

**Evaluation der Äquivalenz  
eines gemeinsamen Satzes an Indikatoren  
in der interkulturell vergleichenden Sozialforschung**

Michael Braun

Zentrum für Umfragen, Methoden und Analysen, Mannheim

*Zusammenfassung*

*Der Beitrag beschreibt verschiedene explorative Verfahren zur Überprüfung der funktionalen Äquivalenz eines gemeinsamen Satzes von Items in der interkulturell vergleichenden Sozialforschung, und zwar sowohl im Hinblick auf interne (Struktur der Indikatoren untereinander in den verschiedenen Ländern) als auch externe Konsistenz (Beziehungen zu Kriteriumsvariablen). Dabei wird der Einsatz verschiedener statistischer Verfahren illustriert.*

*Summary*

*This paper deals with different explorative procedures for the assessment of functional equivalence of a common set of items in interculturality comparative research, both with regard to internal (structure among indicators in different countries) and external consistency (relationships with criterion variables). The use of different statistical techniques is illustrated.*

International vergleichende Umfragen haben in den letzten Jahrzehnten stark an Bedeutung gewonnen. Dies gilt einmal für die Verfügbarkeit entsprechender Datensätze als auch für die auf diesen basierenden Analysen. So sind etwa allein auf der Grundlage der ISSP-Daten bereits knapp 400 Bücher, Buchbeiträge und Zeitschriftenartikel (ohne Beiträge in Newslettern) publiziert worden, bei denen mindestens zwei Länder miteinander verglichen werden. Was sind nun die methodologischen Voraussetzungen solcher Vergleiche?

## **1 Voraussetzung der Vergleichbarkeit: Funktionale Äquivalenz**

Für komparative Studien im Sinne eines Vergleichs über verschiedene Länder und über die Zeit gilt generell, daß sinnvoll nur verglichen werden kann, wenn in den unterschiedlichen Ländern bzw. zu den unterschiedlichen Zeitpunkten 'das Gleiche' gemessen wurde (allgemein dazu Alwin et al. 1994). Dies bezieht sich zunächst nicht auf die manifesten Indikatoren, sondern auf die zugrundeliegenden Dimensionen. Unter funktionaler Äquivalenz wird in einem allgemeinen Sinne verstanden, daß die zugrundeliegenden Dimensionen oder theoretischen Konzepte in den beteiligten Ländern eine zumindest ähnliche Bedeutung haben. Sonst liegt ein 'construct bias' (van de Vijver und Leung 1997) vor: Ein Konstrukt ist in unterschiedlichen Kulturen unterschiedlich breit definiert, d.h. es gehören jeweils unterschiedliche Aspekte dazu. Probleme können beispielsweise auftreten, wenn Traditionalität von Geschlechterrollen allgemein gemessen werden soll, sich das Instrument aber nur auf Erwerbsbeteiligung bezieht. Gibt es dann in einem Land hinsichtlich der Erwerbsbeteiligung keine Unterschiede zwischen den Geschlechtern, wohl aber hinsichtlich anderer Aspekte, so kann das Instrument dort nicht sinnvoll eingesetzt werden. Selbst wenn auf der Konstruktebene keine Probleme bestehen, kann ein 'method bias' vorliegen und die Validität beeinträchtigen: Unvergleichbarkeit der Stichproben oder der Durchführung der Studien in den einzelnen Ländern; unterschiedliche Tendenzen, im Sinne der sozialen Erwünschtheit zu antworten; etc. Schließlich kann ein 'item bias' bestehen, d.h. einzelne Items werden in den einzelnen Ländern unterschiedlich interpretiert. Im folgenden wird nur der 'item-bias' behandelt. Weiterhin erfolgt eine Einschränkung auf die Verwendung eines gemeinsamen Satzes an Indikatoren, da dieser Fall in der sozialwissenschaftlichen Umfrageforschung am häufigsten vorkommen dürfte. Grundsätzlich läßt sich die hier beschriebene Vorgehensweise aber auch auf Instrumente anwenden, bei denen in den einzelnen Ländern auch unterschiedliche Indikatoren verwendet werden (van Deth 1998).

Probleme hinsichtlich der funktionalen Äquivalenz im interkulturellen Vergleich können zumindest analytisch auf zwei verschiedene Ursachen zurückgeführt werden: Unangemessene Übertragung des Master-Fragebogens bzw. einzelner Items in andere Sprachen und Unterschiede in der sozialen Realität in verschiedenen Ländern. Unterscheidet man objektive (Fragen zu Verhalten und demographischen Charakteristika der Befragten) und subjektive Variablen (Fragen zu Wahrnehmungen, Einstellungen und Werten), so ist festzuhalten, daß beide Fehlertypen beide Variablentypen betreffen können. Demographiefragen können genauso falsch übersetzt werden wie Einstellungsfragen. Klassisches Beispiel einer unvollständigen Passung von Erhebungsinstrumenten für sozio-demographische Informationen und der Realität in den betreffenden Ländern ist etwa die Bildungsmessung (Braun und Müller 1997), aber auch Familienstand, Einkommen und sogar Alter können in verschiedenen Gesellschaften eine unterschiedliche Bedeutung haben (Scheuch 1968). Im folgenden werden aber nur Meßfehler diskutiert, die durch unterschiedliche Interpretation von Einstellungsfragen zustandekommen.

Zur Evaluation der funktionalen Äquivalenz bieten sich zwei Strategien an: Überprüfung der internen und der externen Konsistenz (van Deth 1998). Hinsichtlich des Kriteriums der internen Konsistenz gelten die einzelnen Indikatoren dann als äquivalent, wenn die Struktur der Indikatoren untereinander in den verschiedenen Ländern gleich ist. Bei dem Kriterium der externen Konsistenz wird Äquivalenz über gleiche Beziehungen zu externen Kriteriumsvariablen hergestellt. Voraussetzung hierfür ist, daß diese Kriteriumsvariablen in den verschiedenen Ländern funktional äquivalent sind; sonst wird das Problem nur verschoben.

## **2 Ein Beispiel: Die Geschlechterrollen-Batterie des ISSP 1994**

Als Beispiel zur Illustration verwenden wir hier die Geschlechterrollen-Batterie der ISSP-Umfrage (International Social Survey Program, siehe Braun 1994) von 1994 zum Thema 'Familie und sich ändernde Geschlechterrollen' (Zentralarchiv 1997). Das ISSP ist ein Verbund von sozialwissenschaftlichen Instituten aus einer Vielzahl von Ländern, die jährlich eine Umfrage zu einem bestimmten Thema gemeinsam durchführen. An der Umfrage des Jahres 1994 waren insgesamt 23 Länder beteiligt, wobei insgesamt 33.590 Fälle realisiert wurden. Für die folgenden Betrachtungen werden allerdings nur West- und Ostdeutschland, Ungarn, Spanien und Kanada ausgewählt.

Die Geschlechterrollen-Batterie des ISSP basiert auf in den 60er Jahren in den USA entwickelten Fragen (Mason, Czajka und Arber 1976). Sie setzen im Prinzip voraus, daß der männliche Partner voll erwerbstätig ist und unterstellen die Erwerbstätigkeit der Frau als problematisch. In unterschiedlichen Länderkontexten (und bei Zeitvergleichen!) treffen diese Items auf unterschiedliche Realitäten (Ausmaß der Erwerbstätigkeit der Frau, Organisation außerfamiliärer Kinderbetreuung, gesellschaftliche Diskussion, etc.), so daß unterschiedliche Interpretationen der Items nicht überraschen. Neben dem ISSP verwenden auch der World Value Survey, der European Value Survey, die Eurobarometer-Umfragen, der Wohlfahrtssurvey und der ALLBUS ähnliche Items. Von den Methodenproblemen sind alle diese Großumfragen in mehr oder weniger dem gleichem Maße betroffen.

Übersicht 1 zeigt die Geschlechterrollen-Batterie des ISSP 1994 im Überblick; dabei stehen die Items der Dimension Geschlechterrollen-Ideologie im Mittelpunkt der folgenden Betrachtungen. Zu den methodologischen Problemen für internationale Vergleiche, die die Batterie insgesamt aufweist, siehe Braun (1998). Die Items wurden auf einer 5-stufigen Likert-Skala erhoben (von 'stimme voll und ganz zu' zu 'stimme überhaupt nicht zu'). Bei einigen Items ist vermerkt, daß sie für die folgenden Analysen umgepolt wurden.

### **Konsequenzen für die Familie**

Eine berufstätige Mutter kann ein genauso herzliches und vertrauensvolles Verhältnis zu ihren Kindern finden wie eine Mutter, die nicht berufstätig ist (*Verhältnis Mutter-Kind*, umgepolt).

Ein Kind, das noch nicht zur Schule geht, wird wahrscheinlich darunter leiden, wenn seine Mutter berufstätig ist (*Kind leidet*).

Alles in allem: Das Familienleben leidet darunter, wenn die Frau voll berufstätig ist (*Familie leidet*).

Das Familienleben leidet oft, weil Männer sich zu sehr auf ihre Arbeit konzentrieren (*Männer arbeiten zu viel*).

### **Geschlechterrollen-Ideologie**

Einen Beruf zu haben ist ja ganz schön, aber was die meisten Frauen wirklich wollen, ist ein Heim und Kinder (*Präferenz d Frau*).

Hausfrau zu sein ist genauso erfüllend wie gegen Bezahlung zu arbeiten (*Hausfrau erfüllend*).

Es ist nicht gut, wenn der Mann zuhause bleibt und sich um die Kinder kümmert und die Frau außer Haus berufstätig ist (*Rollentausch nicht gut*).

Die Aufgabe des Mannes ist es, Geld zu verdienen, die der Frau, sich um Haushalt und Familie zu kümmern (*Aufgabe Mann-Frau*).

### **Ökonomische Konsequenzen**

Einen Beruf zu haben ist das beste Mittel für eine Frau, um unabhängig zu sein (*Unabhängigkeit*, umgepolt).

Die meisten Frauen müssen heutzutage arbeiten, um ihre Familie zu unterstützen (*Unterstützung*, umgepolt).

Der Mann und die Frau sollten beide zum Haushaltseinkommen beitragen (*Doppelverdienst*, umgepolt).

## **3 Überprüfung der internen Konsistenz**

Zur Überprüfung der internen Konsistenz lassen sich im Prinzip alle Verfahren einsetzen, die für eine Datenanalyse in Frage kommen. In diesem Papier werden aber nur explorative Verfahren dargestellt, die es erlauben, in kurzer Zeit einen anschaulichen Überblick über mögliche Probleme zu erhalten. In den folgenden Analysen sind alle Items so recodiert, daß hohe Werte nicht-traditionelle Einstellungen anzeigen; bei der Dimension Geschlechterrollen-Ideologie gilt dabei durchgehend Ablehnung der Items als nicht-traditionelle Antwort.

### 3.1 Multidimensionale Skalierung

Mit Hilfe der Multidimensionalen Skalierung (MDS, siehe Borg 2000, Borg und Shye 1995, Borg und Groenen 1997) können Ergebnisse in einer intuitiv verständlicheren und viel anschaulicheren Weise als z.B. mit einer Faktorenanalyse präsentiert werden. MDS repräsentiert die Interkorrelationen der Items in einem multidimensionalen Raum. Die Korrelationen entsprechen den Distanzen zwischen den Items. Die Interpretation der MDS läuft über die Entsprechung der geometrischen Charakteristika der Konfiguration und den inhaltlichen Charakteristika der Items. Die Linien, die in die MDS eingetragen werden, sind theoretisch abgeleitete Trennlinien zwischen den Items.

Durchführung: Die MDS läßt sich etwa mit dem Statistik-Programmsystem STATISTIKA (StatSoft 1999) durchführen. Dazu müssen zunächst alle Items in die gleiche Richtung gepolt und eine Korrelationsmatrix berechnet werden. Die Korrelationsmatrix wird innerhalb des Moduls Basic Statistics and Tables mit Hilfe des Submoduls Correlation matrices berechnet. Die Korrelationsmatrix wird dann in das Modul Multidimensional Scaling eingelesen. Dort kann auch eine Teilmenge der in der Korrelationsmatrix berücksichtigten Variablen für die weiteren Analysen ausgewählt werden. Um eine Graphik wie in Abbildung 1 zu erhalten, ist am Schluß das Kommando Graph final configuration, 2D anzuklicken. Die Graphik läßt sich dann innerhalb des STATISTIKA-Programms weiter bearbeiten, wobei insbesondere die theoretisch bestimmten Trennungslinien eingefügt werden können.

Bei der MDS werden alle Items der Geschlechterrollen-Batterie gemeinsam betrachtet, um Auffälligkeiten einzelner Items der Dimension Geschlechterrollen-Ideologie zeigen zu können. Abbildungen 1 bis 3 zeigen die MDS-Repräsentationen der Interkorrelationen der Geschlechterrollen-Items für West- und Ostdeutschland und Kanada; hierbei sind die problematischen Items fett markiert. Die MDS-Repräsentationen für Ungarn und Spanien sind der westdeutschen strukturell äquivalent und werden daher nicht abgebildet.

Während in Westdeutschland, Ungarn und Spanien alle vier in Übersicht 1 unter 'Geschlechterrollen-Ideologie' subsumierten Items auch in der entsprechenden Region liegen, gibt es bei Ostdeutschland und Kanada jeweils ein Item, bei dem dies nicht der Fall ist. In Ostdeutschland ist das Item *Rollentausch nicht gut* kein Teil dieser Region und in Kanada fällt das Item *Hausfrau erfüllend* heraus. Diese Items werden also in den betreffenden Ländern offenbar anders interpretiert als die übrigen Items der Dimension. Zu einer theoretischen Erklärung der Besonderheiten des Items *Hausfrau erfüllend* siehe Braun (2000).

Abbildung 1: MDS für Westdeutschland

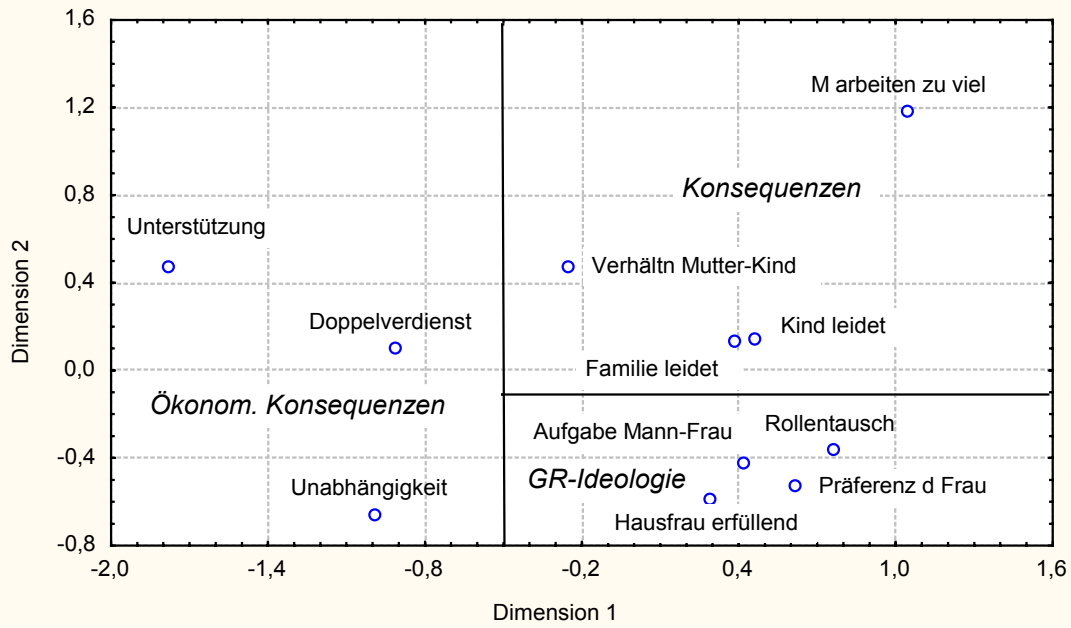
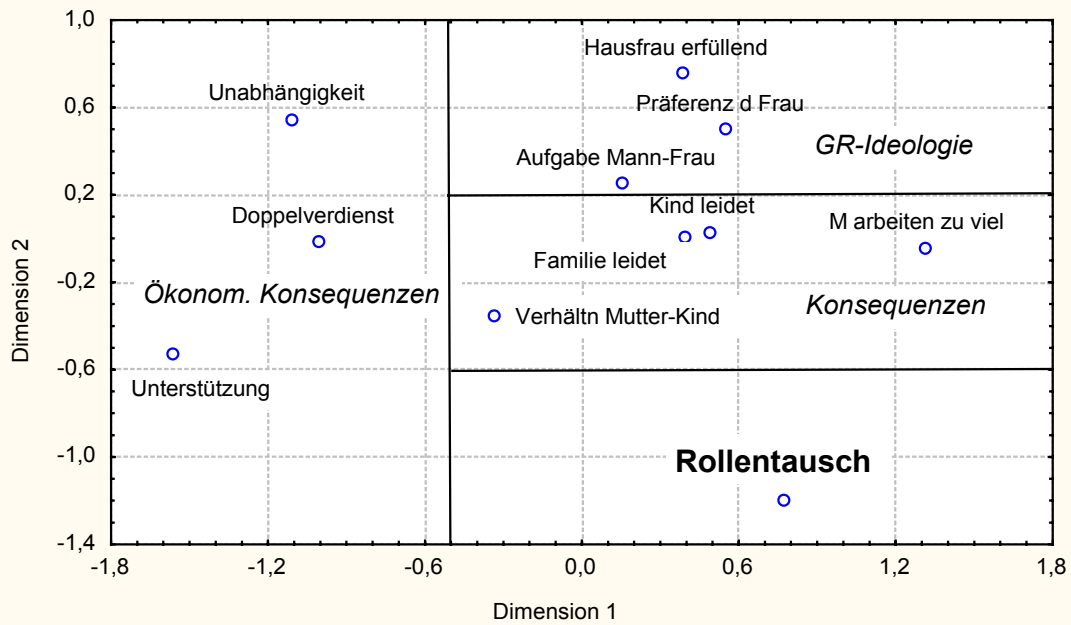
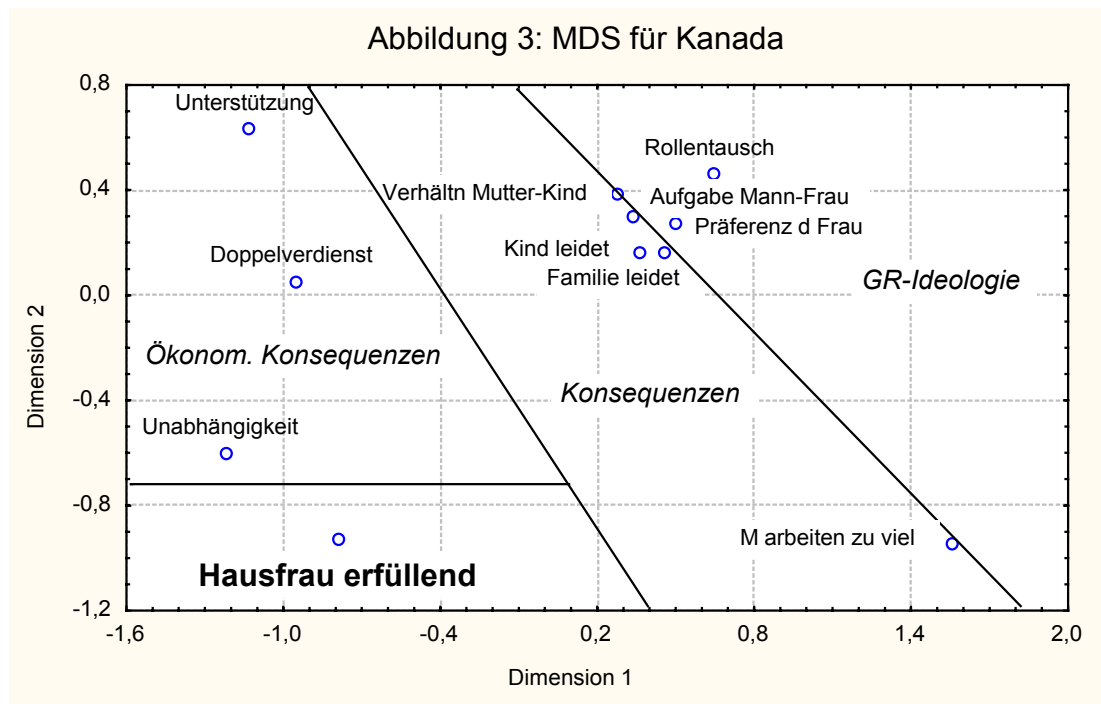


Abbildung 2: MDS für Ostdeutschland





### 3.2 Mittelwertsvergleiche

Ein ganz einfaches Verfahren zur Feststellung von Auffälligkeiten ist der Vergleich von Mittelwerten für die einzelnen Länder über die Items der Dimension. In Tabelle 1 sind die beiden problematischen Fälle jeweils fett gedruckt. Abbildung 4 zeigt die entsprechende graphische Darstellung, bei der die Pfeile diese beiden Fälle markieren.

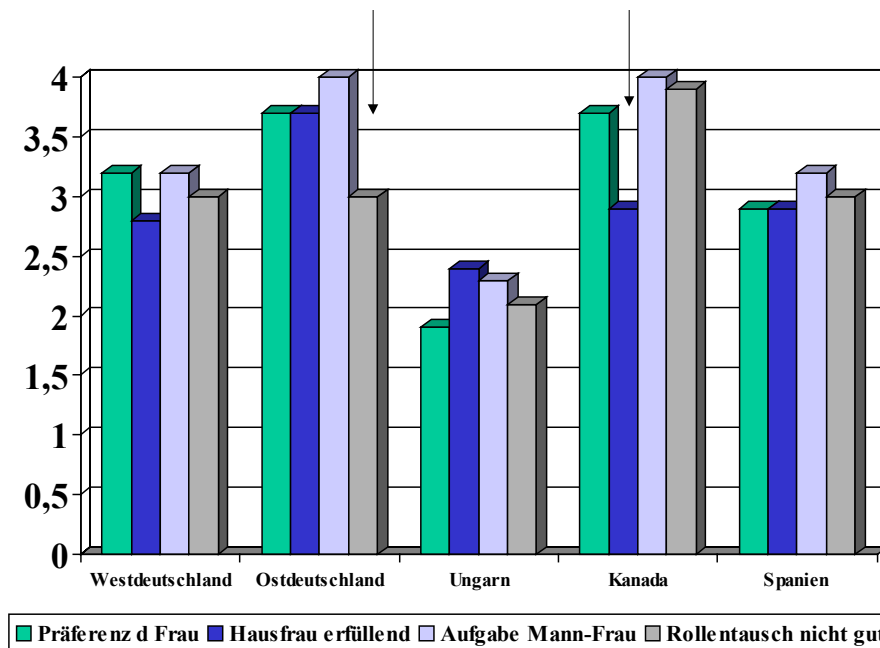
Tabelle 1: Vergleich von Mittelwerten

	<i>Präferenz d Frau</i>	<i>Hausfrau erfüllend</i>	<i>Aufgabe Mann-Frau</i>	<i>Rollentausch nicht gut</i>
Westdeutschland	3.2	2.8	3.2	3.0
Ostdeutschland	3.7	3.7	4.0	<b>3.0</b>
Ungarn	1.9	2.4	2.3	2.1
Kanada	3.7	<b>2.9</b>	4.0	3.9
Spanien	2.9	2.9	3.2	3.0

Der Mittelwert des Items *Hausfrau erfüllend* liegt in Kanada z.B. um 1.1 Skaleneinheiten unter dem für das Item *Aufgabe Mann-Frau*, in den anderen Ländern beträgt diese Differenz maximal .4. Bei dem Item *Rollentausch nicht gut* fällt vor allem der niedrige Wert in Ostdeutschland auf. Während Ostdeutschland bei allen anderen Items der Dimension Geschlechterrollen-Ideologie teilweise deutlich weniger traditionell ist als Westdeutschland, sind beide Teile Deutschlands bei diesem Item auf dem gleichen Niveau.

Die Profile für die einzelnen Länder sind also verschieden: Ostdeutschland und Kanada haben auf den betreffenden Items jeweils 'zu' niedrige Werte. Dieses Ergebnis deutet in die gleiche Richtung wie die MDS, wobei die Übereinstimmung keineswegs trivial ist: Bei der MDS werden nur die Korrelationen zwischen den Items und hier nur die Mittelwerte berücksichtigt.

Abbildung 4: Vergleich von Mittelwerten



### 3.3 Beziehung zwischen Items und der zugrundeliegenden Dimension

Die von den Befragten auf einem Item erzielten Werte werden nun in Beziehung gesetzt mit ihren Werten auf einer unterstellten latenten Dimension und der Zusammenhang zwischen beiden Variablen über verschiedene Länder miteinander verglichen. Als Grundlage der Schätzung der latenten Werte können die Werte dienen, welche die Befragten auf allen übrigen Items erzielen. Wegen der geringen Zahl der verfügbaren Items und der Tatsache, daß nicht nur ein Item problematisch ist, bietet es sich hier jedoch an, lediglich *Aufgabe Mann-Frau* als Vergleichsitem zu wählen. Dadurch wird zwar eine gewisse Beeinträchtigung der Reliabilität in Kauf genommen, aber dafür wird die Benchmark auch nicht aus offensichtlich disparaten oder ebenso problematischen Items gebildet.

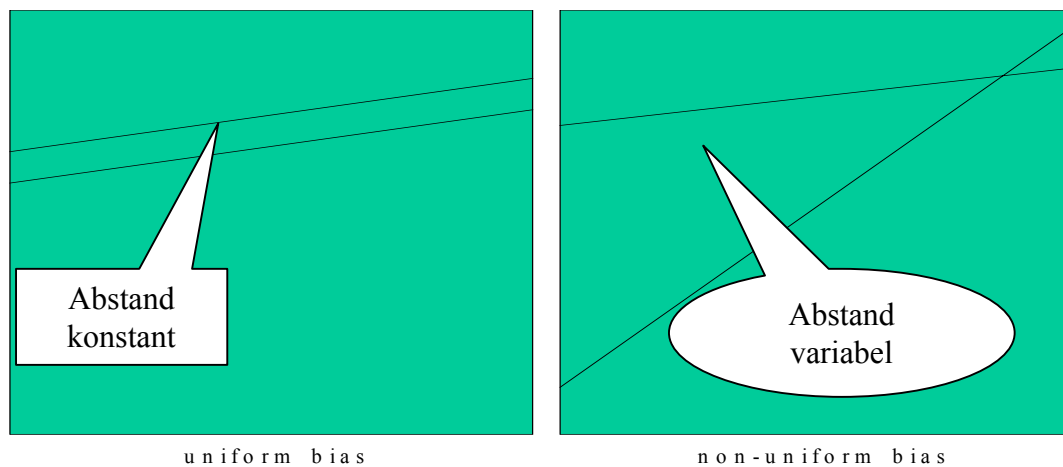
In einer ersten Annäherung läßt sich das hier bezogene Verständnis von (interkultureller) funktionaler Äquivalenz für eine Batterie inhaltlich entsprechender Items in etwas formalerer Weise wie folgt fassen: Sie wäre perfekt gegeben, wenn die Items untereinander in den einzelnen Ländern die gleichen Regressionsmerkmale aufweisen und sich die länderspezifischen Muster für die Regressionsmerkmale der einzelnen Items bei der Regression auf andere Variablen (z.B. den Indikator, der als repräsentativ für die zu messende Dimension angesehen wird) nicht unterscheiden.

Zu diskutieren sind noch die verschiedenen Fälle, in denen diese Bedingungen nicht alle erfüllt sind. Unterscheidet sich nur die Konstante für ein Item in den einzelnen Ländern bei der Regression eines Items auf den Indikator für die zugrundeliegende Dimension, d.h. die 'Schwierigkeit' des betreffenden Items in den verschiedenen Ländern ist unterschiedlich, so liegt ein 'uniform bias' vor (van de Vijver und Leung 1997). Ein 'uniform bias' beeinflusst die



Werte für den gesamten Wertebereich der zugrundeliegenden Dimension in etwa dem gleichen Ausmaß (d.h. lediglich die Identität des Ursprungs ist nicht mehr gegeben). Mittelwertvergleiche zwischen den Ländern sind dann mit den betroffenen Items nicht mehr möglich. Sind (auch) die Regressionskoeffizienten eines Items unterschiedlich, deutet dies noch stärker darauf hin, daß das betreffende Item in den einzelnen Ländern eine unterschiedliche Bedeutung hat. Der dann vorliegende ‘nonuniform bias’ beeinflusst die Werte unterschiedlich über den gesamten Wertebereich (d.h. auch die Identität der Maßeinheit ist nicht mehr gegeben). Abbildung 5 veranschaulicht das Prinzip von ‘uniform’ und ‘non-uniform’ bias.

Abbildung 5: ‘Uniform’ und ‘non-uniform’ bias



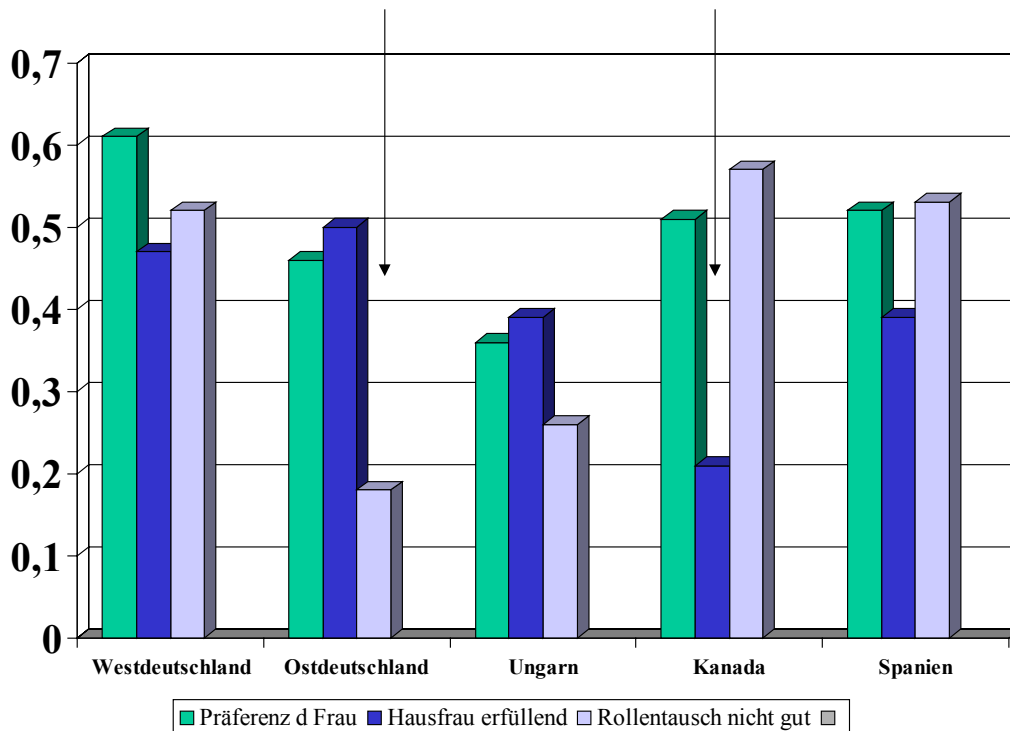
Die folgende Illustration erfolgt über bivariate Korrelationskoeffizienten und Interaktionsplots, die eine besonders anschauliche Darstellung ermöglichen.

Wenn man die Höhe der Korrelationen der drei Items mit dem Item *Aufgabe Mann-Frau* jeweils für die einzelnen Länder vergleicht, so zeigt sich für Kanada ein deutlicher Einbruch bei *Hausfrau erfüllend* und für Ostdeutschland bei *Rollentausch nicht gut* (Tabelle 2 und Abbildung 6 zur Veranschaulichung). Wenn man - wie hier - das Item *Aufgabe Mann-Frau* als Proxy für das zu messende Konzept versteht, sind diese beiden Items in den beiden Ländern jeweils keine besonders validen Indikatoren für dieses Konzept.

Tabelle 2: Korrelationen der übrigen Items mit *Aufgabe Mann-Frau*

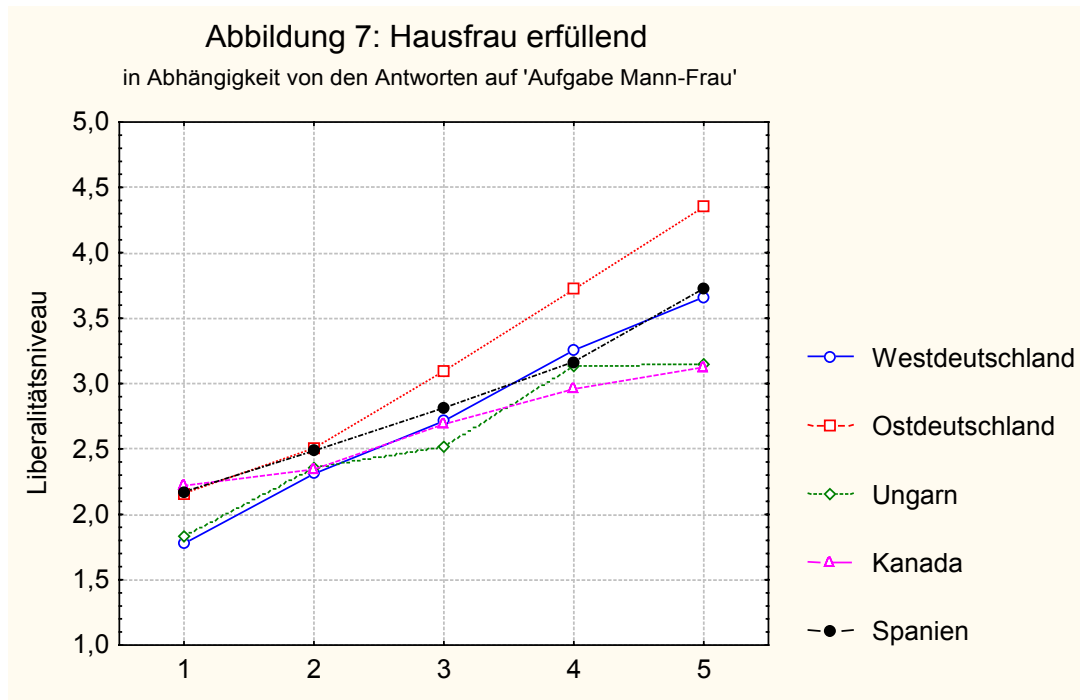
	<i>Präferenz d Frau</i>	<i>Hausfrau erfüllend</i>	<i>Rollentausch nicht gut</i>
Westdeutschland	.61	.47	.52
Ostdeutschland	.46	.50	<b>.18</b>
Ungarn	.36	.39	.26
Kanada	.51	<b>.21</b>	.57
Spanien	.52	.39	.53

Abbildung 6: Korrelationen der übrigen Items mit *Aufgabe Mann-Frau*

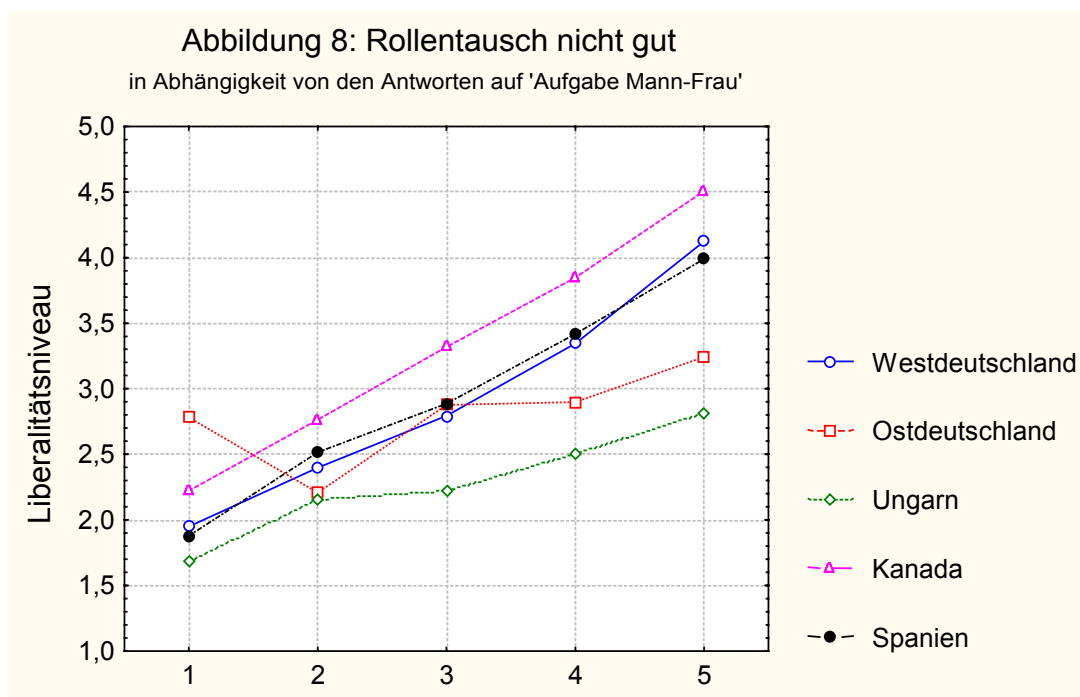


Wenn man nun die Mittelwerte des Items *Hausfrau erfüllend* für die einzelnen Skalenwerte des Items *Aufgabe Mann-Frau* abträgt (Abbildung 7), zeigt Kanada mit den höchsten Mittelwert bei den traditionellen Befragten (gemessen am Item *Aufgabe Mann-Frau*), aber den geringsten Mittelwert für die egalitär eingestellten Befragten. Beide Teile Deutschlands zeigen die größten Unterschiede zwischen den traditionell und den egalitär eingestellten Befragten, Kanada die geringsten, und Spanien sowie Ungarn liegen dazwischen. Vergleicht man Ostdeutschland und Kanada, so läßt sich folgendes sagen: Die traditionellsten Befragten erreichen in beiden Ländern den gleichen Mittelwert auf dem Item *Hausfrau erfüllend*, aber für die am wenigsten traditionellen Befragten liegt der Mittelwert in Ostdeutschland sogar um 1 1/2 Skalenpunkte über dem in Kanada. Dies ist ein beträchtlicher 'non-uniform' bias.

Durchführung: Der in Abbildung 7 dargestellte Interaktionsplot läßt sich etwa mit dem Statistik-Programmsystem STATISTIKA (StatSoft 1999) durchführen. Innerhalb des Moduls Basic Statistics and Tables wird das Submodul Breakdown & one-way ANOVA benutzt und darin die abhängigen und unabhängigen Variablen definiert. Die Graphik wird ausgegeben, wenn das Kommando Interaction plots angeklickt wird.



Hinsichtlich des Items *Rollentausch nicht gut* zeigt Abbildung 8 einen beträchtlichen ‘non-uniform bias’ im Vergleich zwischen den beiden ehemals sozialistischen Ländern auf der einen und den westlichen Ländern auf der anderen Seite. Innerhalb der westlichen Länder besteht dagegen nur ein relativ geringer ‘uniform bias’ (d.h. für die Kanadier erscheint *Rollentausch nicht gut* leichter abzulehnen zu sein als für die Deutschen und Spanier).



## 4 Überprüfung der externen Konsistenz

Für die Überprüfung der externen Konsistenz gilt eine entsprechende Logik wie bei der internen Konsistenz, nur wird hier nicht die Beziehung der Items untereinander und zu (einem als valide geltenden Indikator) der zugrundeliegenden Dimension untersucht, sondern die Beziehungen zu Drittvariablen wie dem Alter.

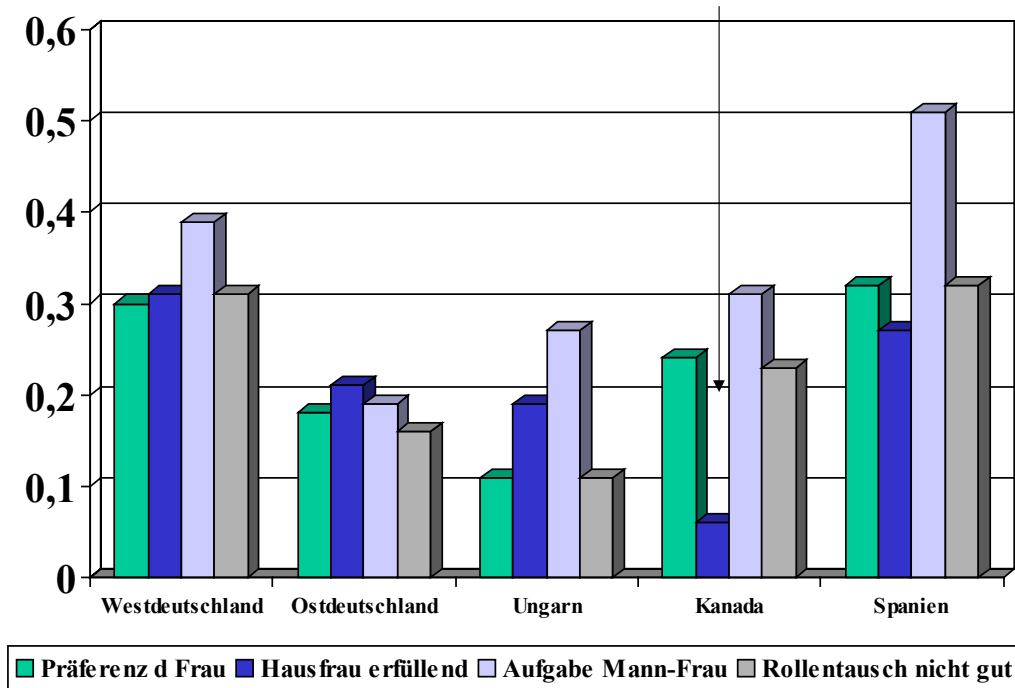
Dabei stellt der Fall, daß sich zwar die absoluten Werte der Regressionskoeffizienten für die betrachteten Items zwischen den Ländern unterscheiden, die Muster aber nicht (z.B. wenn die Regressionskoeffizienten für alle Items in einem Land geringer sind als in den anderen Ländern) kein zwingendes Anzeichen für Probleme der funktionalen Äquivalenz dar. In diesem Fall muß zunächst nach inhaltlichen Erklärungen für die unterschiedliche Stärke der Beziehung zwischen dem untersuchten Konstrukt und den Drittvariablen gesucht werden. Im Falle des Alters als Drittvariable - und einer Interpretation als Kohorteneffekt - wäre etwa zunächst zu prüfen, ob die Sozialisationsbedingungen für die einzelnen Kohorten in den verschiedenen Ländern unterschiedlich waren, so daß eine gleichartige Beziehung zu dem Konstrukt in den einzelnen Ländern schon von daher nicht zu erwarten ist. Interkulturell unterschiedliche Zusammenhänge der Items mit Drittvariablen sollten also nicht mechanisch als Anzeichen fehlender Äquivalenz gedeutet werden. Dies gilt um so mehr, wenn aufgrund bestimmter Charakteristika der zu vergleichenden Länder bereits theoretisch abgeleitet werden kann, daß die Beziehungen gar nicht gleich sein sollten. Dies gilt jedoch nicht für Unterschiede zwischen den Beziehungen der verschiedenen Items mit der Drittvariablen für ein gegebenes Land.

Tabelle 3: Korrelationen der Items mit Alter

	<i>Präferenz d Frau</i>	<i>Hausfrau erfüllend</i>	<i>Aufgabe Mann- Frau</i>	<i>Rollentausch nicht gut</i>
Westdeutschland	-.30	-.31	-.39	-.31
Ostdeutschland	-.18	-.21	-.19	-.16
Ungarn	-.11	-.19	-.27	-.11
Kanada	-.24	<b>-.06</b>	-.31	-.23
Spanien	-.32	-.27	-.51	-.32

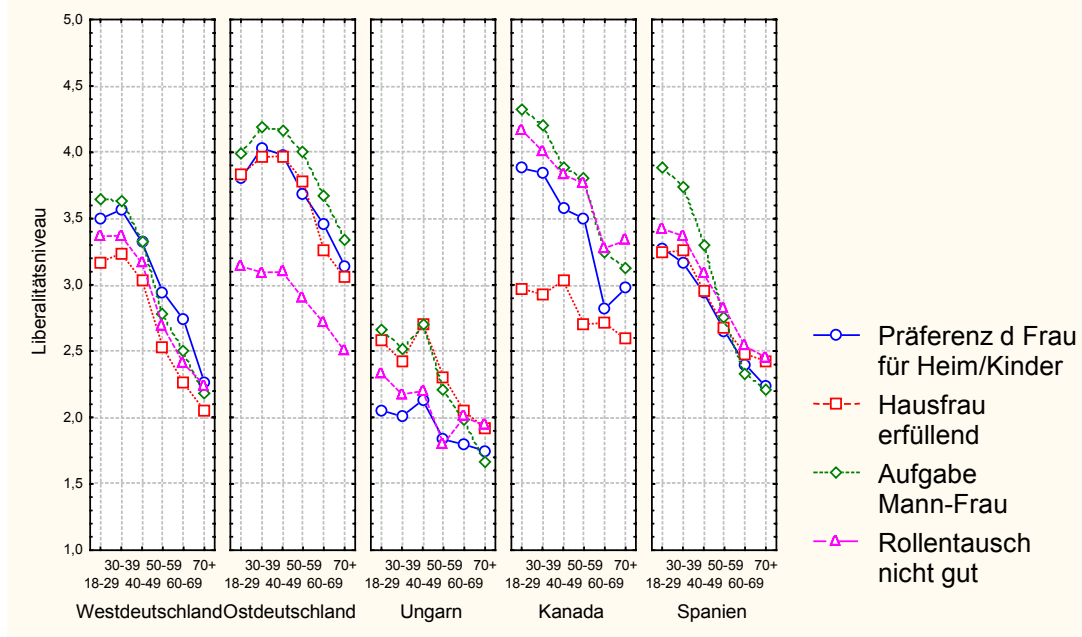
Tabelle 3 und Abbildung 9 zeigen, daß in Kanada *Hausfrau erfüllend* eine - gemessen an den anderen Korrelationen für Kanada - extrem geringe Korrelation mit dem Alter hat. In Ostdeutschland dagegen ist auf der Grundlage der Korrelationen das in den vorherigen Analysen belegte Problem beim Item *Rollentausch nicht gut* nicht nachzuweisen.

Abbildung 9: Korrelationen der Items mit Alter



Woran dies liegt, zeigt Abbildung 10: Der Zusammenhang dieses Items mit dem Alter ist eher linear, für die anderen drei Items aber kurvilinear (mit einer Trendwende im Liberalisierungsprozeß für die jüngeren Kohorten).

Abbildung 10: Items in Abhängigkeit von der Altersgruppe



## 5 Fazit

Eine Evaluation der methodischen Qualität der Meßinstrumente vor den eigentlichen inhaltlichen Analysen ist in jedem Fall sinnvoll. Ein Vergleich von Ländern aufgrund von Indizes, in die nicht vergleichbare Items eingehen, verzerrt die faktisch bestehenden Unterschiede. Es empfiehlt sich in jedem Fall, die Anwendung mehrerer Verfahren zur Überprüfung der funktionalen Äquivalenz, da nicht alle die gleichen Probleme erkennen können. So impliziert insbesondere externe Konsistenz nicht auch automatisch interne Konsistenz (wie in diesem Beispiel am Falle Ungarns sichtbar). Weiterhin kann die Analyse linearer Regressionskoeffizienten in die Irre führen, die Verwendung der graphischen Darstellung vermeidet entsprechende Probleme. Schließlich ist zu erwähnen, daß eine Verwendung der MDS bei zu vielen Fehlern in den Daten nicht sinnvoll ist, da dann die eigentlich interessierenden Probleme im Gesamtbild nicht mehr deutlich werden können.

### Literatur

- Alwin, D.F./M. Braun, M./J. Harkness/J. Scott (1994), Measurement in Multi-National Surveys. In: I. Borg/P.Ph. Mohler (Eds.), *Trends and Perspectives in Empirical Social Research*. Berlin/New York: De Gruyter.
- Borg, I. (2000), Explorative Multidimensionale Skalierung. ZUMA How-to Serie 1. Mannheim: ZUMA.
- Borg, I./P.F.J. Groenen (1997), *Modern Multidimensional Scaling*. New York: Springer.
- Borg, I./S. Shye (1995), *Facet Theory: Form and Content*. Newbury Park, CA: Sage.
- Braun, M. (1994), The International Social Survey Program (ISSP). In: P. Flora et al. (Eds.), *Social Statistics and Social Reporting in and for Europe*. Bonn: Informationszentrum Sozialwissenschaften.
- Braun, M. (1998), Gender roles. In: J.W. van Deth (Ed.), *Comparative Politics. The Problem of Equivalence*. London/New York: Routledge.
- Braun, M. (2000), Einstellung oder Meinung? Änderung der Interpretation von Items im Zuge des sozialen Wandels und ihre Konsequenzen für die interkulturell vergleichende Sozialforschung. In: P.P. Mohler/P. Lüttinger (Hg.), *Querschnitt. Festschrift für Max Kaase*. Mannheim: ZUMA.
- Braun, M./W. Müller (1997), Measurement of education in comparative research. In: L. Mjøset et al. (Eds.), *Methodological Issues in Comparative Social Science*. Comparative Social Research 16. Greenwich, Conn.: JAI Press.
- Mason, K.O./J.L. Czajka/S. Arber (1976), Change in U.S. women's sex-role attitudes, 1964-1974. *American Sociological Review* 41: 573-596.
- Scheuch, E.K. (1968), The cross-cultural use of sample surveys: problems of comparability. In: S. Rokkan (Ed.), *Comparative research across cultures and nations*. Paris: Mouton.
- StatSoft (1999), STATISTIKA for Windows. Tulsa OK: StatSoft Inc.
- van Deth, J.W. (1998), Equivalence in comparative political research. In: J.W. van Deth (Ed.), *Comparative Politics. The Problem of Equivalence*. London/New York: Routledge.

van de Vijver, F.J.R./K. Leung (1997), *Methods and Data Analysis for Cross-Cultural Research*. London: Sage.

Zentralarchiv für Empirische Sozialforschung (1997) *ISSP 1994 - Family and Changing Gender Roles II*. Codebook ZA-No. 2620. Köln: Zentralarchiv.