssystems befassen.

Sprachliche und kulturelle Distanz

Wie weit sind Ausgangssprache und -kultur und Zielsprache und -kultur jeweils voneinander entfernt? Je größer die jeweilige Distanz, umso eher ergibt sich die Notwendigkeit zur Adaptation.

Studienrichtlinien

Welche Vorgaben werden in Studienrichtlinien gemacht? Dies betrifft insbesondere große internationale Studien, die in der Regel den Ländern vorgeben, „mit welchem Freiheitsgrad“ ein Fragebogen in eine neue Sprache übertragen werden soll.

Rechtliche Einschränkungen

Gibt es ein Copyright auf den Fragebogen? Dies betrifft insbesondere Fragen aus der Psychologie oder der Gesundheitsforschung und weniger Fragen aus den sozialwissenschaftlichen Disziplinen. Im Fall von Copyright sind Übersetzungen und darüber hinaus mögliche Adaptationen mit entsprechenden Copyright-Besitzern abzusprechen.

Neben den oben genannten Fragebogenübersetzungen, die auf ein funktionierendes Messinstrument für den neuen Kulturraum abzielen, gibt es zusätzlich noch die „dokumentarische“ Fragebogenübersetzung. So wird zum Beispiel der ALLBUS, die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften in Deutschland, ins Englische übersetzt. Ziel hierbei ist es, Wissenschaftlern, die kein Deutsch sprechen, Einblicke in den deutschen ALLBUS zu geben. Es wird explizit darauf hingewiesen, dass die übersetzten Fragen für einen Einsatz in einem englischsprachigen Land noch einmal überarbeitet werden müssten.

3.8 Welche Formen der Frage(bogen)-Adaptation gibt es?

Der Begriff Adaptation wird nicht nur von unterschiedlichen Wissenschaftlern, sondern auch über Disziplinen hinweg sehr unterschiedlich verstanden (Behr, 2013).

Zum einen wird Adaptation als übergeordneter Begriff für den Gesamtprozess der Übertragung eines Messinstruments von einer Sprache/Kultur in eine andere Sprache/Kultur verwendet. Diese Art der Verwendung trifft insbesondere auf die Gesundheitsforschung sowie die kulturvergleichende psychologische Forschung zu. Der Begriff Adaptation soll hierbei signalisieren, dass eine (semantisch vergleichbare) Übersetzung mitunter nicht ausreicht, sondern Änderungen auf verschiedener Ebene vorgenommen werden müssen. Des Weiteren soll dieser Begriff in seiner übergeordneten Form auch auf die Notwendigkeit der psychometrischen Testung eines neuen Messinstruments hinweisen.

Zum anderen wird Adaptation für unterschiedliche spezifische Änderungen auf Item-Ebene verwendet. In diesem Kontext kann, vereinfacht gesprochen, zwischen Adaptationen im Bereich Kultur, Messung und Sprache unterschieden werden, wobei Adaptationen im Bereich der Sprache vielfach auch das Rüstzeug einer ganz normalen, guten *Übersetzung* darstellen.

Adaptationen im Bereich Kultur

Hierunter fallen Änderungen, die durch unterschiedliche Systeme im weitesten Sinne in den einzelnen Ländern und Kulturen bedingt sind. Konkrete Verweise auf das politische System, das Bildungssystem, das Gesundheitssystem und Ähnliches bedürfen häufig einer Anpassung im Fragebogen. Ebenso gilt dies für Verweise auf Messeinheiten (z.B. miles/meter). Auch Layout-technisch sind häufig Änderungen notwendig, um z.B. unterschiedliche Namens-, Datums- oder Adresskonventionen zu berücksichtigen.

Neben den oben genannten stärker faktisch bedingten Änderungen am Fragebogen gibt es Änderungen, die bedingt sind durch unterschiedliche Werte, Sitten oder Lebensgewohnheiten. Die Frage, ob man schon einmal einen „campaign badge/sticker“ getragen hat, sollte in Ländern, in denen „badge/sticker“ nicht zum Inventar der Symbole politischer Partizipation gehören, nicht in dieser Form gestellt werden. Das Item „How difficult was it to sit without losing your balance?“ misst, inwieweit das Gleichgewicht beim Sitzen auf einem Stuhl gehalten werden kann. Für den japanischen Kontext ist die immer noch übliche Tradition, auf dem Boden zu sitzen, zu berücksichtigen. Die intendierte Messung kann durch eine Spezifizierung „How difficult was it to sit on a chair without losing your balance?“ gewährleistet werden (Breugelmans, 2009).

Weitere Themengebiete, die häufig einer Adaptation bedürfen, sind Familien-bezogene Themen, religiöse Themen, Freizeitaktivitäten, Nahrungsmittel oder Medikamente.

Adaptationen im Bereich Messung

Hierunter fallen Änderungen, die durch unterschiedliche Erfahrungen mit Messinstrumenten und Umfragen bedingt sind. Bei der Übertragung eines Fragebogens in eine neue Sprache und Kultur müssen ggf. Instruktionen/Erklärungen/Bilder zu Skalenlabels und Ähnliches ergänzt werden, wenn die neue Zielgruppe wenig oder keine Erfahrung im Umgang mit Messinstrumenten hat.

Darüber hinaus müssen ggf. Skalen angepasst werden, um Diskursnormen in der neuen Kultur Rechnung zu tragen. Als Beispiel in dieser Hinsicht wird zuweilen die Anpassung der agree-disagree-Skala in der japanischen Sprache zitiert, wobei die englischen Endlabels „strongly agree“/„strongly disagree“ im Japanischen eine weniger starke Ausprägung erfahren (Smith, 2004).

Adaptationen im Bereich Sprache

Hierunter fallen Änderungen, die schlichtweg durch unterschiedliche Sprachsysteme hervorgerufen werden, wie z.B. Änderungen der Satzstellung, der Wortart, Gender-Spezifizierungen usw. Auch wenn diese Änderungen in der Literatur durchaus als „language-driven adaptations“ oder „linguistics-driven adaptations“ (z.B. Harkness, 2008) bezeichnet werden, sollten sie doch als zentrale Merkmale einer guten *Übersetzung* aufgefasst werden. Übersetzung an sich sollte eben nicht als eine rigide, starre Wort-für-Wort-Übertragung, sondern als Übertragung von Bedeutung, unter Berücksichtigung der jeweiligen sprachlichen und idiomatischen Anforderungen der Zielsprache, aufgefasst werden!

Darüber hinaus ist ein Fragebogen im Rahmen einer Übersetzung an die Diskursnormen einer neuen Kultur anzupassen. So sind z.B. unterschiedliche Höflichkeitskonventionen bei der Anrede oder dem Formulieren von Fragen/Instruktionen zu berücksichtigen. Auch diese Änderungen gehören im Wesentlichen zum Rüstzeug einer guten *Übersetzung*. Sie bewirken, dass eine Übersetzung als Original wahrgenommen wird und nicht als Übersetzung.

Handelt es sich bei der neuen Zielgruppe um eine Bevölkerungsgruppe mit geringerer Bildung, als es beim Original-Fragebogen der Fall war, so muss ggf. sprachlich eine Vereinfachung des Fragebogens erfolgen. Auch losgelöst von einem Bildungsunterschied sind Begriffe, wenn sie von der Zielgruppe nicht verstanden werden, zu paraphrasieren oder zu erklären. Hier verschwimmen sicherlich die Grenzen von dem, was die einen noch als *Übersetzung* und die anderen schon als Adaptation bezeichnen.

3.9 Was muss ich bei Projektanträgen berücksichtigen, in deren Rahmen eine internationale bzw. interkulturelle Studie durchgeführt werden soll?

Es war – in der Vergangenheit zumindest – nicht unüblich, dass im Rahmen eines inhaltlichen Forschungsprojektes das Thema Fragebogenübersetzung erst dann Beachtung fand, wenn es im Projekt konkret anstand; Geld oder Zeit wurden für die Fragebogenübersetzung kaum eingeplant.

Vor dem Hintergrund der weitreichenden Konsequenzen, die eine schlechte Übersetzung für die Forschungsergebnisse haben kann, sind bereits bei der Stellung eines Forschungsantrags die Vielschichtigkeit der Übersetzung und der Zusammenhang zwischen guter Fragebogenübersetzung und guter Fragebogenentwicklung zu berücksichtigen. Ist im Rahmen eines neuen Projektantrags eine internationale oder interkulturelle Studie geplant, so gilt:

1. Bereits während der Ausgangstextentwicklung ist die internationale/interkulturelle Perspektive durch die Einbindung von Personen(-gruppen) aus anderen Ländern bzw. Kulturen sowie den Einsatz verschiedener qualitativer und quantitativer Testverfahren zu berücksichtigen.

2. Die Erstellung einer Fragebogenübersetzung erfordert die Einbindung verschiedener Personen(-gruppen) sowie das Zusammenspiel verschiedener Verfahren zur Qualitätssicherung.
3. Eine internationale/interkulturelle Studie hat Auswirkungen auf die zeitliche, personelle und finanzielle Planung eines Forschungsprojektes.

4. Überprüfung von Messinstrumenten ex-post

Zur Überprüfung von Äquivalenz ex-post gibt es eine Reihe quantitativer Verfahren. Das wohl mit Abstand am häufigsten angewandte Verfahren ist die konfirmatorische Faktorenanalyse (Brown, 2006; Vandenberg & Lance, 2000). Für den interkulturellen Vergleich bietet sich dabei ein Gruppenvergleich an („multi-group confirmatory factor analysis“). Dabei wird geprüft, ob ein Messinstrument, das sich in einem Land bewährt hat, auch für die anderen Länder zutrifft. Vergleichbarkeit kann dabei auf unterschiedlichen Ebenen bestehen, die den zuvor diskutierten Äquivalenzkriterien entsprechen: Konfigurale Invarianz ist gegeben, wenn alle Items in den verschiedenen Ländern zur gleichen latenten Dimension gehören. Metrische Invarianz erfordert überall gleiche Ladungen der Items auf dem Faktor, und skalare Invarianz setzt zusätzlich voraus, dass alle Interzepte über die Länder gleich sind. Ein Beispiel für einen solchen Gruppenvergleich mittels einer konfirmatorischen Faktorenanalyse stellen die Arbeiten von Davidov (2008) sowie Davidov, Schmidt & Schwartz (2008) zu allgemeinen Werten dar.

Die Mehrebenenanalyse ist ein statistisches Verfahren, das explizit das Vorhandensein unterschiedlicher Ebenen, d.h. der Befragten und der Länder, berücksichtigt (Hox, 2010; Snijders & Bosker, 2011). Die Mehrebenenanalyse ist besonders dann nützlich, wenn die Zahl der Länder groß ist (20 wird vielfach als Minimum angesehen) und wenn es darum geht, Variablen auf der Länderebene zur Erklärung von Äquivalenzproblemen explizit einzubeziehen. Die Mehrebenenanalyse lässt sich auch mit der konfirmatorischen Faktorenanalyse kombinieren („multi-level confirmatory factor analysis“); dies erfordert allerdings deutlich höhere Fallzahlen auf Länderebene.

Eine besonders in der Psychologie verwendete Verfahren ist die Item Response Theorie (IRT), die die Beziehung der einzelnen Items zu der zugrundeliegenden Dimension analysiert (Embretson & Reise 2000, van de Vijver & Leung, 1997). IRT benötigt eine große Anzahl an Items, was die Anwendbarkeit für viele internationale sozialwissenschaftliche Umfragen einschränkt.

Es gibt eine Reihe weiterer Methoden zur Überprüfung der Äquivalenz (und damit der Vergleichbarkeit über Länder hinweg), die zwar nicht unbedingt die gleiche Präzision mitbringen, aber teilweise deutlich einfacher zu handhaben sind. Statt einer konfirmatorischen Faktorenanalyse lässt sich z.B. eine exploratorische Faktorenanalyse rechnen (van de Vijver & Leung, 1997). Die (multiple) Korrespondenzanalyse (Greenacre & Blasius 2006) stellt praktisch ein Hauptkomponentenverfahren für kategoriale Daten dar. Blasius & Thiessen (2006) wenden die Korrespondenzanalyse auf die Geschlechterrollen-Items des ISSP an. Die multidimensionale Skalierung (Borg & Groenen 2005) stellt die Interkorrelationen zwischen den Items graphisch dar. Die Daten müssen mindestens ordinal sein. Das Kriterium für die Vergleichbarkeit ist, ob die Items in den einzelnen Ländern jeweils in den entsprechenden Regionen verortet sind. Dabei werden die Trennlinien zwischen den Regionen aus der Theorie abgeleitet. Braun & Johnson (2011) wenden eine ganze Reihe – komplexer und weniger komplexer – Verfahren auf das gleiche inhaltliche Problem an und können dadurch zeigen, dass sie in vielen Fällen zu den gleichen Schlussfolgerungen führen.

Ein Nachteil aller quantitativen Verfahren – vielleicht mit der Ausnahme der Mehrebenenanalyse, wenn man Variablen auf der Länderebene aufnimmt, die die Inäquivalenzen (potentiell) erklären können –

ist, dass sie zwar das Vorhandensein von Problemen feststellen, deren eigentliche Ursachen aber nicht erklären können. Hier bieten sich dann – wie bereits bei der Sicherstellung von Äquivalenz ex-ante (also im Rahmen von Pretests) – Probingtechniken an. Da die Durchführung kognitiver Interviews – zumal im internationalen Vergleich – äußerst aufwendig ist, erscheint hier die Durchführung zusätzlicher webbasierter Studien eine mögliche Vorgehensweise. Hierbei können Probingfragen – entsprechend Schumans (1966) Vorschlag für „random probes“ – in einen normalen Web-Fragebogen eingeschaltet werden. Dieses Vorgehen wurde z.B. in den bereits unter Kapitel 2.2 beschriebenen Arbeiten von Behr et al. (2014) sowie Braun, Behr & Kaczmirek (2013) angewendet.

Literaturverzeichnis

- Alegría, M., Vila, D., Woo, M., Canino, G., Takeuchi, D., Vera, M., ... Shrout, P. (2004). Cultural relevance and equivalence in the NLAAS Instrument: Integrating etic and emic in the development of cross-cultural measures for a psychiatric epidemiology and services study of Latinos. *International Journal of Methods in Psychiatric Research*, 13(4), 270-288.
- Behr, D. (2013). On the different uses and users of the term adaptation. Vortrag anlässlich der AAPOR-Konferenz, 16.05.-19.05.2013. Boston, US.
- Behr, D., Braun, M., Kaczmirek, L., & Bandilla, W. (2014). Item comparability in cross-national surveys: Results from asking probing questions in cross-national web surveys about attitudes towards civil disobedience. *Quality & Quantity*, 48, 127-148.
- Biemer, P.P., Groves, R.M., Lyberg, L., Mathiowetz, N.A., & Sudman, S. (Eds.) (1991). *Measurement errors in surveys*. New York: Wiley.
- Blasius, J., & Thiessen, V. (2006). Assessing data quality and construct comparability in cross-national surveys. *European Sociological Review*, 22, 229-242.
- Borg, I., & Groenen, P.J.F. (2005). *Modern multidimensional scaling (2nd ed.)*. New York: Springer.
- Braun, M. (2006). *Funktionale Äquivalenz in interkulturell vergleichenden Umfragen. Mythos und Realität*. Mannheim: ZUMA.
- Braun, M., Behr, D., & Kaczmirek, L. (2013). Assessing cross-national equivalence of measures of xenophobia: Evidence from probing in web surveys. *International Journal of Public Opinion Research*, 25, 383-395. DOI:10.1093/ijpor/eds034.
- Braun, M., & Johnson, T.P. (2010). An illustrative review of techniques for detecting inequivalences. In Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., & Smith, T. (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 375-393). Hoboken, NJ: Wiley.
- Breugelmans, R. (2009). Dangers in using translated medical questionnaires. The importance of conceptual equivalence across languages and cultures in patient-reported outcome measures. *Chest*, 136, 1175-1177.
- Brislin, R. W. (1986). The wording and translation of research instruments. In W.J. Lonner & J. W. Berry (Eds.), *Field Methods in Cross-Cultural Research* (pp. 137-164). Beverly Hills: Sage.
- Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Danner, D. (2015). *Reliabilität – die Genauigkeit der Messung*. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/sdm-sg_011

- Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. *Survey Research Methods*, 2, 33-46.
- Davidov, E., Schmidt, P., & Billiet, J. (2011). *Cross-cultural analysis: Methods and applications*. New York: Routledge.
- Davidov, E., Schmidt, P., & Schwartz, S.H. (2008). Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly*, 72, 420-445.
- Dean, E., Caspar, R., McAviney, G., Reed, L., & Quiroz, R. (2007). Developing a low-cost technique for parallel cross-cultural instrument development: The question appraisal system (QAS-04). *International Journal of Social Research Methodology*, 10, 227-241.
- Dept, S. (2013). Translatability assessment of draft questionnaire items. Paper presented at the 5. Conference of the European Survey Research Association (ESRA), Ljubljana.
<http://www.europeansurveyresearch.org/conference/programme?sess=38&tday=1>
- Dorer, B. (2011). Advance translation in the 5th round of the European Social Survey (ESS). FORS Working Paper Series, paper 2011-4. Lausanne: FORS
- EMT expert group (2009). Competences for professional translators, experts in multilingual and multimedia communication
http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf
- European Social Survey (2014). *ESS Round 7 Translation Guidelines*. London: ESS ERIC Headquarters, Centre for Comparative Social Surveys, City University London.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah NJ: Lawrence Erlbaum.
- Farrall, S., Priede, C., Ruuskanen, E., Jokinen, A., Galev, T., Arcai, M., & Maffei, S. (2012). Using cognitive interviews to refine translated survey questions: An example from a cross-national crime survey. *International Journal of Social Research Methodology*, 15, 467-483.
- Fitzgerald, R., Widdop, S., Gray, M., Collins, D. (2011) Identifying sources of error in cross-national questionnaires: Application of an error source typology to cognitive interview data. *Journal of Official Statistics*, 27, 569-599.
- Fowler, Jr., F. J. (2004). The case for more split-sample experiments in developing survey instruments. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 173-171). Hoboken, NJ: Wiley.
- Goerman, P.L., & Caspar, R. A. (2010). A preferred approach for the cognitive testing of translated materials: Testing the source version as a basis for comparison. *International Journal of Social Research Methodology*, 13, 303-316.
- Greenacre, M...J., & Blasius, J.. (Eds.) (2006). *Multiple correspondence analysis and related methods*. Boca Raton, FL: Chapman & Hall.
- Harkness, J. (2003). Questionnaire Translation. In J. Harkness, F. van de Vijver, & P. Mohler (Eds). *Cross-Cultural Survey Methods* (pp. 35-56). Hoboken, NJ: Wiley.
- Harkness, J. A. (2008). Comparative survey research: Goal and Challenges. In E. de Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 56-77). New York: Lawrence Erlbaum Associates.

- Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., & Smith, T. (Eds.) (2010), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 375-393). Hoboken, NJ: Wiley.
- Harkness, J., van de Vijver, F.J.R., & Johnson, T.P. (2003). Questionnaire design in comparative research. In Harkness, J., van de Vijver, F.J.R., & Mohler, P.P. (Eds.), *Cross-cultural survey methods* (pp. 19-34). Hoboken, NJ: Wiley.
- Harkness, J.A., Edwards, B., Hansen, S.E., Miller, D.R. & Villar, A. (2010). Designing questionnaires for multipopulation research. In Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., & Smith, T. (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (33-58). Hoboken, NJ Wiley.
- Harkness, J., Bilgen, I., Córdova Cazar, A., Hu, M., Huang, L., Lee, S., Mingnan, L., Miller, D., Stange, M., Villar, A., & Yan Ti. (2016). Cross-cultural survey guidelines. *Questionnaire Design*. URL: <http://ccsg.isr.umich.edu/index.php/chapters/questionnaire-design-chapter>
- Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 117-140). Hoboken, NJ: Wiley.
- Harkness, J., van de Vijver, F.J.R., & Mohler, P.P. (Eds.) (2003), *Cross-cultural survey methods*. Hoboken, NJ: Wiley.
- Hox, J.J. (2010). *Multilevel analysis. Techniques and applications* (Second Edition). New York: Routledge.
- Johnson, T.P., & van de Vijver, F.J.R. (2003), Social desirability in cross-cultural research. In Harkness, J., van de Vijver, F.J.R., & Mohler, P.P. (Eds.) (2003), *Cross-cultural survey methods* (pp. 195-204). Hoboken NJ: Wiley.
- Koch, A. & Blohm, M. (2015). Nonresponse bias. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/sdm-sg_004
- Krosnick, J. A. (2011). Experiments for evaluating survey questions. In J. Madans, K. Miller, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: contributing to the science of data collection* (pp. 215-238). Hoboken, NJ: Wiley.
- Lenzner, T., Neuert, C. & Otto, W. (2015). Kognitives Pretesting. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/sdm-sg_010
- Martin, E. (2004). Vignettes and respondent debriefing for questionnaire design and evaluation. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 149-188). Hoboken, NJ: Wiley.
- Martinez, M., Marín V., & Schoua-Glusberg, A. (2006). Translating from English to Spanish: The 2002 national survey of family growth. *Hispanic Journal of Behavioral Sciences*, 28(4), 531-545.
- Miller, K., Fitzgerald, R., Padilla, J.-L., Willson, S., Widdop, S., Caspar, R., Dimov, M., Grey, M., Nunes, C., Prüfer, P., Schöbi, N., & Schoua-Glusberg, A. (2011). Design and analysis of cognitive interviews for comparative multinational testing. *Field Methods*, 23, 379-396.
- Rammstedt, B., Beierlein, C., Brähler, E., Eid, M., Hartig, J., Kersting, M., ... Weichselgartner, E. (2014). Qualitätsstandards zur Entwicklung, Anwendung und Bewertung von Messinstrumenten in der sozialwissenschaftlichen Umfrageforschung. *RatSWD Working Papers*, 230. Retrieved from http://ratswd.de/dl/RatSWD_WP_230.pdf

- Sauer, M. (2008). Methodik und Durchführung von standardisierten Befragungen türkeistämmiger Migranten in Deutschland durch die Stiftung Zentrum für Türkeistudien. In K.-S. Rehberg (Ed.), *Die Natur der Gesellschaft: Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel* (pp. 3901-3907). Frankfurt am Main: Campus Verlag GmbH.
- Schoua-Glusberg, A, Villar, A. (2013). Backtranslation vs. Committee Approach: An Experiment Comparing How They Perform in Questionnaire Translation. Presentation given at FCSM Conference, Washington DC, November 5, 2013.
- Schuman, H. (1966). The random probe: A technique for evaluating the validity of closed questions. *American Sociological Review*, 31, 218-222.
- Sha, M, Park, H., & Pan, Y. (2012). Translation review and cognitive testing of ACS language assistance guides in multiple languages. Retrieved from http://www.census.gov/library/working-papers/2012/acs/2012_Sha_01.html
- Smith, T.W. (2003). Developing comparable questions in cross-national surveys. Harkness, J., van de Vijver, F.J.R., & Mohler, P.P. (Eds.), *Cross-cultural survey methods* (69-91). Hoboken, NJ: Wiley.
- Smith, T. (2004). Developing and evaluating cross-national survey instruments. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 431-452). Hoboken, NJ: Wiley.
- Snijders T.A.B., & Bosker, R.J. (2011). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. 2nd ed. London: Sage.
- Stadler, B. (2010). Herausforderungen der Befragung von Migrantinnen und Migranten im Mikrozensus. *Austrian Journal of Statistics*, 39(1&2), 139-145.
- Thrasher, J.F., Quah, A.C.K., Dominick, G., Borland R., Driezen, P., Awang, R., et al. (2011). Using cognitive interviewing and behavioral coding to determine measurement equivalence across linguistic and cultural groups: An example from the International Tobacco Control Policy Evaluation Project. *Field Methods*, 23, 439-460.
- Vandenberg, R.J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- van Deth, J.W. (Eds.) (2013[1998]). *Comparative politics: The problem of equivalence*. ECPR Classics, Colchester: ECPR Press.
- van de Vijver, F.J.R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks: Sage.