

Das Vergleichen von Vergleichen als Validierungsstrategie: ein Vorschlag für die Sekundärdatenanalyse von Umfragedaten

Beckers, Tilo

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Beckers, T. (2010). Das Vergleichen von Vergleichen als Validierungsstrategie: ein Vorschlag für die Sekundärdatenanalyse von Umfragedaten. *Sozialwissenschaftlicher Fachinformationsdienst soFid*, Methoden und Instrumente der Sozialwissenschaften 2010/1, 19-33. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205051>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Das Vergleichen von Vergleichen als Validierungsstrategie Ein Vorschlag für die Sekundärdatenanalyse von Umfragedaten

Tilo Beckers

„Sociology and political science – when choosing the format of the natural sciences – intend to be generalizing sciences; and yet the time-space coordinates within which empirically found ‘laws’ are valid remain (actually or potentially) a matter of perennial dispute. In lieu of the experiment, observation of the same factors in different contexts is the obvious strategy, either to confirm empirically the generality of the statements or to specify the time-space coordinates of their applicability.” (Scheuch 1993/1968: 129)

1 Einleitung

In den empirischen Sozialwissenschaften, die versuchen, allgemeine Gesetzmäßigkeiten zu erforschen und zu überprüfen, stellt sich ebenso wie in den Naturwissenschaften die Frage, für welche Raum-Zeit-Koordinaten die Aussagen Gültigkeit beanspruchen können. Anstelle des Experiments wird häufig auf die Beobachtung derselben Faktoren in unterschiedlichen Kontexten zurückgegriffen. Dieses Verfahren ist beim internationalen Vergleich auf der Basis von Umfragedaten gegeben, indem die Mittelwerte, Regressionseffekte u.a. Parameter für verschiedene Länderstichproben miteinander verglichen werden. Traditionell begnügen sich die Vergleiche mit folgenden Strategien: (a) aus verschiedenen unabhängigen Umfragen in verschiedenen Ländern werden Mittelwerte u.a. Parameter miteinander verglichen, (b) eine länderübergreifende Datenquelle wie das *Eurobarometer* oder die *European Values Study* (EVS) dient als Quelle für eine einheitliche Messung derselben Sachverhalte. In beiden Fällen wird auf die Reliabilität der Messung vertraut. Genauere Prüfungen finden in weiterführenden Analyseverfahren mit Datenquellen statt, die die Überprüfung von Methodeneffekten etc. mit Hilfe von Strukturgleichungsmodellen (SEM) erlauben, z.B. mit den Daten des *European Social Survey* (ESS; vgl. Billiet 2003). Solche Modelle lassen sich aber nur berechnen, wenn für die interessierenden Variablen mehrere Indikatoren verfügbar sind. Dies ist insbesondere für Untersuchungen, die weniger populäre oder in Umfragereihen noch nicht hinreichend implementierte Forschungsfragen betreffen, nicht immer gegeben. Für diese Themen ist dann eine Kontrolle der Messung mit Hilfe von SEM nicht möglich. Ebenso scheiden bei bestimmten Forschungsthemen mangels vorliegender Studien Metaanalysen (Weiß 2008) als Instrumentarium aus. Für diese Fälle, insbesondere ländervergleichender Studien wird hier ein Vorschlag unterbreitet, wie dennoch die Reliabilität der Messung mit einfachen statistischen Mitteln verglichen werden kann.

Dieser Beitrag behandelt Mittelwertvergleiche und Mehrebenenanalysen, die mit unterschiedlichen Datensätzen und Messinstrumenten im Ländervergleich durchgeführt werden, um die studienübergreifende Stabilität von Stichprobenparametern und insbesondere Effekten in hierarchisch-linearen Modellen zu überprüfen. Als Anwendungsbeispiele dienen die Einstellungen zur Homosexualität in 22 europäischen Ländern, die u.a. im World Values Survey (WVS) und der European Values Study (EVS) untersucht wurden (u.a. 1999-2003). Diese Studienreihen soll hier einheitlich als Wertestu-

dien bezeichnet werden. Zudem sind Einstellungen zu gleichgeschlechtlicher Sexualität im International Social Survey Programme zur Religion untersucht worden (u.a. ISSP 1998).¹

In diesem Beitrag geht es um drei Arten von Fragen. An allererster Stelle geht es um eine statistische Frage: Führen internationale Vergleiche mit unterschiedliche Umfragereihen zu äquivalenten statistischen Ergebnissen in Mittelwertvergleichen und Mehrebenenanalysen? Ergänzend wird gefragt, ob die Zahl der Kontexte sowie die Unterschiede der Messung des Konzepts der interessierenden abhängigen Variable die Richtung und Größe der Koeffizienten beeinflussen. Zudem geht es um eine methodische Frage: Hängen die beobachtbaren Länderunterschiede mit einem abweichenden Verständnis von Homosexualität zusammen, das auf zwei unterschiedliche Messungen der abhängigen Variable zurückführbar ist? Schließlich geht es entlang des Anwendungsbeispiels um eine substantielle Frage: Unter welchen sozialen Bedingungen ist Homosexualität aus der Sicht der Menschen in Europa eigentlich akzeptabel bzw. „zu rechtfertigen“? Die zentrale Hypothese lautet, dass die Länderkontexte jenseits der Einflüsse von Individualvariablen von zentraler Bedeutung sind. Die weiteren Hypothesen und Aspekte der inhaltlichen Frage sollen in diesem methodischen Beitrag aber nicht näher erläutert werden (vgl. ausführlich Beckers 2008a). Im Abschnitt zwei werden die beiden Messinstrumente vorgestellt, in Abschnitt 3 werden die Verteilungen und Mittelwerte der beiden Variablen für die Messreihen Wertstudien 1999 und ISSP 1998 miteinander verglichen. In Abschnitt 4 werden die Mehrebenenanalysen als Fallbeispiel des Vergleichens von Vergleichen als Validierungsstrategie präsentiert. In Abschnitt 5 werden offene Fragen diskutiert und es wird vorgeschlagen, das Verfahren als Standard für die Sekundäranalyse von Umfragedaten zu etablieren.

2 Zwei Messungen der Akzeptanz der Homosexualität in zwei Umfragereihen

Die Akzeptanz der Homosexualität ist bislang relativ selten Gegenstand international vergleichender Studien gewesen (Kelley 2001; Beckers 2008a, 2008b; Andersen/Fetner 2008). In allen Studien werden starke Länderunterschiede in der Akzeptanz vorgefunden. In „Homosexualität und Humanentwicklung“ (Beckers 2008a) werden die Umfragereihen Wertstudien und ISSP miteinander verglichen. Diese Ergebnisse werden hier kurz mit einem Schwerpunkt auf dem methodischen Aspekt des Vergleichs dargestellt und weiter systematisiert.

1 Ein aktuellerer Vergleich der beiden Studienreihen ist zum derzeitigen Zeitpunkt aufgrund der noch nicht publizierten Daten des ISSP 2008 noch nicht möglich. Das hier vorgestellte Verfahren ist im Rahmen eines Promotionsprojektes an der Universität zu Köln erprobt worden: Tilo Beckers (2008): Homosexualität und Humanentwicklung: Genese, Struktur und Wandel der Ablehnung und Akzeptanz gleichgeschlechtlicher Sexualkontakte in einer international vergleichenden Kontextanalyse Untersuchung. Inauguraldissertation. Universität zu Köln. Das Verfahren wird anhand eines anderen Anwendungsbeispiels auch an anderer Stelle vorgestellt: Tilo Beckers (2010): Comparing Comparisons as a Cross-Validation Strategy (CC-CV): A Proposal for Applied Survey Research Using the Example of Attitudes towards Economic Liberalism, erscheint im Sommer 2010 in: Tilo Beckers, Klaus Birkelbach, Jörg Hagenah, Ulrich Rosar (Hg.): Komparative empirische Sozialforschung. Wiesbaden: VS Verlag für Sozialwissenschaften. Die hier vorgestellte Studie basiert auf Konferenzpapieren, die bei der 2. Konferenz der European Survey Research Association 2007 in Prag und beim Kongress der Deutschen Gesellschaft für Soziologie 2008 in Jena vorgestellt wurden.

Die Fragen der beiden Umfragereihen werden hier im Englischen als der Sprache des Quellfragbogens dokumentiert. Die (mindestens problematischen) deutschen Übersetzungen der Antwortskalen sind jeweils in Klammern angegeben. Diese Übersetzungsprobleme sollen an dieser Stelle jedoch nicht weiter thematisiert werden. Der Fragestimulus ist in beiden Studienreihen symmetrisch balanciert formuliert.

Frage zur Akzeptanz der Homosexualität in den Wertstudien (*jhomosex*):

- **„Please tell me for each of the following statements whether you think it can always be justified, never be justified, or something in between, using this card: ‘accepting a bribe’, ‘homosexuality’, (...)**
- **The card indicates a list of numbers from one to ten with the poles being verbally described:**
 - **1 ‘Never justifiable’** [*deutsch*: „niemals in Ordnung“];
 - **2; 3; 4; 5; 6; 7; 8; 9;**
 - **10 ‘Always justifiable’**“ [„immer in Ordnung“]

Diese Messung der Akzeptanz der Homosexualität (*jhomosex*) ist von sehr allgemeiner Art. Die Personen, die in homosexuelle Handlungen involviert sind, werden nicht benannt, sondern das Verhalten an sich soll bewertet werden. Auch die Geschlechter werden dadurch nicht genannt. Für diese Untersuchung soll angenommen werden, dass die Befragten sich bei der Beantwortung auf sexuelle Kontakte oder Beziehungen zwischen zwei Personen gleichen biologischen Geschlechts beziehen. Es bleibt aber offen, inwiefern die Befragten sich durch den Begriff der Homosexualität möglicherweise auch auf die spezielle Variante egalitärer homosexueller Beziehungen und damit implizit auf diese Identität oder ein anderes kulturell etabliertes homosexuelles Identitätsmodell beziehen. Die bisherige Forschung legt nahe, dass die Befragten beim Begriff Homosexualität häufiger schwule Männer als lesbische Frauen assoziieren (Ortner 1981; vgl. auch Brettell 2005). Dies ist sicherlich nicht die Absicht der Verantwortlichen der Wertestudien gewesen.

Frage zur Akzeptanz der Homosexualität im ISSP (*samesex*):

- **„Do you think it is wrong or not wrong if ...”** [Die Liste beginnt mit zwei Fragen zu außerehelichem Geschlechtsverkehr bzw. Sexualkontakt und Untreue]
- **„And what about sexual relations between two adults of the same sex, is it ...**
 - **1. Always wrong**, [*deutsch*: „immer schlimm“]
 - **2. Almost always wrong**, [„fast immer schlimm“]
 - **3. Wrong only sometimes**, [„nur manchmal schlimm“]
 - **4. Not wrong at all** [„nie schlimm“]

Die Messung der Akzeptanz der Homosexualität im ISSP (1991, 1994 und 1998) ist in bestimmter Hinsicht präziser als die Messung in den Wertestudien, da sie *erstens* den Begriff der Homosexualität vermeidet und einen rein verhaltensbezogenen Einstellungsindikator darstellt. *Zweitens* bezieht sich die Frage explizit auf männliche wie weibliche gleichgeschlechtliche Sexualkontakte, ohne jedoch etwa die Begriffe ‚gay‘ oder ‚lesbian‘ zu verwenden und damit die Dimension der Identität einzubeziehen. *Drittens* beschränkt das ISSP-Item die Zahl der Sexualpartner auf zwei und schließt damit z.B. Dreierkonstellationen u.ä. aus. *Fünftens* und *sechstens* werden die Sexualkontakte auf zwei einverständige Personen eingeschränkt, die zudem erwachsen sind. Damit wird bewusst erzwungene Sexualität oder solche in Abhängigkeitsverhältnissen ebenso ausgeschlossen wie Sexualität zwi-

schen Erwachsenen und minderjährigen Heranwachsenden oder gar Kindern. Hiermit wird die Bewertung homosexuellen Verhaltens eingeschränkt auf Verhaltensweisen, die potenziell negative Konnotationen (Gruppensex, Prostitution und Zwang, Pädophilie) definitorisch ausschließen und theoretisch Teil der Bewertung der Homosexualität in den Wertestudien sein können.

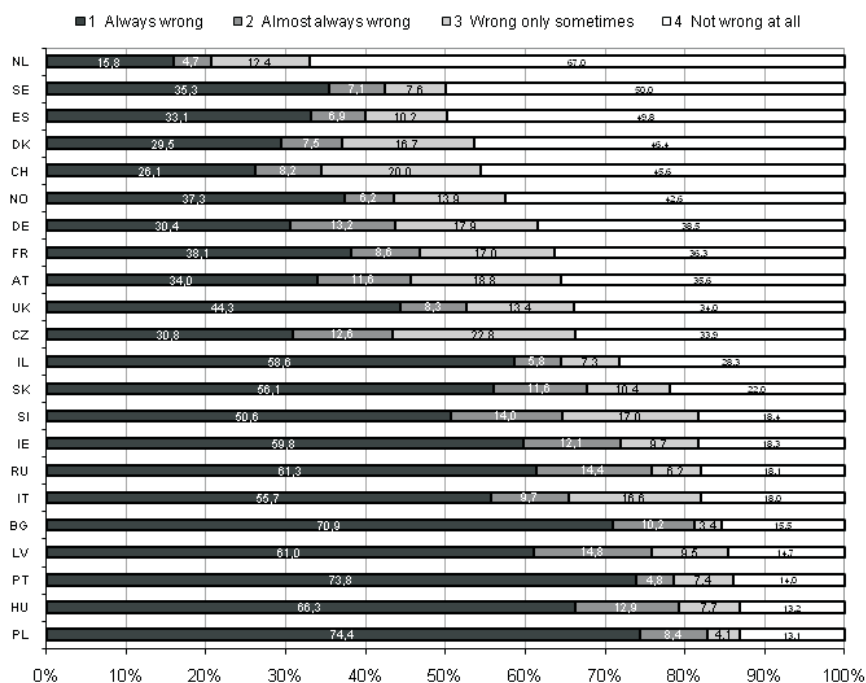
Auf der einen Seite können diese definitorischen Einschränkungen zu einer breiteren Akzeptanz gleichgeschlechtlicher Sexualkontakte bzw. der Homosexualität beitragen, auf der anderen Seite treten im ISSP andere Effekte auf, die umgekehrt die Bewertung negativ beeinträchtigen können. Das vorangehende Item behandelt Untreue und trägt so durch das damit potentiell aktivierte Stereotyp des promiskuen homosexuellen Mannes (zumindest für die ‚schwulen Sexualkontakte‘) zu einem möglichen Halo- bzw. Ausstrahlungseffekt bei. Die negativen moralischen Konnotationen zur sexuellen Untreue können also die Bewertung der gleichgeschlechtlichen Sexualkontakte negativ beeinträchtigen. Diese messtechnischen Probleme schränken einerseits die Vergleichbarkeit der Ergebnisse der Wertestudien und des ISSP ein, andererseits wird aber gerade dadurch ein Vergleich von zwei lexikalisch, semantisch und pragmatisch unterschiedlich formulierten Fragen zur Akzeptanz gleichgeschlechtlicher Sexualkontakte möglich. Smith verweist darauf, dass die Unterschiede in Sprache, Kultur und Sozialstruktur die international vergleichende Forschung einerseits wertvoll machen, sie andererseits aber auch mit einer Vielzahl von Problemen belasten, die die Messäquivalenz betreffen (ebd.): „When using only one measure, it is impossible to determine whether any measured differences (or even a measured non-difference) are societal or merely linguistic” (Smith 2003: 73). Gerade weil in jeder einzelnen hier verfügbaren Studie jeweils nur ein Item die Akzeptanz der Homosexualität misst, ist der Studienvergleich wertvoll, da er eine externe Validierung der Befunde erlaubt (vgl. Beckers 2008a).

3 Mittelwerte der Akzeptanz der Homosexualität in zwei Umfragereihen

Mit den Daten und Fragen der Wertestudien und des ISSP kann ein direkter Vergleich zweier unterschiedlicher Einstellungsmessungen vorgenommen werden, der sich auf repräsentative Stichproben aus denselben Ländern zu beinahe identischen Zeitpunkten bezieht. Damit kann u.a. die Frage geklärt werden, ob die Nennung des Begriffs Homosexualität möglicherweise zu einer Überschätzung von Länderunterschieden führt. Kommt es also bei Vermeidung dieses Begriffs durch die Umschreibung als „Sexualkontakte zwischen zwei Menschen gleichen Geschlechts“ zu geringeren Länderunterschieden? Leider erlauben die Daten wegen der eingeschränkten interkulturellen Variabilität kein vollständiges Systemdifferenzdesign mit rund 60 Ländern wie im World Values Survey 1999-2003 (vgl. Beckers 2008a, 2008b), dass etwa auch Differenzen zwischen religiösen Traditionen wie Christentum und Islam diskutiert (Beckers 2009). Um den hier vorgenommenen Vergleich von zwei Vergleichen im Sinne eines Systemähnlichkeitsdesign (Przeworski/Teune 1970) auf Europa im weiteren Sinne einzuschränken, wurden Australien, Kanada, Chile, Japan, Neuseeland, die Philippinen und die USA hier nicht berücksichtigt, für die ebenfalls Daten in den Wertestudien und dem ISSP vorliegen. Berücksichtigt wurden Israel (IL), Russland (RU) und 20 europäische Staaten: Bulgarien (BG), Dänemark (DK), Deutschland (DE), Frankreich (FR), Großbritannien (UK), Irland (IE), Italien (IT), Lettland (LV), die Niederlande (NL), Norwegen (NO), Österreich (AT), Polen (PL), Portugal (PT), Schweden (SE), die Schweiz (CH), die Slowakei (SK), Slowenien (SI), Spanien (ES), Ungarn (HU) und die Tschechische Republik (CZ). Der Vergleich wird daher also 22 identische Staaten be-

schränkt, die in beiden Studienreihen vorkommen.² Durch diesen konservativen Vergleich sollen mögliche Selektionseffekte kleiner Länderstichproben vermieden werden.

Ein bereits für die Daten der Wertestudien an andere Stelle (Beckers 2008a) nachgewiesenes Phänomen der Messung der Einstellungen zur Homosexualität bzw. gleichgeschlechtlichen Sexualkontakten zeigt sich auch für die Analyseergebnisse des ISSP im direkten Vergleich mit den Wertestudien: die starke Polarisierung der Meinungen zur Homosexualität in Befürworter und Gegner, also solche, die Akzeptanz und solche, die Ablehnung äußern. Diese Polarisierung ist zum einen auf der Personenebene, zum anderen auf der Länderebene erkennbar. Die Verteilungen der Zielvariablen same-sex im ISSP und jhomosex in den Wertestudien zeigen in Abbildung 1 eine starke Ähnlichkeit in der Polarisierung sowohl in den Ländern als auch zwischen den Ländern. Werte sind jeweils absteigend ranggeordnet nach der relativen Häufigkeit in der höchsten Akzeptanzkategorie. Die Summe der Werte 1 bis 4 an den Skalenden ist in allen 22 Ländern des ISSP größer als 65 Prozent und in 15 Ländern sogar größer als 75 Prozent. In den Wertestudien sind diese Zahlen durch die Existenz einer fünften Kategorie (bzw. 10 Ausprägungen in der Originalskalierung) etwas kleiner.



2 Daten für einige Länder der Wertestudien sind nicht aus dem Jahre 1999, sondern aus anderen Jahren und z.T. aus der dritten Welle entnommen, um die Länderüberlappung, also die Zahl der Länder für den Vergleich zu verbessern (Schweiz: 1996; Ungarn: 1998; Israel: 2001; Norwegen: 1996). Daten für einige Länder im ISSP sind faktisch im Jahre 1999 erhoben worden (Bulgarien, Schweiz, Polen, Portugal).

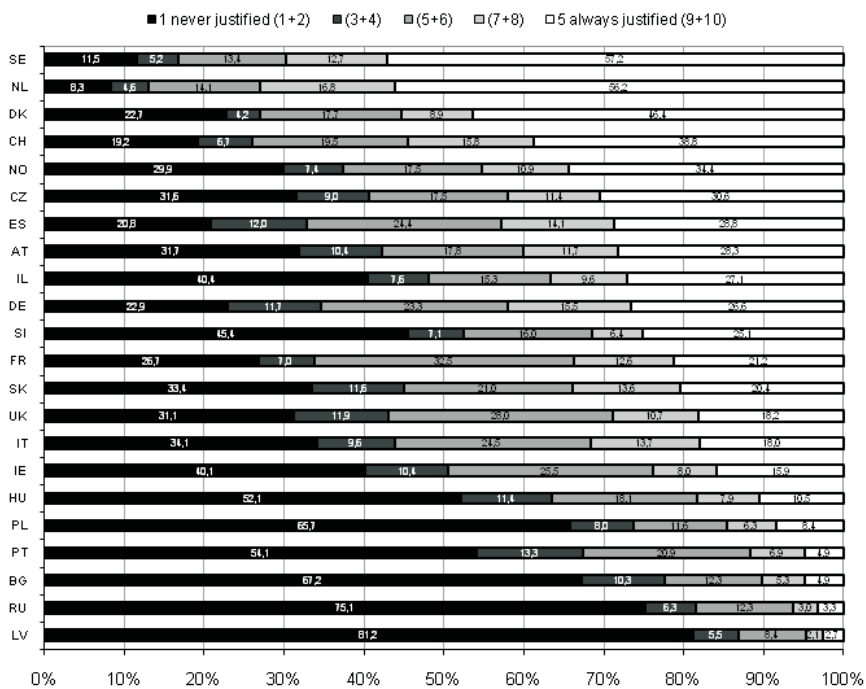


Abb. 1: Verteilungen von same-sex (ISSP 1998, oben) und homosex (Wertestudien 1999, unten) über 22 Staaten im Vergleich.

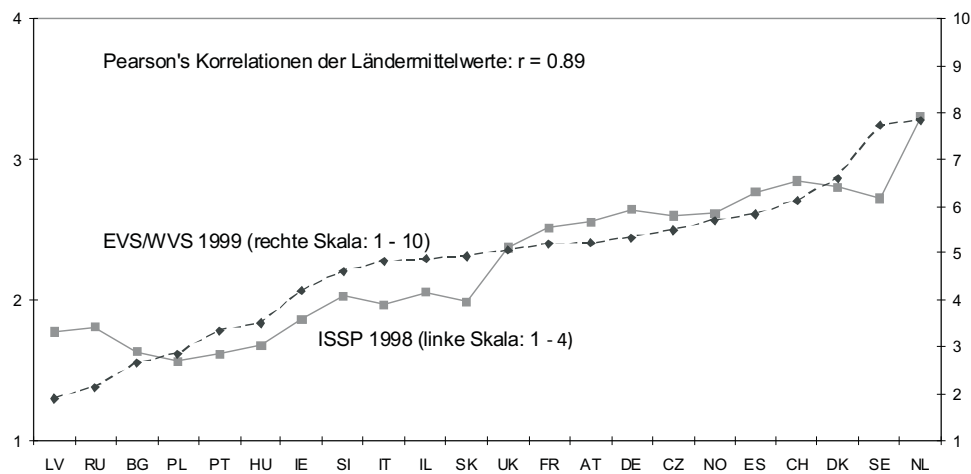


Abb. 2: Mittelwerte der Akzeptanz der Homosexualität in 22 Staaten (Wertestudien 1999/ISSP 1998)

Abbildung 2 belegt darüber hinaus die starke Ähnlichkeit der beiden Rangfolgen nach arithmetischen Mittelwerten, die eine Pearsons-Korrelation von $r = 0,89$ aufweisen. Die Abbildung zeigt überdies grafisch die für weiteren Analysen vorgenommene Lineartransformation der Skalenwerte des ISSP an die Skala der Wertestudien, die für die Tabelle 1 die Vergleichbarkeit der Mittelwerte und Standardabweichungen zwischen den Erhebungsreihen erleichtert. Der auf der Basis von Äquivalenzgewichten über alle Länder aggregierte Mittelwert ist danach praktisch identisch. Die Standardabweichung fällt in den Wertestudien etwas höher aus, die Polarisierung ist dort also stärker ausgeprägt.

Insgesamt zeigt der einfache Mittelwertvergleich zwischen den beiden Studienreihen eine hohe Ähnlichkeit der Ergebnisse unabhängig vom Messinstrument. Die Verwendung der abweichenden Frageformulierungen erzeugt hier also keine gravierenden Unterschiede, obwohl es hinsichtlich weiterer Parameter (Item-Non-Response) durchaus erhebliche Differenzen gibt, wie ergänzende Analysen zeigen.

4 Mehrebenenanalysen der Akzeptanz der Homosexualität mit zwei Umfragereihen

In der ländervergleichenden Forschung hat sich die Mehrebenenanalyse (MLA) in den vergangenen 10-15 Jahren als Standard in variablenorientierten Forschungsdesigns entwickelt. In aller Regel wird aber nur eine Studienreihe zur Überprüfung theoretischer Annahmen verwendet und im weiteren auf existierende Literatur mit abweichenden Messungen, Zielsetzungen und ggf. auch Ergebnissen verwiesen. Anstelle der Replikation, die eine ex post-Strategie der Ergebnisvalidierung darstellt, soll hier auch für die MLA das direkte ex ante Vergleichen von Vergleichen als Validierungsstrategie vorgeschlagen werden.

Tabelle 1: Vergleich der Verteilungs- und Modellparameter zwischen Wertestudien 1999 und ISSP 1998

	N	n	Mittel (min)	Mittel (max)	Mittel alle	SD alle	ANOVA: eta	HLM: ICC
jhomosex	58	54389	1.0	7.8	3.2	3.15	0.60***	0.39
jhomosex	29	27680	1.5	7.8	4.3	3.45	0.56***	0.34
jhomosex	22	20682	1.9	7.8	4.9	3.53	0.45***	0.25
samesex	22	19068	2.7	7.9	4.8	3.98	0.39***	0.15

Anmerkungen: Eta-Werte mit *** = $p < 0,001$; alle n-Werte für ungewichtete Stichproben, alle Daten wurden für die Länder als Fälle äquivalenzgewichtet mit $n=1000$; die Deskriptivmaße beziehen sich alle auf die Skala 1-10.

Tabelle 2: Äquivalenztabelle der Messinstrumente in den Wertestudien 1999 und im ISSP 1998

	ISSP 1998	EVS/WVS 1999
<i>Abhängige Variablen:</i> accepths	(Stimulus beginnt in v7: Do you think it is wrong or not wrong if ...) v9 (samesex): And what about sexual relations between two adults of the same sex, is it ... (Please tick one box only): 1. Always wrong, 2. Almost always wrong, 3. Wrong only some-times, 4. Not wrong at all.	f118 (jhomosex): Please tell me for each of the following statements whether you think it can always be justified, never be justified, or some-thing in between, using this card: (...) Homosexuality (...) – [card] 1 ‘Never justifiable’; 2; 3; 4; 5; 6; 7; 8; 9; 10 ‘Always justifiable’.
<i>Unabhängige Variablen:</i>		
cohort – Alter in Kategorien	v201: Alter rekodiert auf 7 Kategorien, 18-24, 25-34 ... 65-74, 75+	x003: Alter rekodiert auf 7 Kategorien, 18-24, 25-34 ... 65-74, 75+
atndchu – Kirchgang	v218: How often do you attend religious services?; Kirchgang in 6 Kategorien, Skala umgekehrt.	f028: rekodiert von 8 auf 6 Kategorien, und Skala umgekehrt.
godimp – Wichtigkeit von Gott	v46: life is meaningful only because god exists (5-Punkte-Skala): 1. Strongly agree, 2. Agree, 3. Neither agree nor disagree, 4. Disagree, 5. Strongly disagree	f063: korrespondierendes Item nur im EVS/WVS 1990, nicht in 1999; anstelle dessen: how important is god in your life; rekodiert von 10- auf 5-Punkte-Skala.
educat – Höchster erreichter Bildungsabschluss	v205: Höchster erreichter Bildungsabschluss in 6 Kategorien	x025: Bildung rekodiert von 8 auf 6 Kategorien
sex – Geschlecht	Geschlecht: 1. Male, 2. Female, rekodiert als Regressionsdummy zu 0: male, 1: female.	Sex: 1 ‘Male’, 2 ‘Female’, rekodiert als Regressionsdummy zu 0: male, 1: female.
gndreq – Geschlechteregalitarismus	v13: “Agree or disagree: All in all, family life suffers when the woman has a full-time job” (5-Punkte-Skala): 1. Strongly agree, 2. Agree, 3. Neither agree nor disagree, 4. Disagree, 5. Strongly disagree, Skala umgekehrt.	d056, d056a: “A working mother can establish just as warm and secure a relationship with her children as a mother who does not work”, rekodiert von 4 auf 5 categories durch Einbeziehung der N.A./D.K als Mittelkategorie, außer in Österreich und Irland (5 Punkte-Skala gegeben), alle Skalen umgekehrt.

Damit wird die Glaubwürdigkeit von Forschungsergebnissen wirksam gesteigert und neben der gängigen Replikation ein weiteres Validierungsverfahren etabliert.

Weitet man den Vergleich zusätzlich über das oben vorgestellte konservative Design auf andere Länderstichproben aus, so zeigt sich für die hier diskutierte Akzeptanz der Homosexualität das folgende Bild. Tabelle 1 zeigt, dass eine größere Länderstichprobe in unserem Beispiel die Varianz der Akzeptanz der Homosexualität zwischen Ländern stark erhöht. Werden für die Wertestudien die weiteren europäischen Länder herangezogen (N=29), so werden die *eta*-Werte in der Varianzanalyse ebenso wie die Intraklassenkorrelationskoeffizienten (*ICC*) im leeren Modell der Mehrebenenanalyse erheblich gesteigert. Die entsprechenden Vergleichswerte zwischen den Wertstudien und dem ISSP für die oben genannten 22 Länder unterscheiden sich ebenfalls. Wie auf der Basis der Frageformulierungen erwartet wurde, zeigen sich für das ISSP deutlich geringere Länderunterschiede (*eta*, *ICC*). Die Messung beeinflusst also tatsächlich die Vergleichsparameter stärker, als dies auf der Basis der Anschauung der Verteilungen und Mittelwerte zu erwarten gewesen wäre.

In die Mehrebenenanalyse fließen das Niveau der Humanentwicklung auf der Länderebene (*Human Development Indicator/HDI*; www.undp.org) und verschiedene Prädiktoren auf der Individualebene ein, die in Tabelle 2 mit ihren jeweiligen Messungen und den Quellvariablen bzw. -fragenummern aus den Originalfragebögen dokumentiert sind. Die semantischen Abweichungen können in den Modellen nicht statistisch kontrolliert werden. Allerdings stellen die Abweichungen wiederum wie schon bei der abhängigen Variable einen interessanten Testfall dar für die Beeinflussung der Vergleichsergebnisse. Für das Vergleichen von Vergleichen in den Mehrebenenanalysen wird zudem die ISSP-Skala rekodiert auf die Werte 1, 4, 7, 10, um die Interpretation der unstandardisierten Effektstärken zu erleichtern.

Die Mehrebenenanalyse als spezieller Fall der so genannten *Random Effects*-Modelle erlaubt die Spezifikation von festen bzw. fixen und freien bzw. Zufallseffekten (vgl. Snijders 1999 ; Hox 2002). Wie in allen Zufallseffektmodellen können verschiedene Effekte variiert werden: *erstens* die Effekte der Kontextvariablen auf die abhängige Variable der Individualebene, also unterschiedliche Niveaus der abhängigen Variablen in Abhängigkeit vom Kontext (*random intercept* bzw. Zufallskonstante); *zweitens* die Effekte der unabhängigen Variablen der Individualebene, also unterschiedliche Steigungsparameter in Abhängigkeit vom Kontext (*random slope* bzw. Zufallssteigungen). Sofern angenommen wird, dass der Effekt der Individualvariablen in Abhängigkeit eines bestimmten Kontextmerkmals variiert, können diese Einflussbeziehungen als *cross-level interactions* bzw. *ebenenübergreifende* Interaktionen statistisch modelliert werden.

Alle Stichproben in den dargestellten Analysen wurden auf $n=1.000$ äquivalenzgewichtet, alle HLM werden mit *Restricted Maximum Likelihood*-Schätzungen gerechnet, wobei die *Chi Quadrat*-Vergleiche auf *Full Maximum Likelihood* basieren. Die Individualmodelle werden hier als *Fixed Effect only*-Modelle geschätzt, die vollständigen HLM beinhalten auch *Random Effects*. Dabei werden die random slopes geprüft und Cross-Level-Interaktionen hinzugefügt.

Tabelle 3: Hierarchisch-lineare Zwei-Ebenenmodelle der Wertestudien (*homosex*) und des ISSP (*samesex*): Leere Modelle und Individualmodelle

	Verteilungsmasse und Vorhersage		Leeres Modell		Kohorte und Religiosität		+ Bildung		+ Geschlechtergleichheits			
	WS m	ISSP m	WS sd	ISSP sd	WS 0	ISSP 0	WS 1.1	ISSP 1.1	WS 1.2	ISSP 1.2	WS 1.3	ISSP 1.3
Intercept	4,81	4,92	3,48	4,01	4,78 (0,39)	4,63 (0,32)	4,72 (0,38)	4,81 (0,27)	4,73 (0,38)	4,82 (0,27)	4,59 (0,39)	4,34 (0,23)
Individualeffekte												
<i>Kohorte</i>												
cohort(7)	3,64	3,65	1,74	1,75	-	-	-0,44 (0,03)	-0,49 (0,05)	-0,34 (0,02)	-0,42 (0,04)	-0,33 (0,02)	-0,36 (0,04)
<i>Religiosität</i>												
atrichu(7)	3,09	2,93	1,98	1,84	-	-	-0,19 (0,03)	0,29 (0,04)	-0,20 (0,03)	-0,31 (0,04)	-0,20 (0,03)	-0,32 (0,04)
godimp(5)	3,24	2,49	1,48	1,31	-	-	-0,19 (0,03)	-0,53 (0,06)	-0,14 (0,03)	-0,48 (0,06)	-0,19 (0,03)	-0,44 (0,05)
<i>Bildung</i>												
educat(6)	3,42	3,56	1,38	1,36	+	+			0,51 (0,05)	0,35 (0,04)	0,52 (0,05)	0,31 (0,04)
<i>Geschlechtergleichheits</i>												
sex(2)	0,54	0,54			+	+					0,89 (0,11)	0,91 (0,10)
gendrequ(5)	3,77	2,84	1,20	1,32	+	+					0,16 (0,05)	0,42 (0,04)
Varianzkomponenten												
τ_0^2 (Interceptvar., Level 2)					2,89	2,09	2,72	1,48	2,75	1,49	2,69	1,23
σ_1^2 (Residualvar., Level 1)					8,87	12,81	7,82	10,91	7,44	10,80	7,12	10,35
ICC in %					24,59	14,05	25,81	11,94	26,99	12,12	27,42	10,62
Modellgüte												
D devianz					144238,81	137110,15	131018,56	105235,30	128828,48	103479,90	113392,26	100789,91
R^2_{Person} (Level 1) $R^2_{Person/Docter}$							0,10	0,17	0,13	0,18	0,17	0,22
n_{Person} (Level 2)					22	22	21	20	21	20	19	20
n_{Person} (Level 1) ungeachtet					28713	25429	26753	20112	26572	19814	23599	19454

Anmerkung: Die Tabelle weist robuste unstandardisierte Koeffizienten für *restricted maximum likelihood* (RML) Schätzungen mit robusten Standardfehlern aus, die in Klammern hinter den Effektkoeffizienten angegeben sind. Die Daten sind äquivalenzgewichtet nach Ländern mit $n = 1000$. Die Gewichte berücksichtigen disproportionale Stichproben in Deutschland und Großbritannien. Alle unabhängigen Variablen sind am *grand mean* zentriert worden (Ausnahme: unzentrierte Dummyvariablen). Alle Berechnungen sind als *two-level hierarchical linear model* mit HLM6.03 durchgeführt worden. Variablen: siehe Tabelle.

Tabelle 4: Hierarchisch-lineare Zwei-Ebenenmodelle der Wertestudien (*homosex*) und des ISSP (*same-sex*): Kontexteffekte und Cross-Level-Interaktionen

	Verteilungsmasse und Vorhersagen		Random Intercept, Random Slopes		Random IC, Cross-level Interaktionen		Random IC, Cross-level Interaktionen		Random IC, reduziertes Modell	
	WS m	ISSP WS sd	ISSP WS sd	ISSP WS sd	WS 2.0	ISSP 2.0	WS 2.1	ISSP 2.1	WS 2.2	ISSP 2.2
<i>Intercept</i>	4,81	4,77	3,48	3,98	4,35 (0,18)	4,27 (0,19)	4,38 (0,19)	4,28 (0,20)	4,42 (0,17)	4,28 (0,17)
<i>Random Effekte</i>										
HDI (0-10)	8,80	8,72	0,55	0,55	1,91 (0,21)	1,80 (0,20)	2,37 (0,23)	2,06 (0,22)	2,51 (0,27)	1,82 (0,23)
<i>Individualeffekte & Cross-level-Interaktionen</i>										
cohort (7)	3,64	3,65	1,74	1,75	-0,32 (0,03)	-0,35 (0,04)	-0,32 (0,02)	-0,35 (0,03)	-0,36 (0,02)	-0,44 (0,04)
atndichu(7)	3,09	2,93	1,98	1,84	-0,19 (0,03)	-0,33 (0,04)	-0,19 (0,03)	-0,32 (0,03)	-0,09	-0,14
HDI*atndichu					-0,09 (0,06)		-0,09 (0,06)			
godimp (5)	3,24	2,49	1,48	1,31	-0,19 (0,03)	-0,43 (0,04)	-0,17 (0,03)	-0,42 (0,04)	-0,06	-0,13
HDI*godimp							-0,07 (0,03)	-0,25 (0,03)		-0,39 (0,04)
educat (6)	3,42	3,56	1,38	1,36	0,52 (0,05)	0,30 (0,03)	0,53 (0,04)	0,30 (0,03)	0,17	0,10
HDI*educat										-0,15 (0,04)
sex (2)	0,54	0,54			0,83 (0,12)	0,87 (0,11)	0,83 (0,11)	0,87 (0,11)	0,87 (0,10)	0,80 (0,11)
gndrequ (5)	3,77	2,84	1,20	1,32	0,14 (0,05)	0,39 (0,04)	0,15 (0,04)	0,38 (0,03)	0,04	0,12
HDI*gndrequ							0,12 (0,04)	0,19 (0,04)		
<i>Varianzkomponenten</i>										
τ_{00} (Interceptvar., Level 2)					0,64	0,81	0,66	0,87	0,58	0,63
τ_{11} (Var. von agecat)					0,01	0,02	0,01	0,02		
τ_{22} (Var. von atndichu)					0,01	0,03	0,01	0,02		
τ_{33} (Var. von godimp)					0,01	0,03	0,01	0,02	0,04	0,05
τ_{44} (Var. von educat)					0,04	0,01	0,04	0,01	0,04	0,01
τ_{55} (Var. von sex)					0,22	0,19	0,22	0,20		
τ_{66} (Var. von gndrequ)					0,04	0,03	0,03	0,01		
σ^2 (Residualvar., Level 1)					6,91	10,03	6,91	10,03	7,33	10,37
Konditionaler ICC in %					8,48	7,47	8,72	7,98	7,33	5,73
<i>Modelleüte</i>										
Deviantz					121601,31	100323,67	121605,27	100310,48	134711,94	119717,41
R^2_{Land} (L., 2) Bsp/Randomsubs					0,78	0,61	0,77	0,58	0,80	0,70
R^2_{Person} (L., 1) Spjater/Bocher					0,36	0,27	0,36	0,27	0,33	0,26
n_{Land} (Level 2)					20	20	20	20	22	22
n_{Person} (Level 1) ungew.					25430	19454	25430	19454	27850	23091

Anmerkung: Vgl. Tabelle 3; HDI = Human Development Index des United Nations Development Programme (2001).

In dieser Untersuchung werden alle Variablen am Gesamtmittel zentriert, um Abweichungen durch die Parameterkombinationen (z.B. niedrige Bildung und hohe Religiosität) stets auf den Durchschnittsbefragten über alle Kontexteinheiten zu beziehen. Für die Dummyvariablen ist eine Zentrierung nicht erforderlich.

In den Tabellen 3 und 4 werden die hierarchisch-linearen Zwei-Ebenenmodelle für die Akzeptanz der Homosexualität in den beiden Studienreihen über 22 Länder dokumentiert. Neben den Verteilungsmaßen sind auch die theoretischen Vorhersagen angegeben, die sämtlich für beide Datenquellen bestätigt werden. Für die leeren Modelle zeigen sich in Tabelle 3 die bereits diskutierten Unterschiede mit geringeren Kontextvarianzen für das ISSP. Die schrittweise aufgebauten Vergleichsmodelle für die Individualebene zeigen Unterschiede in den Effektstärken.

So ist der negative Effekt der Religiosität und der positive Effekt des Geschlechteregalitarismus auf die Akzeptanz der Homosexualität im ISSP jeweils deutlich stärker. Umgekehrt ist der positive Effekt der Bildung in den Wertestudien deutlich höher. Sehr prononcierte Unterschiede ergeben sich auch für die konditionalen ICC, die im ISSP fast nur ein Drittel der Stärke der Wertestudien erreichen. In den Modellen in Tabelle 4 nähern sich die konditionalen ICCs bei Hinzunahme von Kontextmerkmal und Cross-Level-Interaktionen wieder stärker an. Die positiven Effekte des Niveaus der Humanentwicklung bzw. des sozio-ökonomischen Wohlstands (HDI) auf die Akzeptanz fällt ebenfalls in den Wertestudien deutlich stärker aus und ist zudem abhängig von dem auf der Individualebene realisierten Modell.

Die Kontextvarianz, die erklärten Varianzen auf Ebene 1 und 2, die Bedeutung des Lands als Kontext und die Kontextvariablen werden also je nach verwendetem Modell und analysierter Datenreihe unterschiedlich eingeschätzt. Dennoch ist der Vergleich zufriedenstellend, da die Effektrichtungen identisch vorhergesagt werden und die Effektstärken sich zwar unterscheiden, aber die Gesamtbeurteilung ähnlich ausfallen würde.

5 Schlussfolgerungen und ein Vorschlag für die Sekundäranalyse von Umfragedaten

Die Resultate dieser Untersuchung untermauern substantiell die Bedeutung des Länderkontexts neben Individualeinflüssen auf die Akzeptanz der Homosexualität, zeigen aber insbesondere im Vergleich mit erweiterten Modellen mit mehr Ländern (Beckers 2008a, 2009), dass nicht nur die Kontextvarianz, sondern damit einhergehend auch die Bedeutung der Kontextvariablen in Modellen mit weniger Ländern beträchtlich schwinden kann. Ländervergleichende Mehrebenenanalysen sollten aus methodischer Sicht mit Blick auf die möglichen substantiellen Schlussfolgerungen also mit Bedacht und Vorsicht analysiert und interpretiert werden. Insbesondere für etablierte Forschungsthemen werden Vergleiche mit verschiedenen Umfragen und ggf. abweichenden Messinstrumenten nahegelegt. Das Vergleichen von Vergleichen als Validierungsstrategie führt vor allen Dingen zum Schluss, dass ungeachtet der Ähnlichkeiten in Mittelwertvergleichen in den Mehrebenenanalysen zwar Effektrichtungen, nicht aber Effektstärken eindeutig interpretierbar sind. Selbst wenn die Unterschiede auf die Abweichungen in der Messung der unabhängigen Individualvariablen zurückgeführt werden können, bedeutet dies, dass die Verfügbarkeit eines bestimmten Indikators in einer bestimmten Umfragereihe in erheblichem Maße dessen relative Effektstärke in einem Modell bestimmen kann und nicht unbedingt das latente theoretische Konstrukt per se von starker Bedeutung ist.

Neben den hier vorgestellten Schritten ist ergänzend auch ein Vergleich von Korrelationen oder anderen Regressionskoeffizienten, z.B. aus OLS- oder Logitmodellen vorstellbar. Ebenso kann das Verfahren durch die Angabe von Konfidenzintervallen der Parameter und statistische Signifikanztests ergänzt werden.

In jedem Falle erhöht das Verfahren die Glaubwürdigkeit von Untersuchungsergebnissen aus der stets mit Unschärfen arbeitenden Umfrageforschung. Ein weiteres wichtiges Argument ist, dass Forscher B nur selten bereit ist, eine Replikation einer Studie von Forscher A ohne leichte Modellveränderungen durchzuführen. Zudem sind Replikationen rar und bestätigen nur Resultate innerhalb einer Studienreihe. Das Vergleichen von Vergleichen hingegen schafft Klarheit ex ante und verringert damit ein für die Umfrageforschung bedeutendes Problem: die relative Unsicherheit von Forschungsergebnissen in Fällen, in denen nur ein Indikator pro Umfragereihe zur Messung verfügbar ist und SEM nicht anwendbar sind. Der Forscher wird durch den Vergleich zwischen Studienreihen auch sensibilisiert für mögliche Alternativen der Operationalisierung, für die relative Arbitrarität seiner bisherigen Indikatorenauswahl und (bei Durchführung des vorgestellten Verfahrens) für die möglichen Abweichungen in den Parametern seiner Modelle bei der Sekundäranalyse von Umfragedaten. Dieser Mehraufwand sollte die Sache wert sein, trägt er doch mit einfachen Mitteln zu einer Verbesserung der Forschungsqualität und -kreditabilität bei.

Literatur

- Andersen, Robert/Tina Fetner 2008: Economic Inequality and Intolerance: Attitudes toward Homosexuality in 35 Democracies. In: *American Journal of Political Science* 52: 942-958.
- Beckers, Tilo 2008a: Homosexualität und Humanentwicklung. Genese, Struktur und Wandel der Ablehnung und Akzeptanz gleichgeschlechtlicher Sexualkontakte in einer international vergleichenden Kontextanalyse. Inauguraldissertation zur Erlangung des Doktorgrades der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln.
- Beckers, Tilo 2008b: Das Recht und der homosexuelle Körper in einem weltweiten Vergleich der Zusammenhänge von Rechtsordnung, sozialen Chancen und Einstellungen, in: Karl-Siegbert Rehberg (Hg.): *Die Natur der Gesellschaft. Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel 2006*. S. 1097-1113.
- Beckers, Tilo 2009: Islam and the Acceptance of Homosexuality: the Shortage of Socio-Economic Well-being and Responsive Democracy, in: Samar Habib (ed.): *Homosexuality and Islam* (2 Bände). Praeger.
- Beckers, Tilo 2010: Comparing Comparisons as a Cross-Validation Strategy (CC-CV): A Proposal for Applied Survey Research Using the Example of Attitudes towards Economic Liberalism, erscheint im Sommer 2010 in: Tilo Beckers, Klaus Birkelbach, Jörg Hagenah, Ulrich Rosar (Hg.): *Komparative empirische Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Billiet, Jaak 2003: Cross-Cultural Survey Equivalence with Structural Equation Modelling, in: Janet A. Harkness/Fons J.R. van de Vijver/Peter Ph. Mohler: *Cross Cultural Survey Methods*. Hoboken, NJ: Wiley-Interscience. S. 247-264.
- Brettell, Caroline B./Carolyn F. Sargent (Hg.) 2005: *Gender in Cross-Cultural Perspective*. Upper Saddle River, New Jersey: Pearson.

- Hox, Joop J. 2002: *Multilevel Analysis, Techniques and Applications*. Lawrence Erlbaum Associates.
- Kelley, Jonathan 2001: Attitudes towards Homosexuality in 29 Nations. In: *Australian Social Monitor*, Jg. 4, H. 1. Online verfügbar unter <http://melbourneinstitute.com/sm/number13.html>, zuletzt geprüft am 25.10.2006.
- Ortner, Sherry B./Harriet Whitehead 1981: *Sexual Meanings: The Cultural Construction of Gender and Sexuality*. Cambridge, London u.a.: Cambridge University Press.
- Przeworski, Adam/Henry Teune 1970: *The Logic of Comparative Social Inquiry*. New York u.a.: Wiley-Interscience
- Raudenbush, Stephen W./Anthony S. Bryk 2002: *Hierarchical Linear Models. Applications and Data Analysis Methods*. Newbury Park, CA: Sage.
- Scheuch, Erwin K. 1993 (1968): The Cross-Cultural Use of Sample Surveys: Problems of Comparability. *Historical Social Research/Historische Sozialforschung* 18/1: 104-138.
- Smith, Tom W 2003: Developing Comparable Questions in Cross-National Surveys, in: Janet A. Harkness/Fons J.R. van de Vijver/Peter Ph. Mohler: *Cross Cultural Survey Methods*. Hoboken, NJ: Wiley-Interscience. S. 69-91.
- Snijders, Tom A.B./Bosker, Roel 1999: *Multilevel Analysis*. London: Sage.
- Weiß, Bernd 2008: *Meta-Analyse als Verfahren der Forschungssynthese in der Soziologie. Dargestellt anhand zweier Fallbeispiele zum Schulabsentismus*. Inauguraldissertation zur Erlangung des Doktorgrades der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln.

Datenquellen:

- EUROPEAN AND WORLD VALUES SURVEYS FOUR-WAVE INTEGRATED DATA FILE, 1981-2004, v.20060423, 2006. The European Values Study Foundation and World Values Survey Association. Aggregate File Producers: ASEP/JDS, Madrid, Spain/Tilburg University, Tilburg, the Netherlands. Aggregate File Distributors: ASEP/JDS and ZA, Cologne, Germany. Daten als Download verfügbar unter: <http://www.worldvaluessurvey.org/>
- International Social Survey Programme: ISSP "Religion II" 1998. ZA-Nr. 3190. Verfügbar über GESIS: <http://www.gesis.org/dienstleistungen/daten/umfragedaten/issp/modules-study-overview/religion/1998/>
- United Nations Development Programme 2001: *Human Development Reports* (<http://hdr.undp.org/> bzw. <http://hdr.undp.org/statistics/data/>).

Zur Person

Tilo Beckers, Dr. rer. pol., ist Akademischer Rat für Soziologie und Methoden der empirischen Sozialforschung am Institut für Sozialwissenschaften der Heinrich-Heine-Universität Düsseldorf. Er hat 2008 an der Universität zu Köln mit einer international vergleichenden Kontext- und Mehrebenenanalyse über Einstellungen zur Homosexualität im Fach Soziologie promoviert und zuvor in Düsseldorf und New York (New School for Social Research) studiert. In seinem laufenden Habilitationsprojekt analysiert er u.a. die Einstellungen zur Sterbehilfe.

Neueste Publikationen: „Komparative empirische Sozialforschung“, hg. zus. mit Klaus Birkelbach, Jörg Hagenah und Ulrich Rosar. Wiesbaden: VS-Verlag. (2010, im Erscheinen); „Islam and the Acceptance of Homosexuality: the Shortage of Socio-Economic Well-being and Responsive Democracy“, in: Samar Habib (ed.): *Islam and Homosexuality* (in two volumes). Praeger/ABC-Clio. (2009); „Ökonomische Moral im europäischen Vergleich“, in: Jan van Deth/Silke Keil (Hg.): *Deutschlands Metamorphosen. Analysen mit dem European Social Survey*. Baden-Baden: Nomos. (erscheint voraussichtlich im Herbst 2010).