SSOAR

Open Access Repository
www.ssoar.info

Der internationale Vergleich von Meßmodellen unter verallgemeinerten Verteilungsbedingungen

Faulbaum, Frank

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:
Faulbaum, F. (1990). Der internationale Vergleich von Meßmodellen unter verallgemeinerten Verteilungsbedingungen. *ZUMA Nachrichten*, *14*(26), 56-71. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-209903

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.

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Der internationale Vergleich von Meßmodellen unter verallgemeinerten Verteilungsbedingungen

Von Frank Faulbaum

In diesem Aufsatz soll anhand eines einfachen internationalen Vergleichs zwischen der BRD und den USA eine Methodologie vorgestellt werden, die es ermöglicht, die Invarianz von Modellannahmen über mehrere Stichproben (Nationen) hinweg auch dann zu überprüfen, wenn die übliche Normalverteilungsbedingung verletzt ist. Letzteres ist gerade in der sozialwissenschaftlichen Einstellungsforschung häufig der Fall.

1. Einleitung

Die im ALLBUS 1980, im ALLBUS 1982 und im General Social Survey (GSS) 1982 verwendete Itembatterie zur Erfassung von Einstellungen zum Beruf war wiederholt Gegenstand verschiedener statistischer und meßtheoretischer Untersuchungen (vgl. Borg 1986; Faulbaum 1983, 1984, 1986; Schmidt 1983; Wegener 1984). Besondere Aktualität gewinnt dieses Instrument dadurch, daß es im Rahmen des ISSP 1989 (International Social Survey Programme) wiederum – wenn auch in einer verkürzten Form – Teil einer großen internationalen Studie geworden ist, deren Ergebnisse in gewissem Umfang auch Vergleiche mit Ergebnissen der Studien der Jahre 1980 und 1982 zulassen werden. Dies ist Anlaß genug, noch einmal rückblickend zu fragen, welche Items der Berufswerteskala sich international als besonders stabil erwiesen haben. Wir wollen dabei insbesondere einer Hypothese nachgehen, für deren Plausibilität sich in explorativen Studien (vgl. Faulbaum 1984) Hinweise ergeben haben und deren adäquate Untersuchung durch neuere Entwicklungen in der Methodik der Mehrstichprobenanalyse von Strukturgleichungsmodellen ermöglicht wird.

Diese Hinweise besagen, daß von den drei der Itembatterie zugrundeliegenden Konstrukten "Extrinsische Orientierung", "Intrinsische Orientierung"und "Soziale Orientierung"das Konstrukt "Extrinsische Orientierung"durch eine besondere Stabilität über die Nationen BRD und USA hinweg gekennzeichnet ist. Unter Stabilität verstehen wir dabei den Sachverhalt, daß das Konstrukt in beiden Ländern durch ähnliche Items gekennzeichnet ist. Diese Items sind:

BRD

- Sichere Berufsstellung (SICHER)
- Hohes Einkommen (EINKOM)
- Gute Aufstiegsmöglichkeiten (AUFSTG)
- Ein Beruf, der anerkannt und geachtet wird (GEACHT)

USA

- Job Security (JOBKEEP)
- High Income (JOBPAY)
- Good opportunities of advancement (JOBRISE)
- An occupation that is recognized and respected (JOBHONR)

Sie wurden von den Befragten hinsichtlich ihrer Wichtigkeit für die berufliche Arbeit und den Beruf auf einer 7–stufigen Wichtigkeitsskala beurteilt. Die Frage lief in dieser Form jeweils in einer Splitversion des ALLBUS 1982 und des GSS 1982 mit 1504 bzw. 747 Befragten. Die anderen Splitversionen enthielten eine Alternativversion der Frage.

Die Stabilitätshypothese beinhaltet eine Aussage zum Vergleich der beiden Nationen BRD und USA bezüglich eines aus den genannten Items bestehenden Meßmodells für das Konstrukt "Extrinsische Orientierung". Im schwächsten Fall besagt die Hypothese, daß es sich in beiden Nationen um ein kongenerisches Meßmodell handelt, ohne daß damit eine Aussage über die Gleichheit der Ladungen oder der Meßfehler verbunden wäre. Das Ähnlichkeitsmaximum wäre erreicht, wenn alle Items in beiden Nationen gleich reliabel wären.

Untersuchungen dieser Art lassen sich mit Hilfe von Mehrstichprobenanalysen (Multipler Gruppenvergleich, Multisample Analysis) im Rahmen der Analyse von Kovarianzstrukturen durchführen. Diese Methode erlaubt es insbesondere, Einschränkungen von Parametern (wie etwa die Gleichheit von Ladungen) in einfachen und komplexen Modellen mit und ohne latente Variablen über Stichproben hinweg zu testen. Erst neuerdings ist es darüber hinaus möglich, solche Vergleiche auch unter verallgemeinerten Verteilungsbedingungen durchzuführen, was bedeutet, daß die Normalverteilungsannahme in den einzelnen Populationen nicht mehr erfüllt sein muß. Anwendungen dieser Methodologie sind noch sehr selten, so daß nur geringe Erfahrungen auf diesem Gebiet vorhanden sind (vgl. Muthén 1989).

Die Entwicklung von Schätzverfahren unter verallgemeinerten Verteilungsbedingungen ist insbesondere für die sozialwissenschaftliche Einstellungsforschung von Bedeutung, wo die Variablen sehr häufig nicht normalverteilt sind. Dies gilt insbesondere für die Wichtigkeitseinstufungen der Berufsmerkmale. Tabelle 1 zeigt Mittelwerte, Schiefewerte und Kurtosiswerte für beide Nationen. Es zeigt sich, daß insbesondere die USA stark erhöhte Schiefe– und Kurtosiswerte aufweisen. Die univariaten Kurtosiswerte sind für große Stich-

Tabelle 1: Univariate Statistiken für die Stichproben BRD und USA

Items	Mittelwert		Schiefe		Kurtosis	
	BRD[1)]	USA[2)]	BRD	USA	BRD	USA
SICHER (JOBKEEP)	6.104	6.268	-1.987	-2.244	3.107	5.059
EINKOM (JOBPAY)	5.399	5.783	- .862	-1.265	.573	1.918
AUFSTG (JOBRISE)	5.194	6.056	- .891	-1.767	.115	3.417
GEACHT (JOBHONR)	5.413	5.871	- .986	-1.277	.252	1.336

[1)] Anzahl der Beobachtungen nach fallweiser Elimination der fehlenden Werte: N=1488.

[2)] Anzahl der Beobachtungen nach fallweiser Elimination der fehlenden Werte : N=728.

proben ($N > 1000$) mit einem Standardfehler von $(24/N)^{1/2}$ normalverteilt. Die Standardfehler der Schiefewerte betragen für Stichproben > 200 $(6/N)^{1/2}$. Unter Zugrundelegung dieser Standardfehler sind alle Schiefe- und Kurtosiswerte in Tabelle 1 signifikant von Null verschieden. Allerdings muß dieser Sachverhalt mit einer gewissen Distanz betrachtet werden, da bei großen Stichproben wie sie in unserem Fall vorliegen, bereits geringfügige Abweichungen von der Normalverteilung in Bezug auf Schiefe und Kurtosis signifikant werden.

Verschiedene Untersuchungen (z.B. Harlow 1985) zeigen, daß auf der Normalverteilungstheorie beruhende Schätzungen (GLS und ML) und χ^2 - Indizes gegenüber moderaten Abweichungen von der Normalverteilungsannahme bemerkenswert robust sind. Dennoch ist es im konkreten Fall schwer zu entscheiden, ob die Robustheitsbedingungen erfüllt sind. Angesichts der beobachteten Verteilungsanomalien ist es sinnvoll sich anzusehen, wie sich die Bedingungen für die Akzeptanz oder Zurückweisung von Parametereinschränkungen über Stichproben hinweg bei Unterstellung verallgemeinerter Verteilungsbedingungen verändern.

Die dieser Arbeit zugrundeliegenden Analysen wurden mit dem Programm EQS (Version 3.0; vgl. Bentler 1989) durchgeführt. Sie hätten in wesentlichen Teilen gleichermaßen auch mit LISREL 7 (in Verbindung mit PRELIS; vgl. Jöreskog/Sörbom 1988) oder mit LISCOMP (vgl. Muthén 1988) durchgeführt werden können. Allerdings enthalten diese Programme keinen multivariaten Lagrange–Multiplikator–Test, mit dem sich die statistischen Konsequenzen der simultanen Freisetzung von Parametereinschränkungen über Stichproben hinweg prüfen lassen (zur Theorie vgl. Bentler 1989; Bentler/Dijkstra 1985; Lee/Bentler 1980; Satorra 1989). Insbesondere die Zerlegung des multivariaten Tests in eine Folge univariater Tests, mit der sich die Beiträge einzelner Einschränkungen zum multivariaten χ^2 aller Einschränkungen prüfen lassen, stellt eine wesentliche statistische Hilfe bei der Modellmodifikation dar (vgl. Bentler/Chou 1986).

2. Mehrstichprobenvergleiche unter verallgemeinerten Verteilungsbedingungen

Mehrstichprobenvergleiche lassen sich im allgemeinsten Fall auch dann durchführen, wenn die Verteilungen in den verschiedenen Populationen unterschiedlich sind (Bentler/Lee/Weng 1987). So könnte man die Parameter in den einzelnen Populationen z.B. auch unter der Annahme schätzen, daß die Meßwerte in einer Population normalverteilt, in einer zweiten elliptisch verteilt und in einer dritten arbiträr verteilt sind.

Nehmen wir an, wir hätten m Stichproben $g = 1, \ldots, m$ vorliegen. Dann werden die Parameter bei einer Mehrstichprobenanalyse so geschätzt, daß die folgende Funktion ein Minimum ergibt:

$$T = \sum_{g=1}^{m} n_g T_g. \tag{1}$$

Dabei ist n_g die Stichprobengröße der g–ten Stichprobe und

$$T_g = (s_g - \sigma_g)' W_g (s_g - \sigma_g) \qquad (g = 1, \ldots, m). \tag{2}$$

s_g ist hier der Vektor der empirischen Kovarianzen der Gruppe g, σ_g ist der Vektor der theoretisch abgeleiteten Kovarianzen und W_g stellt eine Gewichtsmatrix dar. In den von uns unter unterschiedlichen Verteilungsbedingungen durchgeführten Analysen nimmt T_g die folgenden Formen an:

A. Maximum Likelihood (ML; Normalverteilungsannahme):

$$T_g = 2^{-1} tr((S_g - \Sigma_g) W_g)^2. \tag{3}$$

Hierbei ist S_g die empirische Kovarianzmatrix, Σ_g die theoretische Kovarianzmatrix und

$$W_g = \hat{\Sigma}_g^{-1} \qquad (\hat{\Sigma}_g\text{: geschätzte Kovarianzmatrix}). \tag{4}$$

B. Generalized Least Squares (GLS; Normalverteilungsannahme):

T_g wie unter ML, $W_g = S_g^{-1}$.

C. ADF-Schätzer (Asymptotically distribution-free estimator; vgl. Browne 1984):

$$T_g = (s_g - \sigma_g)' W_g (s_g - \sigma_g). \tag{5}$$

mit $W_g = V_g^{-1}$, wobei ein typisches Element von V_g bei Weglassen des Gruppenindexes g die folgende Form hat:

$$v_{ij,kl} = s_{ij,kl} - s_{ij} s_{kl}. \tag{6}$$

s_{ijkl} ist das vierte multivariate Produktmoment der beobachteten Variablen um ihre Mittelwerte.

Der ADF-Schätzer heißt in EQS *AGLS*-Schätzer. In LISREL 7 kann man sich die Gewichtsmatrix mit PRELIS erstellen lassen. Bei Anwendung der Option WLS und Angabe der PRELIS-erzeugten Gewichtsmatrix erhält man eine ADF-Schätzung in LISREL. In LISCOMP arbeitet man mit der Option WF, um diesen Schätzer zu realisieren.

Man kann den ADF-Schätzer verwenden, wenn die Messungen beliebig verteilt sind. Seine Anwendung erfordert von der statistischen Theorie her große Stichproben, die bei Umfragen in der empirischen Sozialforschung jedoch häufig vorliegen. Das Problem liegt vor allem darin, daß die statistischen Eigenschaften dieser Schätzer nur asymptotisch gelten, d.h. nur für große Stichproben bewiesen sind. Einige Studien (vgl. Tanaka 1984; Harlow 1985) legen Stichprobengrößen von mindestens 400–500 Fällen nahe. PRELIS lehnt eine Berechnung der Gewichtsmatrix für Stichprobengrößen, die kleiner als 200 sind, ab. EQS weist die Berechnung des AGLS-Schätzers für Stichproben kleiner $p(p+1)/2$ (p: Anzahl der beobachteten Variablen) zurück.

EQS enthält zusätzlich zu den ADF-Schätzern noch Schätzer für den Fall elliptischer Verteilungsbedingungen, auf dessen Darstellung wir hier verzichten (vgl. Bentler 1989: 213). Bei den elliptischen Verteilungen handelt es sich um eine Klasse von Verteilungen, die sich von der Normalverteilung durch die Schwere bzw. Leichtheit der Verteilungsenden unterscheiden können und die die Normalverteilung als Spezialfall mit einem multivariaten Kurtosisparameter von Null enthalten. Die Anwendung der elliptischen Verteilungstheorie ist in unserem Fall problematisch, da bei Vorliegen einer multivariaten elliptischen Verteilung die Kurtosisparameter der univariaten Verteilungen der einzelnen gemessenen Variablen eigentlich gleich sein müßten, was nach Tabelle 1 nicht erfüllt zu sein scheint.

Die Minimierung der Funktion (1) führt zu einer χ^2-Statistik der Form

$$n_1 \hat{T}_1 + n_2 \hat{T}_2 + \ldots + n_m \hat{T}_m \qquad (\hat{T}_g = min T_g), \tag{7}$$

verteilt mit $p(p+1)/2 - q + r$ Freiheitsgraden (p: Anzahl der Varianzen und Kovarianzen in allen Gruppen; q: Anzahl der freien Parameter in allen Gruppen; r: Anzahl der Gleichheitsrestriktionen innerhalb und zwischen den Gruppen). Gl. (7) offenbart, daß die

Werte der Anpassungswerte nach der Stichprobengröße gewichtet in die Formel eingehen, so daß größere Stichproben einen entsprechend größeren Einfluß auf die Gesamtanpassung haben.

Lagrange–Multiplikator–Test

Mit Hilfe eines multivariaten Lagrange–Multiplikator–Tests läßt sich prüfen, ob die simultane Aufhebung der Gleichheitsrestriktionen zwischen den Gruppen zu einer signifikanten Modellverbesserung, gemessen an der Abnahme des χ^2–Wertes führt. Außerdem läßt sich der Beitrag jeder einzelnen Restriktion zum multivariaten Gesamt–χ^2 prüfen.

Der Lagrange–Multiplikator–Test entspricht einem D–Test (χ^2– Differenzentest; vgl. Anderson/Gerbing 1988; Bentler/Bonett 1980; Steiger/Shapiro/Browne 1985) für den Vergleich von einem Ausgangsmodell mit einem Modell, bei dem gewisse Restriktionen zwischen den Parametern aufgehoben sind.

3. Parameterrestriktionen zwischen den Gruppen und Strategie der Modellselektion

Ein Vergleich von BRD und USA hinsichtlich des Meßmodells für das Konstrukt "Extrinsische Orientierung"muß zunächst mit der Untersuchung der Frage beginnen, ob es sich in beiden Nationen überhaupt um ein Modell mit nur einem Konstrukt handelt, was wiederum zur Frage führt, ob die Hypothese eines kongenerischen Modells für beide Nationen akzeptiert werden kann.

An dieser Stelle sei noch einmal kurz die Definition eines *kongenerischen Meßmodells* in Erinnerung gebracht. Von einem kongenerischen Meßmodell (vgl. Alwin/Jackson 1980; Faulbaum 1983; Jöreskog 1971; Jöreskog/Sörbom 1988: 76; Steyer 1989) spricht man nach Jöreskog dann, wenn die einer Menge gemessener Variablen $x_1, x_2, \ldots, x_p$ zugeordneten Variablen der korrespondierenden wahren Werte $\tau_1, \tau_2, \ldots, \tau_p$ untereinander paarweise mit 1.0 korrelieren. In diesem Fall gibt es eine gemeinsame latente Variable F, mit $\tau_i = \lambda_i F$. Ein kongenerisches Meßmodell ist einem 1–Faktor–Modell äquivalent, wobei zusätzlich angenommen wird, daß die Fehlervariablen unkorreliert sind. In Abbildung 1 ist das Diagramm eines kongenerisches Meßmodells für das Konstrukt "Extrinsische Orientierung"abgebildet. Die $\lambda_i (i = 1, 2, 3, 4)$ bezeichnen die Faktorladungen, die $\epsilon_i (i = 1, 2, 3, 4)$ bezeichnen die Fehlervariablen.

Lord/Novick (1968:76) nennen eine Menge von Messungen τ–äquivalent, wenn die Messungen den gleichen wahren Wert besitzen. Parallelität liegt nach ihrer Definition dann vor, wenn die Messungen zusätzlich auch noch gleiche Fehlervarianzen aufweisen. In diesem Fall sind die Messungen in Bezug auf das Konstrukt auch gleich reliabel. Jöreskog

(1971 op. cit.) hat gezeigt, daß die τ–Äquivalenz gleichbedeutend mit der Parameterrestriktion

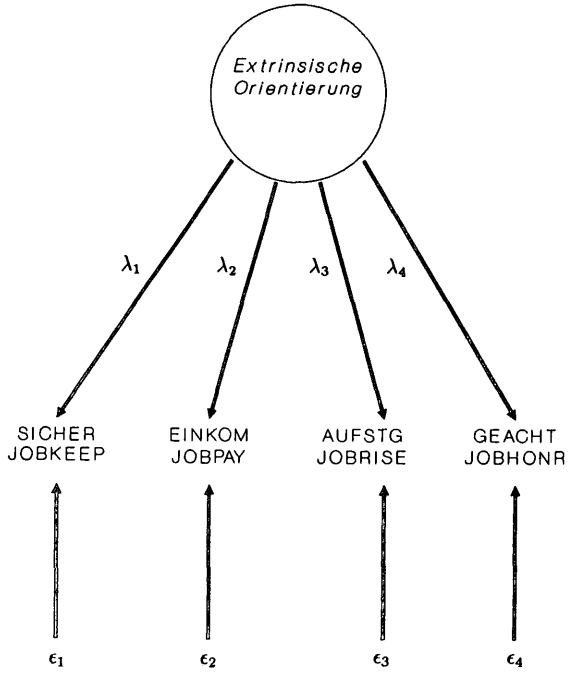
$$\lambda_1 = \lambda_2 = \ldots = \lambda_p \tag{8}$$

in einem kongenerischen Meßmodell mit p beobachteten Variablen ist und daß die Annahme der Parallelität gleichbedeutend mit der zusätzlichen Restriktion

$$\sigma^2_{\epsilon_1} = \sigma^2_{\epsilon_2} = \ldots = \sigma^2_{\epsilon_p} \tag{9}$$

ist. τ–äquivalente und parallele Meßmodelle sind also Spezialfälle kongenerischer Meßmodelle.

Abbildung 1: Kongenerisches Meßmodell



Die obigen Annahmen kann man bekanntlich im Rahmen der Analyse von Kovarianzstrukturen unter Verwendung der auf dem Markt befindlichen Computerprogramme überprüfen. Hat man in mehreren Stichproben, z.B. BRD und USA, die kongenerische Eigenschaft bestätigt, so kann man darüber hinaus weiter prüfen, ob die Ladungen und/oder die Fehlervarianzen über die Stichproben hinweg invariant sind. Der eingeschränkteste Fall liegt dann vor, wenn zusätzlich noch die Invarianz der Faktorvarianz behauptet wird. Letztere Einschränkung impliziert die Gleichheit der Reliabilitäten, da die Reliabilität einer Messung x_i in einer Gruppe g definiert ist durch

$$\frac{\sigma^2_{\tau_{i,g}}}{\sigma^2_{x_{i,g}}} = \frac{\lambda^2_{i,g}\sigma^2_{F,g}}{\sigma^2_{x_{i,g}}} = \frac{\lambda^2_{i,g}\sigma^2_{F,g}}{\lambda^2_{i,g}\sigma^2_{F,g} + \sigma^2_{\epsilon_{i,g}}} \tag{10}$$

und man alle Größen in dieser Formel durch die Restriktion über die Gruppen hinweg gleichgesetzt hat.

Wir sind in unserer Untersuchung davon ausgegangen, daß das uns eigentlich interessierende Modell (häufig als M bezeichnet) das Modell ist, welches die Invarianz der Reliabilitäten behauptet. Dieses Modell ist durch die folgenden Einschränkungen charakterisiert:

$$\begin{aligned}
\lambda_{i,BRD} &= \lambda_{i,USA} \quad (i = 1,2,3,4) \\
\sigma^2_{\epsilon_i,BRD} &= \sigma^2_{\epsilon_i,USA} \quad (i = 1,2,3,4) \\
\sigma^2_{F,BRD} &= \sigma^2_{F,USA} \quad (i = 1,2,3,4).
\end{aligned}$$

Wir haben dann unter der Voraussetzung, daß dies das wahre Modell ist, den multivariaten Lagrange–Multiplikator–Test angewendet. Ergab der Test, daß die Aufhebung einiger Restriktionen die Modellanpassung signifikant verbessert, so haben wir diese Aufhebung vorgenommen. Das resultierende Modell wurde unter der Voraussetzung akzeptiert, daß der dem ADF–Schätzer (AGLS) korrespondierende χ^2–Wert dieses weniger eingeschränkten Modells mindestens auf dem 1%–Niveau angepaßt werden kann und weiterhin die übrigen inkrementellen Anpassungsindizes von Bentler/Bonett, die das betrachtete Modell mit dem Unabhängigkeitsmodell (beobachtete Variablen unkorreliert; Nullmodell) vergleichen, sowie der komparative Fitindex von Bentler ebenfalls eine gute Anpassung signalisieren (zur Erklärung der Indizes vgl. Abschnitt 3). Auch die durchschnittlichen standardisierten Residuen in beiden Gruppen sollten möglichst klein sein (auf jeden Fall kleiner .050). Die Anwendung des Lagrange-Multiplikator-Tests entspricht durchaus den Empfehlungen anderer Autoren, wie z.B. Anderson/Gerbing (1988), über die Verwendung von χ^2–Differenzentests im Prozeß der Modellselektion.

4. Ergebnisse

Die Tabellen 2 und 3 zeigen zunächst die Ergebnisse der getrennten Einzelanalysen für beide Nationen, wobei zunächst die Frage im Mittelpunkt stand, ob es sich in beiden Nationen um ein kongenerisches Meßmodell handelt. Die zur Modellbeurteilung herangezogenen Gütekriterien seien an dieser Stelle nur kurz und ohne Angabe der Formeln informell erklärt. Eine Zusammenfassung der formalen Definitionen findet sich u.a. in Bentler (1989).

Die Definition von χ^2 setzen wir als bekannt voraus. NFI (Normed Fit Index), NNFI (Nonnormed Fit Index) und CFI (Comparative Fit Index) sind inkrementelle Anpassungsindizes. NFI und NNFI vergleichen die Werte der Anpassungsfunktionen an der Stelle der geschätzten Parameter für das betrachtete Ausgangsmodell mit den Werten für ein Modell, bei dem alle beobachteten Variablen unkorreliert sind (Nullmodell), wobei NNFI noch zusätzlich die Anzahl der Freiheitsgrade in Betracht zieht. Der NNFI gehört zu einer Klasse von Indizes, von denen Marsh/Balla/McDonald (1988) zeigen konnten, daß sie relativ unabhängig von der Stichprobengröße sind (Indizes vom Typ 2). Der CFI vergleicht die Nichtzentralitätsparameter. Er nimmt ebenso wie der NFI bei perfekter Modellanpassung den Wert 1.0 an. Der NNFI kann gelegentlich über 1.0 liegen. FI (FIT) und AFI (ADJUSTED FIT) schließlich, die den Indizes GFI und AGFI in LISREL 7 entsprechen, setzen das Minimum der Anpassungsfunktion vor und nach der Modellanpassung zueinander in Beziehung, wobei beim AFI auf die Anzahl der Freiheitsgrade relativiert wird. Die in den Tabellen mit AASR bezeichneten Werte stellen die durchschnittlichen absoluten standardisierten Residuen (average absolute standardized residuals) zwischen empirischer Kovarianzmatrix und modellimplizierter Kovarianzmatrix dar.

In die Tabellen 2 und 3 mit aufgenommen wurden ebenfalls die GLS–Schätzungen unter elliptischen Verteilungsbedingungen (EGLS–Schätzungen). Betrachten wir die Tabellen einzeln, so ergibt sich, daß wir im Fall der BRD die Nullhypothese eines kongenerischen Meßmodells auf dem 1%–Niveau nicht zurückweisen können, gleichgültig, ob wir die Schätzverfahren unter Normalverteilungsbedingung oder verallgemeinerten Ver-

Tabelle 2: Anpassungsindizes–Kongenerisches Modell BRD

Indizes	ML	GLS	EGLS	AGLS
χ^2	5.122 ($df = 2; p = .077$)	5.058 ($df = 2; p = .080$)	3.108 ($df = 2; p = .211$)	3.583 ($df = 2; p = .167$)
NFI	.998	1.000	1.000	.992
NNFI	.996	.999	1.000	.992
CFI	.999	1.000	1.000	.996
FI	-	-	-	.997
AFI	-	-	-	.987
AASR	.005	.006	.005	.006

Tabelle 3: Anpassungsindizes–Kongenerisches Modell USA

Indizes	ML	GLS	EGLS	AGLS
χ^2	12.686 ($df = 2; p = .002$)	11.674 ($df = 2; p = .003$)	5.645 ($df = 2; p = .060$)	5.361 ($df = 2; p = .069$)
NFI	.978	.997	.996	.943
NNFI	.945	.992	.993	.886
CFI	.982	.997	.998	.962
FI	-	-	-	.982
AFI	-	-	-	.936
AASR	.018	.022	.020	.026

teilungsbedingungen betrachten. Außerdem sieht man, daß sich die Modellanpassung im letzteren Fall, wo den Verteilungsanomalien Rechnung getragen wird, stark verbessert. Auch die übrigen Anpassungsindizes liegen durchweg sehr hoch.

Etwas anders liegt die Situation, wenn wir die Werte für die USA betrachten. Hier müßten wir die Hypothese eines kongenerischen Meßmodells, gemessen an den χ^2–Werten unter Normalverteilungsbedingungen (ML und GLS) eigentlich zurückweisen. Ein Blick auf die unter verallgemeinerten Verteilungsbedingungen ermittelten χ^2–Anpassungsindizes zeigt aber, daß diese relativ zur Anzahl der Freiheitsgrade nahezu halbiert werden und wir die Hypothese eines kongenerischen Modells unter diesen Bedingungen sogar auf dem 5%–Niveau akzeptieren können. Die Vermutung liegt nahe, daß dieses Ergebnis auf die stärkeren Verteilungsanomalien der USA–Stichprobe zurückzuführen ist. Die übrigen Indizes liegen ebenfalls deutlich über .900, so daß wir die Hypothese eines kongenerischen Modells für die Stichprobe der USA aufrechterhalten. Insgesamt jedoch ist die Anpassung nicht so gut wie im Fall der BRD, was sich insbesondere auch in der Höhe der standardisierten Residuen widerspiegelt. Bei dem schwachen Wert für den NNFI bei AGLS müssen wir bedenken, daß dieser Index in Simulationsstudien eine beträchtliche Varianz aufweist. Abnormes Verhalten wurde außerdem bei kleinen Stichproben beobachtet, wo dieser Index im Gegensatz zu den anderen klein ausfällt (vgl. Anderson/Gerbing 1984).

Weder für die BRD noch für die USA ließen sich τ–äquivalente Modelle anpassen.

Tabelle 4 enthält die Anpassungswerte für den Gruppenvergleich BRD–USA unter der Restriktion gleicher Reliabilitäten (oberer Teil), sowie unter der Annahme, daß die Faktorvarianzen und die Ladung von EINKOM auf dem Faktor in beiden Gruppen unterschiedlich sind (unterer Teil).

Wenden wir uns dem oberen Teil der Tabelle zu, so erkennen wir, daß die χ^2–Werte alle zu einer Ablehnung des Modells auf dem 1%–Niveau führen müßten. Die in den letzten beiden Zeilen enthaltenen durchschnittlichen standardisierten Residuen für BRD (AASR1) und USA (AASR2) sind noch sehr hoch, so daß wir trotz der relativ hohen inkrementellen Indizes eher zu einer Ablehnung des Modells mit Invarianz der Reliabilitäten tendieren. Hinweise darüber, welche der zwischen den Gruppen eingeführten Restriktionen zu einer signifikanten Verbesserung des Modells beitragen könnten, lassen sich aus den Ergebnissen

Tabelle 4: Gruppenvergleich BRD/USA

Invarianz der Reliabilitäten			
Indizes	ML	GLS	AGLS
χ^2	121.713 ($df = 12; p < .001$)	134.354 ($df = 12; p < .001$)	78.428 ($df = 12; p < .001$)
NFI	.960	.992	.849
NNFI	.964	.993	.870
CFI	.964	.993	.870
FI	-	-	.956
AFI	-	-	.927
AASR1	.098	.197	.167
AASR2	.296	.159	.198
Faktorvarianzen und Ladungen von EINKOM frei geschätzt			
Indizes	ML	GLS	AGLS
χ^2	24.464 ($df = 10; p = .007$)	23.559 ($df = 10; p = .009$)	13.600 ($df = 10; p = .192$)
NFI	.992	.999	.974
NNFI	.994	.999	.992
CFI	.995	.999	.993
FI	-	-	.992
AFI	-	-	.985
AASR1	.008	.008	.009
AASR2	.029	.029	.036

des multivariaten Lagrange-Multiplikator-Tests entnehmen.

Tabelle 5 enthält die χ^2-Indizes unter ML und AGLS bei simultaner Freisetzung aller signifikanten Restriktionen, sowie die univariaten Beiträge zum multivariaten χ^2, die unter mindestens einer der beiden genannten Methoden signifikant sind. Die signifikanten univariaten Beiträge betreffen die Faktorvarianz sowie die Ladung von EINKOM auf dem Konstrukt.

Nach Freisetzung dieser Parameter ergaben sich die Ergebnisse des unteren Teils von Tabelle 4. Das modifizierte Modell ist unter AGLS sehr gut angepaßt, die Residuen haben sich nunmehr stark reduziert und die inkrementellen Indizes liegen relativ nahe an 1.0.

Dieses Ergebnis deutet darauf hin, daß sich die Gruppen vor allem in der Bedeutung der Einkommensvariablen für das Konstrukt unterscheiden. Eine Inspektion der Ladungen (Tabelle 6) zeigt, daß die Variable "Hohes Einkommen"in der USA-Stichprobe einen höheren Stellenwert aufweist als in der BRD-Stichprobe. Die Tabelle zeigt, daß die

Tabelle 5: Ergebnisse des Lagrange-Multiplikator-Tests

	ML	AGLS
Multivariates Gesamt-χ^2	86.523 ($p = .000$)	33.848 ($p = .000$)
Univariater Beitrag von		
σ^2_F	68.728 ($p = .000$)	27.239 ($p = .000$)
λ_{EINKOM}	11.939 ($p = .001$)	4.130 ($p = .042$)

Einkommensvariable in der BRD-Stichprobe die geringste Bedeutung für das Konstrukt besitzt.

Die Unterschiedlichkeit beider Stichproben hinsichtlich der Variablen EINKOM legt eine Reanalyse unter Weglassung dieser Variablen nahe. Die Ergebnisse einer solchen Analyse unter Annahme der Invarianz von Ladungen und Fehlervarianzen sind in Tabelle 7 wiedergegeben. Die Ergebnisse zeigen eine sehr gute Anpassung des Modells, wenn auch die durchschnittlichen Residuen eine etwas schlechtere Anpassung in der USA-Stichprobe signalisieren. Ein mit LISREL 7 durchgeführter Vergleich der latenten Mittelwerte mit ML und unter der Bedingung der Invarianz der Interzeptstruktur war allerdings nicht signifikant. Er ergab eine signifikante Mittelwertsdifferenz des Konstrukts von $-.671$ (Referenzgruppe USA), was jedoch zumindest einen Hinweis darauf liefert, daß die Stichprobe der BRD deutlich geringer extrinsisch orientiert ist als die Referenzgruppe USA.

Bei Analysen der Ähnlichkeiten von Meßmodellen muß damit gerechnet werden, daß sich bei Disaggregation der totalen Population nach sozialen Teilgruppen gewisse Ähnlichkeiten verstärken, andere abschwächen. Ohne daß wir die Ergebnisse hier aus Platzgründen tabellarisch darstellen können, gilt z.B., daß in der BRD-Stichprobe der männlichen Befragten (N=678) die Hypothese des kongenerischen Modells nur mit Einschränkungen aufrechterhalten werden kann, und auch in den USA (N=221) ist die Anpassung keineswegs zufriedenstellend. Bei den weiblichen Befragten der BRD (N=810) und der USA (N=507) dagegen sind die kongenerischen Modelle nahezu perfekt angepaßt.

Entsprechende Unterschiede zeigen sich auch bei Vergleichen BRD-USA innerhalb der

Tabelle 6: Itemladungen BRD und USA (Einzelanalysen; ML-Schätzung)

Items	BRD			USA		
	ust[1)]	st[2)]	Rang	ust	st	Rang
SICHER	1.183	.752	2	.771	.590	3
EINKOM	1.014	.708	4	.889	.679	2
AUFSTG	1.415	.850	1	.981	.766	1
GEACHT	1.215	.741	3	.696	.517	4

[1)] unstandardisierte Werte

[2)] standardisierte Werte

Tabelle 7: Vergleich BRD/USA unter Annahme der Invarianz von Fehlervarianzen und Ladungen

Indizes	ML	GLS	EGLS	AGLS
χ^2	7.711 ($df = 5; p = .173$)	7.263 ($df = 5; p = .202$)	5.714 ($df = 5; p = .335$)	3.890 ($df = 5; p = .565$)
NFI	.996	.999	.999	.946
NNFI	.998	1.000	1.000	1.000
CFI	.999	1.000	1.000	1.000
FI	-	-	-	.997
AFI	-	-	-	.994
AASR1	.005	.004	.010	.004
AASR2	.019	.029	.021	.057

Tabelle 8: Vergleich BRD/USA innerhalb der Geschlechtergruppen mit freigesetzter Faktorvarianz und freigesetzten Ladungen von EINKOM

Männer			
Indizes	ML	GLS	AGLS
χ^2	23.854 ($df = 9; p = .005$)	22.691 ($df = 9; p = .007$)	23.350 ($df = 9; p = .005$)
NFI	.981	.997	.909
NNFI	.984	.998	.922
CFI	.988	.998	.941
FI	-	-	.972
AFI	-	-	.938
AASR1	.014	.014	.023
AASR2	.043	.050	.077
Frauen			
Indizes	ML	GLS	AGLS
χ^2	14.247 ($df = 9; p = .162$)	15.343 ($df = 9; p = .120$)	7.690 ($df = 9; p = .659$)
NFI	.992	.999	.974
NNFI	.997	.999	1.010
CFI	.998	.999	1.000
FI	-	-	.993
AFI	-	-	.985
AASR1	.011	.022	.014
AASR2	.035	.034	.031

einzelnen Gruppen. Tabelle 8 zeigt die Ergebnisse eines Vergleichs in Bezug auf das Modell mit freigesetzter Faktorvarianz und freigesetzte Ladungen für die Variable EINKOM. Wie man sieht, ist das Modell für die Gruppe der weiblichen Befragten deutlich besser angepaßt. Unter Gesichtspunkten der Modelltestung müßte man das Modell für die Gruppe der Männer zurückweisen. Unter Anpassungsgesichtspunkten wäre diese Entscheidung angesichts der hohen inkrementellen Indizes und der relativ niedrigen durchschnittlichen standardisierten Residuen jedoch nicht unproblematisch.

5. Zusammenfassung

Die Ergebnisse des Vergleichs BRD–USA bezüglich des Konstrukts "Extrinsische Orientierung"brachten deutliche Hinweise dafür, daß wir die Aussage der Stabilität der extrinsischen Orientierung über die Nationen hinweg einschränken müssen. Die Einschränkung betrifft die Konstruktvarianz (Varianz der wahren Werte) einerseits und die Ladung für das Item "Hohes Einkommen"andererseits, die das Konstrukt in beiden Nationen auch über bestimmte soziale Teilgruppen hinweg diskriminieren.

Ob es sich bei diesen Befunden um zeitstabile Konstanten innerhalb der Nationen und um zeitstabile Unterschiede zwischen den Nationen handelt, ist eine Frage, auf die die Analyse der 1989 erhobenen ISSP–Daten möglicherweise eine Antwort geben kann. Insbesondere hat man dann die Möglichkeit zu untersuchen, wie sich diese Befunde in den Kontext anderer westlicher und östlicher Nationen einordnen lassen.

Bei aller Betonung der Unterschiede überwiegen jedoch die Ähnlichkeiten. Offensichtlich unterscheiden sich beide Nationen nicht in den Einflußgrößen und den Meßfehlern der übrigen Variablen. Weder in der Frage des Strebens nach Aufstieg, noch in der Bewertung der Wichtigkeit der Sicherheit, noch in dem Streben nach Anerkennung schien es im Jahre 1982 zwischen den Staaten nennenswerte Unterschiede zu geben. Ob es sich hier jedoch um grundlegende Urteilsinvarianten westlicher Gesellschaften handelt, ist eine Frage, die man unter empirischen Gesichtspunkten auf der Basis dieser Untersuchung noch nicht entscheiden kann. Unter meßtheoretischen Gesichtspunkten haben die Analysen Hinweise dafür ergeben, wie die Meßmodelle mit extrinsischen Items in umfassenderen Kausalmodellen strukturiert sein könnten. Erst bei der Analyse solcher Modelle wird sich zeigen, ob die Struktur des Meßmodells auch in komplexeren Modellen erhalten bleibt. Die Analysen haben darüber hinaus Hinweise dafür erbracht, daß bei Verletzungen der Normalverteilungsannahme die Anpassung erheblich verbessert werden kann, wenn man die Gruppenvergleiche unter verallgemeinerten Verteilungsbedingungen durchführt.

Literatur

Alwin, D.F./Jackson, D.J., 1980: Measurement models for response errors in surveys: Issues and applications. S.68–119 in: K. Schuessler (ed.), Sociological Methodology 1980. San Francisco: Jossey Bass.

Anderson, J.C./Gerbing, D.W., 1984: The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analyses. Psychometrika 49: 155–173.

Anderson, J.C./Gerbing, D.W., 1988: Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin 103: 411–423.

Bentler, P.M., 1989: EQS. Structural Equations Program Manual. Los Angeles, California: BMDP Statistical Software.

Bentler, P.M./Bonett, D.G., 1980: Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 88: 588–606.

Bentler, P.M./Chou, C.-P., 1986: Statistics for parameter expansion and contraction in structural equation models. Paper presented at American Educational Research Association Meeting, San Francisco, California.

Bentler, P.M./Dijkstra, T., 1985: Efficient estimation via linearization in structural models. S. 9–42 in: P.R. Krishnaiah (Ed.), Multivariate Analysis VI. Amsterdam: North Holland.

Bentler, P.M./Lee, S.-Y./Weng, L.-J., 1987: Multiple population covariance structure analysis under arbitrary distribution theory. Communications in Statistics – Theory 16: 1951–1964.

Borg, I., 1986: A cross-cultural replication of Elizur's facets of work values. Multivariate Behavioral Research 21: 401—-410.

Faulbaum, F., 1983: Konfirmatorische Analysen von Wichtigkeitseinstufungen beruflicher Merkmale. ZUMA-Nachrichten 14: 46–59.

Faulbaum, F., 1984: Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982. ZUMA-Arbeitsbericht Nr. 04/84.

Faulbaum, F., 1986: Gruppenvergleiche latenter Mittelwerte von Berufsorientierungen. ZUMA-Nachrichten 18: 48–62.

Harlow, L.L., 1985: Behavior of some elliptical theory estimators with nonnormal data in a covariance structure framework: A Monte Carlo study, Ph.D. Thesis, University of California, Los Angeles.

Jöreskog, K.G., 1971: Structural analysis of sets of congeneric tests. Psychometrika 36: 109–133.

Jöreskog, K.G./Sörbom, D., 1988: LISREL 7. A Guide to the Program and Applications. Chicago, Ill.: SPSS Inc.

Lee, S.-Y./Bentler, P.M., 1980: Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. South African Statistical Journal 14: 121–136.

Lord, F.M./Novick, M.E., 1968: Statistical Theories of Mental Test Scores. Reading: Addison-Wesley Publ. Co.

Marsh, H.W./Balla, J.R./McDonald, R.P., 1988: Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin 103: 391–410.

Muthén, B.O., 1988: LISCOMP. A Program for Advanced Research. Mooresville, IN: Scientific Software.

Muthén, B.O., 1989: Multiple–group structural modelling with non–normal continuous variables. British Journal of Mathematical and Statistical Psychology 42: 55–62.

Satorra, A., 1989: Alternative test criteria in covariance structure analysis. Psychometrika 54: 83–90.

Schmidt, P., 1983: Messung von Arbeitsorientierungen: Theoretische Fundierung und Test alternativer Meßmethoden. Analyse und Kritik 2: 115–153.

Steiger, J.H./Shapiro, A./Browne, M.W., 1985: On the multivariate asymptotic distribution of sequential chi–square statistics. Psychometrika 50: 253–264.

Steyer, R., 1989: Experiment, Regression und Kausalität. Habilitationsschrift. Universität Trier.

Tanaka, J.S., 1984: Some results on the estimation of covariance structure models, Ph.D. Thesis, University of California, Los Angeles.

Wegener, B., 1984: Ergebnisse der Reliabilitätsstudie 1982. ZUMA–Arbeitsbericht, Mannheim.