

## Zur Konstruktion von Einstellungsskalen im internationalen Vergleich

Krebs, Dagmar; Schuessler, Karl F.

Veröffentlichungsversion / Published Version  
Forschungsbericht / research report

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Krebs, D., & Schuessler, K. F. (1986). *Zur Konstruktion von Einstellungsskalen im internationalen Vergleich*. (ZUMA-Arbeitsbericht, 1986/01). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-66348>

### 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.

Zur Konstruktion von Einstellungsskalen  
im internationalen Vergleich

Dagmar Krebs  
Karl F. Schuessler

ZUMA-Arbeitsbericht Nr. 86/01

Zentrum für Umfragen, Methoden  
und Analysen e. V. (ZUMA)  
Postfach 59 69  
D-6800 Mannheim 1  
Tel. (06 21) 18 00 40



## Einleitung

Dieser Bericht stellt den Teil einer interkulturell vergleichenden Studie dar, der sich in erster Linie mit der Konstruktion von Meßinstrumenten zur interkulturellen Anwendung befaßt. Bei diesen Instrumenten handelt es sich um Skalen zur Erfassung eines allgemeinen sozialen Lebensgefühls. Hierzu zählen Empfindungen wie Entfremdung, Machtlosigkeit und Zynismus ebenso wie interne/externe Kontrollerwartungen, Vertrauen, Zukunftsorientierung und Optimismus. Das "allgemeine soziale Lebensgefühl" umfaßt sowohl Einschätzungen gesellschaftlicher und sozialer Verhältnisse als auch Beurteilungen der subjektiv wahrgenommenen Stimmungen und affektiven Zustände, die als Konsequenzen des Lebens unter solchen Verhältnissen auftreten können.

Die Besonderheit dieser Skalenkonstruktion besteht darin, daß die Items aus einer Vielzahl von Skalen "abstammen", die im Laufe einer über 30jährigen Forschungstradition zur Erfassung individueller und kollektiver Reaktionen auf strukturell bedingte schwierige Lebenssituationen und damit zur Messung der subjektiven Reflektion gesellschaftlich-struktureller Gegebenheiten benutzt wurden. Die Konstruktion derartiger Skalen war - historisch betrachtet - häufiger das Resultat einer Reaktion auf gesellschaftliche Problemsituationen als das Resultat theoretischer Deduktion.

Der Ausgangspunkt für die Studie, von der hier nur der die Skalenkonstruktion betreffende Teil behandelt wird, lag in der Beobachtung, daß es im Bereich der Erfassung subjektiver Reaktionen auf gesellschaftliche Zustände eine Vielzahl von Meßinstrumenten gibt, die "... in some cases (are) nominally identical but substantively different and, in other cases (are) nominally different but substantively identical" (Schuessler, 1982:10). Da durch ständige Neukonstruktion von Meßinstrumenten eine Kumulation von Erfahrungen und Erkenntnissen unterbunden wird, begann Schuessler 1974 der Frage nachzugehen, ob es möglich ist, aus einer Vielzahl von ad hoc-Skalen (zur Messung des allgemeinen sozialen Lebensgefühls) einige wenige Skalen zu konstruieren, die für die Zielsetzung der meisten Untersuchungen, Beziehungen zwischen dem allgemeinen sozialen Lebensgefühl und sozialen Ausgangsbedingungen darzustellen, ausreichen würden.

Aus einem nach Durchsicht der relevanten Literatur (anglo-amerikanische Zeitschriftenaufsätze und Bücher) gewonnenen Itempool von über 1800 Items extrahierte Schuessler (1982) auf der Basis unterschiedlicher Analysen 237 Items, die er einer repräsentativen Stichprobe von N=1522 amerikanischen Befragten zur Beantwortung vorlegte. Die Auswertung der Antworten erbrachte eine weitere Reduktion der Itemzahl auf 120. Diese 120 Items verteilten sich in der amerikanischen Stichprobe auf 12 SLF (Social Life Feeling)-Skalen und 2 Skalen zur Messung von Antworttendenzen (social desirability und response desirability). Nach Voruntersuchungen zur deutschen Anwendung dieser Skalen wurden weitere 25 Items eliminiert, so daß 95 Items in die deutsche Studie eingingen. Diese 95 Items bilden 9 SLF-Skalen und 2 Skalen zur Messung von Antworttendenzen. Im vorliegenden Bericht werden nur die 9 SLF-Skalen behandelt, was den Itempool auf 80 Items reduziert. Die Skalen zur Messung von Antworttendenzen werden bei Krebs & Schuessler (1987) dokumentiert. Dort werden auch die Effekte demographischer Befragtenmerkmale auf das allgemeine soziale Lebensgefühl und die Antworttendenzen sowie die Beziehung zwischen Antworttendenzen und allgemeinem sozialen Lebensgefühl behandelt.

#### Die Entwicklung von Skalen zur Messung des allgemeinen sozialen Lebensgefühls in der amerikanischen und der deutschen Stichprobe

Das allgemeine soziale Lebensgefühl (Social Life Feeling) ist das Resultat einer Kombination von Wahrnehmungen, Beurteilungen und Bewertungen über die Beschaffenheit der sozialen Umwelt einerseits und von affektiven Zuständen, die sich als Konsequenz des Lebens unter konkreten Bedingungen in eben dieser sozialen Umwelt ergeben, andererseits. Die konkreten Items beinhalten sowohl Feststellungen zur Person, z.B. subjektives Befinden des Befragten, Bewertung seiner eigenen Situation, Lebensphilosophie des Befragten, als auch Feststellungen über die Gesellschaft, z.B. Ansichten über die Menschen allgemein oder über die Institutionen, insbesondere die Behörden. Die hier vorliegenden Items sind in der soziologischen Forschung als Indikatoren für die unterschiedlichsten theoretischen Konstrukte verwendet worden, indem sie von verschiedenen Forschern zu ad hoc-Skalen zusammengesetzt wurden. Diese ad hoc, jeweils für die Zwecke eines speziellen Forschungsvorhabens entwickelten Skalen waren weder nominell noch substantiell einheitlich.

Das Anliegen dieser Studie ist es nun, einen vorhandenen umfangreichen Itempool auf wenige zentrale Dimensionen zur Messung des allgemeinen sozialen Lebensgefühls zu reduzieren. Auf diese Weise sollen Skalen mit einheitlichem Bedeutungsgehalt geschaffen werden, die in verschiedenen Untersuchungen - auch interkulturell - zur Messung des allgemeinen sozialen Lebensgefühls eingesetzt werden können. In diesem Sinne bezeichnen wir die hier entwickelten Skalen als synthetische Skalen, die eine Synthese aus Items verschiedener Abstammungsskalen darstellen und in denen nur die Items enthalten sind, die den unterschiedlichsten Prüfungen statistischer und inhaltlicher Art standgehalten haben.

Von den ursprünglich in der amerikanischen Untersuchung enthaltenen 237 Items wurden zu Beginn der deutschen Studie nur noch die 120 Items in Betracht gezogen, die bestimmten statistischen Auswahlkriterien in der amerikanischen Studie genügten (siehe unten). Die deutsche Studie begann mit Übersetzung und Rückübersetzung der Items (die Ergebnisse der Rückübersetzung sind in Krebs & Schuessler, 1987, dargestellt). Diese 120 Items wurden in einem Pretest von N=150 zufällig ausgewählten erwerbstätigen Personen der Mannheimer Bevölkerung in einem mündlichen Interview und von N=405 Studenten der Universitäten Gießen und Hamburg in einer schriftlichen Befragung beantwortet. Als Ergebnis dieser Voruntersuchungen wurden die an der amerikanischen Stichprobe entwickelten "social life feeling"(SLF)-Skalen 8 ("People Cynici"), 11 ("Feeling Demoralised") und 12 ("Career Concern") sowie 6 Items der RD ("responding desirably")-Skala aus dem Itempool eliminiert, so daß in der deutschen Befragung an einer Repräsentativstichprobe von N=2003 Personen noch 95 Items zum Einsatz kamen. Diese 95 Items verteilen sich in der amerikanischen Stichprobe auf 9 Dimensionen des allgemeinen sozialen Lebensgefühls und auf 2 Skalen zur Messung von Antworttendenzen.

In einem ersten Schritt soll überprüft werden, ob sich die Items, die in der deutschen Stichprobe zum Einsatz kamen, zu Dimensionen zusammenfassen lassen, die mit denen der amerikanischen Stichprobe identisch oder diesen vergleichbar sind. Die an der amerikanischen Stichprobe entwickelten Skalen werden hier kurzerhand als USA-Skalen bezeichnet.

## Statistische Kriterien:

### 1. Eindimensionalität

Die Skalen (Dimensionen des allgemeinen sozialen Lebensgefühls) sind aus dichotom zu beantwortenden Items zusammengesetzt. Sie werden als eindimensional angesehen, wenn die Interitem-Produkt-Moment-Korrelationen nur einen Eigenwert  $> 1.0$  haben und wenn mindestens 90% der beobachteten Korrelationen dem 1. Faktor zuzuschreiben sind. Das erstgenannte Kriterium soll gewährleisten, daß nur ein Faktor wesentlich zur Aufklärung der Gesamtvarianz beiträgt und somit deutlich über den Varianzanteilen anderer Faktoren liegt. Das zweitgenannte Kriterium bietet Gewähr dafür, daß Faktoren, die nach der auf Einfachstruktur ausgerichteten Hauptachsenfaktorenanalyse über den 1. Faktor hinaus extrahiert werden, insgesamt weniger als 10% zu den beobachteten Korrelationen beitragen.

Zur Erreichung des erstgenannten Kriteriums wurden exploratorische Faktorenanalysen durchgeführt, wobei die Faktoren nach dem Hauptachsenverfahren extrahiert wurden. Die Berechnung einer Faktorenanalyse ist bei dichotomen Daten dann zulässig, wenn je zwei zu korrelierende Items einigermaßen ausgeglichene Randverteilungen aufweisen, so daß eine Phi-Korrelationsmatrix berechnet werden kann, die den Ausgangspunkt für die Faktorenanalyse darstellt. Dieses Vorgehen ist deshalb zulässig, weil der Phi-Koeffizient bei dichotomen Daten dem Produkt-Moment-Korrelationskoeffizienten nach Pearson vergleichbar ist und weil die Faktorenanalyse keine Verteilungsannahmen macht. Zur Erreichung des zweitgenannten Kriteriums wurden konfirmatorische Faktorenanalysen durchgeführt, um möglichst exakte Angaben über die durch die extrahierten Dimensionen aufgeklärte Varianz in den Daten zu erhalten. Während die exploratorische Faktorenanalyse darauf anzielt, die Interkorrelationen der Items durch eine Anzahl von Faktoren zu reproduzieren, die geringer ist als die Anzahl der Items, ist es das Ziel der konfirmatorischen Faktorenanalyse, die Gesamtvarianz der Items zu reproduzieren. Faktoren der exploratorischen Faktorenanalyse können zwar die Interkorrelationen der Items exakt reproduzieren, sie erklären jedoch einen geringeren Anteil der Gesamtvarianz als dieselbe Anzahl der Dimensionen in einer konfirmatorischen Faktorenanalyse (Jöreskog, 1979:5-20). Gibt man an, in welchem Ausmaß die

beobachteten Korrelationen dem 1. Faktor zuzuschreiben sind, so impliziert das den Vergleich einer statistischen Prüfgröße, die unter der Hypothese der Nichtexistenz eines gemeinsamen Faktors berechnet wird, mit einer Prüfgröße, die unter der (Modell-)Annahme eines gemeinsamen Faktors berechnet wird (Tucker & Lewis, 1973). Das Arbeiten mit derartigen Prüfgrößen setzt jedoch die Normalverteilung der beobachteten Variablen voraus. Da unsere Daten diese Anforderungen nicht erfüllen, erfolgt hier keine probabilistische Interpretation der Prüfgrößen und der aus ihnen abgeleiteten Maßzahlen.

## 2. Reliabilität

Bei der Zusammenfassung der Items zu Dimensionen wurde also besonders auf Einfachststruktur geachtet. Bei der Entscheidung über die Dimensionalität wurde – wie oben ausgeführt – nicht nur ein Kriterium zugrunde gelegt, sondern es wurde von Fall zu Fall eine Gegenüberstellung der idealerweise wünschenswerten Kriterien und der real erreichten Ergebnisse vorgenommen. In diesem Prozeß wurden die Ideal-Kriterien oft verletzt und die Ansprüche zurückgeschraubt. Das gilt besonders für das Reliabilitätsmaß Alpha. Der Reliabilitätskoeffizient Alpha liefert eine eher konservative Schätzung der Zuverlässigkeit bzw. der internen Konsistenz eines Meßinstrumentes. Die auf den Dimensionen aufbauenden Skalen sollten nicht nur eindeutig hinsichtlich der zugrundeliegenden latenten Variablen sein, sie sollten auch zuverlässig sein in dem Sinne, daß bei wiederholten Messungen wenn nicht identische, so doch weitgehend ähnliche Antworten zu erwarten sind. Idealerweise sollte Alpha Werte von 0.80 und höher annehmen. Die mittlere Interitem-Korrelation innerhalb einer Dimension sollte mindestens 0.25 betragen. Dieser niedrige Wert berücksichtigt die Inkonsistenz im Antwortverhalten der Befragten, die zwischen dem mit 0 und 1 vercodeten Items hin- und herpendeln. Wäre diese Inkonsistenz der Antworten im Hinblick auf die Polung der Items nicht vorhanden, so gäbe es nur extreme Summenwerte von 0 und p ( $p$  = Anzahl der Items pro Dimension). Es ist daher unrealistisch, bei relativ kurzen Skalen, bei denen die Befragten sich auf alle Summenwertgruppen verteilen (sollen), eine wesentlich höhere mittlere Interitem-Korrelation als 0.25 und einen Alpha-Koeffizienten  $> 0.75$  zu erwarten. Anzumerken ist hier allerdings, daß die Akzeptabilität der numerischen Ausprägung des Alpha-Wertes davon abhängt, zu



welchen Aussagen dieser Koeffizient herangezogen wird. Geht es z.B. bei einem Geschicklichkeitstest darum, daß die Person mit dem höchsten Wert DM 100,- erhält, dann sollte die Test-Retest-Reliabilität (die Stabilität der Ergebnisse bei wiederholten Messungen) sehr hoch sein. Wird jedoch - wie in soziologischen Untersuchungen üblich - die Abhängigkeit der Skalensummenwerte von sozialen und demographischen Merkmalen der befragten Personen untersucht, so kann für diesen Zweck ein erheblich geringerer Alpha-Wert als akzeptabel abgesehen werden. Selbst wenn die Summenwerte bei wiederholten Messungen unstabil sind, bleibt das Muster der Beziehungen zwischen diesen und den demographischen Merkmalen erhalten.

#### SLF-Skalen der amerikanischen Stichprobe

Inwiefern sind nun die an der amerikanischen Stichprobe entwickelten Skalen in der deutschen Stichprobe reproduzierbar? Bei Abwägung aller genannten Kriterien konnte für die USA-Skalen selbst in der amerikanischen Stichprobe kein befriedigendes Ergebnis erzielt werden (Tabelle 1). Im linken Bereich dieser Tabelle sind die Kennwerte der USA-Skalen für die amerikanische und die deutsche Stichprobe aufgeführt. Die Anteile der erklärten Interitem-Korrelationen (je Skala) bewegen sich in der amerikanischen Stichprobe zwischen 0.86 und 0.96 bei einem Durchschnitt von 0.92; die Eigenwerte für den 2. Faktor liegen nahe bei 1.0, einige sogar darüber. Beide Kriterien weisen auf die Existenz eines zusätzlichen Faktors hin und werfen indirekt die Frage auf, ob es überhaupt möglich ist, SLF-Skalen zu entwickeln, die nur eine Dimension, d.h. eine latente Variable abbilden. Die Angaben in Tabelle 1 zeigen, daß die Skalenkriterien der USA-Skalen in der deutschen Stichprobe in noch geringerem Ausmaß erfüllt werden als in der amerikanischen Stichprobe. Die USA-Skalen sind also in der deutschen Stichprobe nicht ohne weiteres reproduzierbar.

Um nun eine Annäherung an die Skalenkriterien in der deutschen Stichprobe zu erreichen, wurden die USA-Skalen durch Eliminierung von Items modifiziert. Eliminiert wurden die Items, bei denen die Differenz zwischen den beobachteten und den (unter der Annahme eines einfaktoriellen Meßmodells) geschätzten Korrelationen am höchsten war. Auf diese Weise entstanden die USAM-Skalen, deren Kennwerte im rechten Bereich der Tabelle 1 enthalten sind. Die

**Tabelle 1:** Skalenskennwerte der an der amerikanischen Stichprobe entwickelten USA-Skalen und deren Modifikation. USAM-Skalen nach Anwendung in der deutschen Stichprobe

Skalen-Nr.	Skalenname	Anzahl Items	USA-Skalen				USAM-Skalen				
			AM: N=1522 PEC <sup>1)</sup>	$\lambda_2^2)$	GE: N=2003 PEC	$\lambda_2$	AM: N=1522 PEC	$\lambda_2$	GE: N=2003 PEC	$\lambda_2$	
1	Externe Kontrollorientierung	14	.94	1.02	.86	1.10	9	.90	.97	.94	.93
2	Mißtrauen	8	.86	.92	.72	1.14	6	.86	.98	.80	.90
3	Depression	10	.96	.99	.93	1.04	7	.97	.98	.98	.90
4	Arbeitszufriedenheit	9	.90	.97	.83	1.11	8	.92	.95	.80	.96
5	Politische Partizipation	10	.92	1.01	.83	1.17	7	.95	.99	.92	.96
6	Optimismus	11	.94	.97	.89	1.08	7	.94	.99	.79	.89
8	Politische Desillusionierung	9	.96	.96	.93	1.04	8	.96	.99	.95	.94
9	Zukunftsperepektiven	12	.96	.96	.90	1.03	10	.92	.98	.94	.96
10	Ökonomische Selbstbestimmung	5	.88	.93	.90	.94	5	.90	.98	.88	.94

1) Anteil der durch ein einfaktorielles Modell aufgeklärten Varianz (Korrelation)

2) Eigenwerte des 2. Faktors

Eigenwerte für den 2. Faktor und die Anteile der dem jeweils 1. Faktor zuzuschreibenden Korrelationen indizieren, daß die USAM-Skalen in der deutschen Stichprobe den Skalenkriterien eher entsprechen als das für die USA-Skalen der Fall ist.

Durch die Eliminierung der Items, die am meisten gegen das Kriterium der Eindimensionalität verstießen, entsprechen die USAM-Skalen auch in der amerikanischen Stichprobe den Skalenkriterien besser als die nicht modifizierten USA-Skalen. Allerdings gibt es hier auch den Fall (z.B. bei der USAM-Skala 2), daß die Reduktion der Itemzahl den Anteil der dem 1. Faktor zuzuschreibenden Interitem-Korrelationen verringert. Die Eindimensionalität einer Skala wird also nicht zwangsläufig durch Eliminierung von Items verbessert.

Der rechte Bereich der Tabelle 1 gibt also Auskunft darüber, daß die USAM-Skalen in beiden Stichproben die statistischen Kriterien für Eindimensionalität erfüllen, was als Indikator dafür gewertet wird, daß jeder dieser Skalen eine dominante Dimension des allgemeinen sozialen Lebensgefühls zugrunde liegt. Die Frage der Relation zwischen diesen Dimensionen innerhalb der amerikanischen und der deutschen Stichprobe wird durch die Angaben in Tabelle 2 beantwortet. Dort sind in der oberen rechten Dreiecksmatrix die Korrelationen der deutschen und in der linken unteren Dreiecksmatrix die Korrelationen der amerikanischen Stichprobe für die USAM-Skalen aufgeführt. Die Korrelation von 0.79 zwischen den einander entsprechenden Tabellenfeldern reflektiert einen akzeptablen Grad an Proportionalität bzw. Entsprechung zwischen den beiden Dreiecken der Matrix und ist als Indikator für eine weitgehend ähnliche Korrelationsstruktur der USAM-Skalen in beiden Stichproben anzusehen. Die Korrelationen für die USA-Skalen werden hier nicht gesondert aufgeführt, da sie einerseits ein ähnliches Beziehungsmuster aufweisen wie die USAM-Skalen, andererseits zum Vergleich zwischen den beiden Stichproben kaum herangezogen werden.

Tabelle 2: Korrelationen der USAM-Skalen. Die Angaben für die amerikanische Stichprobe (AM: N=1522) finden sich unterhalb, die Angaben für die deutsche Stichprobe (GE: N=2003) oberhalb der Diagonalen.

		GE-Stichprobe							
		SLFS 1	SLFS 2	SLFS 3	SLFS 5 <sup>1)</sup>	SLFS 6	SLFS 8	SLFS 9	SLFS 10
	SLFS 1		.25	.30	.13	.32	.26	.37	.12
	SLFS 2	.41		.23	.20	.20	.36	.35	.16
	SLFS 3	.37	.25		.14	.42	.22	.28	.17
AM-	SLFS 5 <sup>1)</sup>	.15	.26	.16		.17	.39	.37	.27
Stichprobe	SLFS 6	.44	.27	.43	.24		.23	.30	.25
	SLFS 8	.41	.45	.25	.37	.25		.69	.36
	SLFS 9	.49	.51	.29	.36	.38	.70		.41
	SLFS10	.12	.16	.17	.20	.25	.15	.20	

Korrelation zwischen den GE und AM Korrelationskoeffizienten:  $r=.79$

1) Da SLF-Skala 4 nur für eine Subgruppe der Population (die zum Zeitpunkt der Befragung erwerbstätigen Personen) zutrifft, wird diese Skala bei der Berechnung der Interskalenkorrelationen nicht berücksichtigt.

#### SLF-Skalen der deutschen Stichprobe

Bei dem bisher beschriebenen Verfahren waren die an der amerikanischen Stichprobe entwickelten Skalen der Ausgangspunkt der Analyse. Nachdem für diese Eindimensionalität in beiden Stichproben erreicht war, blieb die Frage zu klären, welche Dimensionen sich in der deutschen Stichprobe ergeben würden, wenn die Gesamtzahl der Items die Basis der Analyse darstellen würden. Zur Klärung dieser Frage wurde der Itempool um die Items der SLF-Skala 4 (Arbeitszufriedenheit; diese Skala richtet sich nur an Erwerbstätige) sowie um die Items der beiden Skalen zur Messung der Antworttendenzen (insgesamt 24 Items) reduziert, so daß 71 Items übrigblieben.

Diese wurden einer Faktorenanalyse (Hauptachsenverfahren) unterworfen, die nach der Varimax-Rotation acht interpretierbare Faktoren ergab. Für jeden dieser acht Faktoren wurden jeweils die Items ausgewählt, die eine Ladung von  $> 0.20$  aufwiesen. Jede dieser so gewonnenen Itembatterien wurde dann

einer gesonderten Prüfung auf Eindimensionalität unterzogen, was wiederum mittels Faktorenanalyse erfolgte. Zufriedenstellende Eindimensionalität wurde dann angenommen, wenn lediglich ein Eigenwert  $> 1$  vorhanden war. Bei den Itembatterien, die dieses Kriterium nicht erfüllten, wurden die Items mit hohen numerischen Werten für die Residuen erster Ordnung eliminiert. Die so gewonnenen Skalen überschneiden sich weitgehend mit den Skalen, die an der amerikanischen Stichprobe gewonnen wurden, sind jedoch nicht identisch mit diesen. Die Benennung dieser Skalen richtet sich nach der nächstverwandten USAM-Skala (indiziert durch den Korrelations-Koeffizienten).

Die Kennwerte für die deutsche Fassung der Skalen, kurz: BRD-Skalen genannt, sind in Tabelle 3 aufgeführt. Im Vergleich zu den amerikanischen Originalen sind die deutschen Skalen durchweg (mit Ausnahme der Skala 8) kürzer, d.h. sie enthalten weniger Items. Für alle BRD-Skalen besteht in der deutschen Stichprobe Eindimensionalität, d.h. es existiert pro Skala nur ein Faktor mit einem Eigenwert  $> 1$ , und dieser Faktor erklärt bei allen Skalen mehr als 90% der Interitem-Korrelationen.

In der amerikanischen Stichprobe haben 2 der BRD-Skalen mehr als einen Eigenwert  $> 1.0$ . Während die an der amerikanischen Stichprobe entwickelten Skalen (mit Ausnahme der Skala 10) alle modifiziert werden mußten, damit das Eigenwert-Kriterium der Eindimensionalität in der deutschen Stichprobe gewährleistet war, ist bei der deutschen Fassung der Skalen eine Modifikation nur bei zwei Skalen (6 und 8) notwendig, um das formale Kriterium für Eindimensionalität in der amerikanischen Stichprobe zu gewährleisten. Diese Skalen sind mit dem Kürzel BRDM versehen (ihre Kennwerte befinden sich im rechten Teil der Tabelle 3).

Auch für die an der deutschen Stichprobe entwickelten Skalen wurden die Interkorrelationen berechnet. Diese sind in Tabelle 4 aufgeführt. Der Koeffizient von  $r=0.87$  indiziert ein hohes Ausmaß an Proportionalität der Interkalenrelationen in den beiden Stichproben. Mit anderen Worten: die Dimensionen des allgemeinen sozialen Lebensgefühls weisen in beiden Stichproben eine weitgehende Vergleichbarkeit hinsichtlich ihrer Korrelationsstruktur auf.

**Tabelle 3:** Skaleneckwerte der an der deutschen Stichprobe entwickelten BRD-Skalen und deren Modifikation.  
BRDM-Skalen nach Anwendung in der deutschen Stichprobe

Skalen- Nr.	Skalen- name	Anzahl Items	BRD-Skalen				BRDM <sup>3)</sup> -Skalen				
			AM: N=1522 PEC <sup>1)</sup>	$\lambda_2^{2)}$	GE: N=2003 PEC	$\lambda_2$	Anzahl Items	GE: N=2003 PEC	$\lambda_2$	AM: N=1522 PEC	$\lambda_2$
1	Externe Kontrollorientierung	9	.96	.91	.95	.97	9	.95	.97	.96	.91
2	Mißtrauen	6	.90	.87	.91	.95	6	.91	.95	.90	.87
3	Depression	9	.97	.91	.94	.92	9	.94	.92	.97	.91
4	Arbeitszufriedenheit	8	.80	.96	.92	.95	8	.92	.92	.80	.96
5	Politische Partizipation	5	.93	.90	.91	.93	5	.91	.93	.93	.90
6	Optimismus	10	.89	1.09	.95	.99	5	.95	.92	.90	.99
8	Politische Desillusionierung	10	.90	1.04	.98	.99	6	.98	.90	.97	.98
9	Zukunftsperepektiven	10	.95	.97	.94	.98	10	.94	.98	.95	.97
10	Ökonomische Selbstbestimmung	5	.88	.94	.90	.98	5	.90	.98	.88	.94

1) Anteil der durch ein einfaktorielles Modell aufgeklärten Varianz (Korrelation)

2) Eigenwerte des 2. Faktors

3) Bei den BRD-Skalen 1,2,3,5,9 und 10 war keine Modifikation erforderlich

Tabelle 4: Korrelationen der BRDM-Skalen. Die Angaben für die amerikanische Stichprobe (AM: N=1522) finden sich links unterhalb, die Angaben für die deutsche Stichprobe (GE: N=2003) rechts oberhalb der Diagonalen.

		GE-Stichprobe							
		SLFS 1	SLFS 2	SLFS 3	SLFS <sub>1)</sub> 5	SLFS 6	SLFS 8	SLFS 9	SLFS 10
AM- Stichprobe	SLFS 1		.46	.34	.24	.49	.44	.64	.22
	SLFS 2	.56		.22	.19	.18	.31	.35	.12
	SLFS 3	.39	.26		.10	.49	.21	.31	.18
	SLFS 5 <sup>1)</sup>	.25	.19	.13		.14	.53	.33	.25
	SLFS 6	.50	.25	.50	.18		.21	.34	.31
	SLFS 8	.50	.41	.24	.49	.24		.54	.42
	SLFS 9	.72	.54	.35	.31	.43	.56		.41
	SLFS10	.16	.14	.18	.15	.31	.29	.19	

Korrelation zwischen den GE und AM Korrelationskoeffizienten:  $r=.87$

1) Da SLF-Skala 4 nur für eine Subgruppe der Population (die zum Zeitpunkt der Befragung erwerbstätigen Personen) zutrifft, wird diese Skala bei der Berechnung der Interskalenkorrelationen nicht berücksichtigt.

Die Kombination der an der amerikanischen und der deutschen Stichprobe entwickelten SLF-Skalen

Um eine Dezentrierung der Skalen, d.h. eine Loslösung von ihrem Entstehungskontext zu erreichen, wurden die COMB-Skalen so gebildet, daß sie jeweils eine Kombination der modifizierten Skalen darstellen. Diese Kombination erfolgte auf der Basis der Korrelationen der modifizierten USAM-Skalen und der BRD bzw. BRDM-Skalen (da in der deutschen Version nur zwei Skalen modifiziert werden mußten) innerhalb der beiden Stichproben. Diese sind den Tabellen 5a und 5b abgebildet: In Tabelle 5a finden sich die Korrelationen für die deutsche, in Tabelle 5b die Korrelationen für die amerikanische Stichprobe. Wenn die modifizierten Skalen in beiden Stichproben dasselbe messen, dann müßte man erwarten, daß die jeweils identischen Skalen die höchsten Korrelationswerte in der Matrix aufweisen. Idealerweise würde man für die

Tabelle 5a: Korrelationen zwischen den USAM-Skalen und den BRD- bzw. BRDM-Skalen für die deutsche (GE: N=2003) Stichprobe

		BRD-/BRDM-Skalen								
		SLFS 1	SLFS 2	SLFS 3	SLFS 4	SLFS 5	SLFS 6	SLFS 8	SLFS 9	SLFS 10
USAM- Skalen	SLFS 1	.79	.27	.33	.20	.17	.46	.31	.41	.12
	SLFS 2	.44	.94	.24	.19	.19	.20	.31	.36	.16
	SLFS 3	.29	.20	.91	.26	.10	.45	.19	.27	.17
	SLFS 4	.22	.16	.28	1.00	.17	.31	.24	.24	.18
	SLFS 5 <sup>1)</sup>	.20	.18	.14	.18	.81	.17	.47	.36	.27
	SLFS 6	.35	.16	.47	.32	.11	.70	.19	.32	.25
	SLFS 8	.42	.33	.24	.23	.41	.25	.83	.61	.36
	SLFS 9	.57	.33	.30	.21	.35	.33	.60	.91	.41
	SLFS10	.22	.12	.18	.18	.25	.31	.42	.41	1.00

Tabelle 5b: Korrelationen zwischen den USAM-Skalen und den BRD- bzw. BRDM-Skalen für die amerikanische (AM: N=1522) Stichprobe

		BRD-/BRDM-Skalen								
		SLFS 1	SLFS 2	SLFS 3	SLFS 4	SLFS 5	SLFS 6	SLFS 8	SLFS 9	SLFS 10
USAM- Skalen	SLFS 1	.84	.43	.40	.25	.17	.52	.41	.57	.12
	SLFS 2	.55	.96	.26	.21	.19	.26	.40	.53	.16
	SLFS 3	.35	.25	.92	.27	.13	.49	.22	.32	.17
	SLFS 4	.24	.21	.28	1.00	.17	.18	.21	.21	.13
	SLFS 5 <sup>1)</sup>	.25	.26	.17	.21	.88	.21	.46	.35	.19
	SLFS 6	.42	.26	.46	.32	.20	.72	.25	.42	.25
	SLFS 8	.51	.46	.25	.22	.35	.25	.81	.66	.15
	SLFS 9	.61	.52	.31	.20	.32	.38	.60	.91	.20
	SLFS10	.16	.14	.18	.13	.16	.34	.29	.19	1.00

1) Da SLF-Skala 4 nur für eine Subgruppe der Population (die zum Zeitpunkt der Befragung erwerbstätigen Personen) zutrifft, wird diese Skala bei der Berechnung der Interskalenkorrelationen nicht berücksichtigt.



Diagonalelemente Werte von 1.0 annehmen. Wie jedoch ein kurzer Blick auf die Tabellen 5a und 5b offenbart, geht es hier so ideal nicht zu. In beiden Stichproben liegen die Werte der Diagonalelemente zwischen 0.72 (0.70) und 0.96 (0.94). Ideal ist dieses Ergebnis nicht, aber wir können damit leben, denn immerhin korrelieren 8 (bzw. 7) der 9 Skalen mit einem Wert oberhalb von 0.80, was zwar keine Identität, aber doch weitgehende Vergleichbarkeit an Stichproben aus unterschiedlichen kulturellen Kontexten entwickelten (USAM- und BRD-) Skalen indiziert.

Die durch Kombination der Items dieser Skalen erhaltenen Itembatterien wurden nochmals auf Eindimensionalität in beiden Stichproben überprüft und – wenn nötig – durch Eliminierung derjenigen Items mit hohen Residuen erster Ordnung modifiziert. Diese Skalen stellen die COMB-Skalen dar. Die Itemformulierungen und –kennwerte dieser Skalen sind bei Krebs & Schuessler (1987) dokumentiert.

Alle COMB-Skalen erfüllen in beiden Stichproben die oben genannten Kriterien für Eindimensionalität, d.h. daß es für die zu einer Skala gehörenden Items jeweils nur einen dominanten Faktor gibt. Die Kennwerte der COMB-Skalen sind in Tabelle 6 aufgeführt. Die COMB-Skalen kommen der Vorstellung von – für die Verwendung bei deutschen und amerikanischen Befragten – standardisierten Meßinstrumenten am nächsten. Die Korrelationsstrukturen der COMB-Skalen weisen mit einem Koeffizienten von  $r=0.83$  eine durchaus akzeptable Übereinstimmung zwischen den beiden Stichproben auf. Die Ähnlichkeit der Zusammenhänge zwischen den Dimensionen des allgemeinen sozialen Lebensgefühls läßt eine weitgehende Vergleichbarkeit dieser Dimensionen in den beiden Stichproben vermuten. Das bedeutet jedoch nicht, daß die Itemkennwerte homogen sind. Tatsächlich ist es üblicherweise so, daß sich die Itemkennwerte wie z.B. Itemmittelwerte in Gruppen mit unterschiedlichen demographischen, sozialen und linguistischen Merkmalen unterscheiden, selbst wenn die Skalenkennwerte wie z.B. Mittelwerte und Standardabweichungen einander ähnlich sind. Zur Demonstration derartiger skaleninterner Itemunterschiede und ihrer Bedeutung in vergleichenden Untersuchungen werden hier die Items der COMB-Skala 9 (Zukunftsperspektiven) angeführt. Die Itemmittelwerte und die (biserialen) Item-Gesamt-Korrelationen sind den Maßen für Itemschwierigkeit und für Item-

diskriminierung (sowie den Item-"intercepts" und den Itemsteigungen) vergleichbar, auf die üblicherweise in der testtheoretischen Literatur Bezug genommen wird (Lord & Novick, 1968; Andersen, 1980).

Tabelle 6: Skalenkennwerte der COMB-Skalen in der deutschen und der amerikanischen Stichprobe

Skalen-Nr.	Skalen-name	Anzahl Items	AM: N=1522		GE: N=2003	
			PEC <sup>1)</sup>	$\lambda_2$ <sup>2)</sup>	PEC	$\lambda_2$
1	Externe Kontrollorientierung	12	.92	.99	.90	.98
2 <sup>a</sup>	Mißtrauen	6	.90	.87	.91	.95
3 <sup>b</sup>	Depression	9	.97	.91	.94	.92
4	Arbeitszufriedenheit	8	.80	.96	.92	.95
5 <sup>c</sup>	Politische Partizipation	7	.92	.96	.95	.99
6	Optimismus	8	.94	.98	.95	.99
8	Politische Desillusionierung	8	.94	.96	.90	.99
9	Zukunftsperepektiven	12	.94	.96	.90	.99
10 <sup>d</sup>	ökonomische Selbstbestimmung	5	.88	.94	.90	.98

- a BRDM2
- b BRDM3
- c USAM5
- d USAM10=BRDM10

- 1) Anteil der durch ein einfaktorielles Modell aufgeklärten Varianz (Korrelation)
- 2) Eigenwerte des 2. Faktors

Item- versus Skalenkennwerte

In Tabelle 7 sind die Skalen- und Itemkennwerte der COMB-Skala 9 aufgeführt. Zunächst ist festzustellen, daß die deutschen und die amerikanischen Befragten ähnliche Mittelwerte bei dieser Skala erreichen (6.45 zu 6.68). Beide Stichproben ähneln sich weiterhin in der Struktur ihrer Interitem-Korrelationen, was in den auf der Korrelationsmatrix der Items basierenden Maßen (Eigenwerte, Anteil der durch den 1. Faktor erklärten Korrelationen, Cronbach's Alpha) zum Ausdruck kommt. Andererseits bestehen jedoch erhebliche Unterschiede zwischen den Stichproben auf Itemebene. Diese Unterschiede

Tabelle 7: COMB-Skala 9: Zukunftsperspektiven;  
Itemformulierungen und Kennwerte sowie Skalenskennwerte für die  
deutsche (GE: N=2003) und die amerikanische (AM: N=1522) Stich-  
probe

	Item- mittelwert		Item-Gesamt- Korrelation	
	GE	AM	GE	AM
1. In den nächsten Jahren wird es noch mehr Arbeitslose geben.	.82	.75	.34	.30
2. Obwohl sich ständig sehr viel ändert, weiß man im großen und ganzen doch, was man zu erwarten hat.	.28	.55	.21	.29
3. Die Zukunft dieses Landes ist ungewiß.	.59	.72	.49	.49
4. Die Zukunft sieht sehr trübe aus.	.52	.46	.56	.55
5. Meiner Meinung nach ist dieses Land kaputt.	.16	.43	.33	.38
6. Die Zukunft ist zu ungewiß, als daß man weit im voraus planen könnte.	.63	.50	.46	.45
7. Wir verlieren langsam unsere persönliche Freiheit, weil sich die Behörden immer weiter ausdehnen.	.54	.65	.37	.42
8. Das Leben des Normalbürgers wird schlechter, nicht besser.	.59	.53	.48	.45
9. So, wie die Zukunft aussieht, ist es unfair, Kinder in die Welt zu setzen.	.29	.35	.34	.37
10. Viele Grundsätze unserer Eltern gehen den Bach runter.	.72	.77	.24	.36
11. Ich habe zur Zeit wenig Vertrauen zur Regierung.	.62	.58	.44	.45
12. Ich finde, daß die Zukunft recht gut aussieht.	.70	.40	.45	.38
	Mittelwert	6.45	6.68	
	Standardabweichung	2.88	3.06	
	Eigenwert des 2. Faktors	0.99	0.96	
	PEC	0.90	0.94	
	Alpha	0.75	0.77	
	Durchschnittlicher Itemmittelwert	0.54	0.56	

Zehn der zwölf Items beziehen sich auf gegenwärtige oder antizipierte Trends in der Entwicklung der sozialen Lebensqualität. Acht dieser Items beschreiben eine Verschlechterung und zwei (2, 12) indizieren eine Verbesserung oder zumindest keine Verschlechterung in der Qualität der sozialen Lebensbedingungen. Zwei (5, 11) der Items haben keinen direkten Bezug zur Einschätzung gegenwärtiger oder zukünftiger Trends der Lebensqualität. Hohe Summenwerte ergeben sich dann, wenn die positiv formulierten Items abgelehnt werden und den negativen Items zugestimmt wird, was eine düstere Erwartung zukünftiger Entwicklungen zum Ausdruck bringt. Niedrige Summenwerte ergeben sich dann, wenn umgekehrt die negativ formulierten Items abgelehnt werden und den positiv formulierten Items zugestimmt wird, wodurch sich eine positive, eher optimistische Zukunftserwartung ausdrückt. Der durchschnittliche Item-Mittelwert liegt bei 0.54 bzw. 0.56, was besagt, daß etwa 50% der Items in beiden Stichproben eine Antwort hervorrufen, die düstere Zukunftsperspektiven bekräftigt.

werfen die Frage nach Unterschieden im allgemeinen sozialen Lebensgefühl auf, die bei der Aggregation der Itemscores verschwinden. Derartige Unterschiede können einerseits substantielle Verschiedenheiten im allgemeinen sozialen Lebensgefühl reflektieren. Sie können aber auch Unterschiede im Antwortverhalten, Unterschiede in den demographischen Merkmalen, Zeitunterschiede sowie durch Ungenauigkeiten bei der Übersetzung verursachte Bedeutungsunterschiede spiegeln. Unabhängig davon, welche(r) dieser Faktoren zur Erklärung der Differenzen in den Itemkennwerten herangezogen wird, ist es wichtig, diesen Differenzen bei der Untersuchung sozialer und kultureller Unterschiede zwischen Gruppen nachzugehen. In dem Beispiel in Tabelle 7 ist also einerseits auf die ähnlichen Skalenmittelwerte und andererseits auf die Differenz von fast 0.30 bei den Mittelwerten der Items 2 und 12 zu achten. Diese beiden Items indizieren eine Verbesserung oder zumindest keine Verschlechterung in der Qualität der sozialen Lebensbedingungen. Während die amerikanischen Befragten eine höhere Erwartungsunsicherheit in einer sich ständig ändernden Welt zum Ausdruck bringen als die deutschen Befragten, lehnen diese eine pauschal positive Zukunftssicht im Vergleich zu den Amerikanern vehement ab. Die Mittelwertunterschiede bei den Items 2 und 12 lenken die Aufmerksamkeit darauf, daß die Amerikaner bei Items, die die Zukunft allgemein ansprechen, niedrigere Werte - also eine positivere Sichtweise - haben als die Deutschen, daß sie dagegen mehr Pessimismus hinsichtlich zukünftiger Entwicklungen zum Ausdruck bringen, wenn es um konkrete Themen wie das Allgemeinwohl, ihr Land und die Tradition ihres Landes geht. Sowohl die Deutschen als auch die Amerikaner haben also insgesamt gesehen eine eher negativ gefärbte Zukunftssicht, die sich bei den deutschen Befragten eher an der Einschätzung der Zukunft im allgemeinen, bei den amerikanischen Befragten eher an einer auf konkrete Bereiche bezogenen Einschätzung festmacht. Die Einschätzung zukünftiger Lebensqualität wird in beiden Stichproben an unterschiedlichen Schwerpunkten festgemacht, was auf die Existenz unterschiedlicher Deutungsmuster in den beiden Kulturbereichen hinweist.

#### Probabilistische Skalierungsmodelle

Üblicherweise wird ein beobachteter Wert als Schätzwert eines "wahren" Wertes angesehen, der seinerseits als der statistische Erwartungswert des beobachteten Wertes definiert ist. Die Differenz zwischen beobachtetem und

wahrem Wert wird dem Meßfehler zugeschrieben, wobei die Korrelation zwischen beobachtetem und wahrem Wert als Index für den Meßfehler herangezogen wird. In der neueren meßtheoretischen Diskussion - "item response"-Theorie - richtet sich das Interesse weniger auf die Relationen zwischen beobachtetem und wahrem Wert als vielmehr auf die Beziehungen zwischen der Beantwortung von Items und einer latenten Variablen. Diese Relation basiert entsprechend der "item response"-Theorie auf einem Wahrscheinlichkeitsmodell, dessen Anpassung an die Daten (goodness of fit) getestet werden kann. Wird das Modell nicht zurückgewiesen, so können Personen auf einem latenten Kontinuum niedrige oder hohe Positionen zugewiesen werden.

Zur Illustration werden hier drei probabilistische Modelle vorgestellt, die kurzerhand als das Muthén-Modell, das Bartholomew-Modell und das Rasch-Modell bezeichnet werden. Das Muthén-Modell (Muthén, 1978) geht davon aus, daß die Wahrscheinlichkeit einer positiven Antwort auf ein Item als Normalogive darstellbar ist und daß die latente Variable in der Population normal verteilt ist. Das Modell enthält gleichzeitig die Annahme, daß die Wahrscheinlichkeit von  $k$  positiven Antworten das Produkt aus den individuellen Wahrscheinlichkeiten aller  $k$  positiven Antworten ist. Letzteres entspricht dem Postulat der lokalen Unabhängigkeit, d.h. der Annahme, daß ohne die Existenz einer latenten Variable keine Interitem-Korrelationen vorhanden wären. Im Muthén-Modell wird die Popularität (Schwierigkeit) eines Items als Schwelle  $\tau_j$ , der Zusammenhang zwischen Item und latenter Variable als Faktorenladung  $\lambda_j$  dargestellt. Die  $\tau_j$  dienen zur Kategorisierung der Items nach Popularität, die  $\lambda_j$ s zur Kategorisierung der Items entsprechend ihrer Abhängigkeit von der latenten Variablen.

Tabelle 9, Spalte 3, enthält die Lösung des Muthén-Modells für die COMB-Skala 2 (Mißtrauen), deren Item- und Skalenkennwerte in Tabelle 8 aufgeführt sind. Die Itempopularitäten ( $\tau_j$ ) sind in den beiden Stichproben sehr unterschiedlich. Dasselbe gilt für die Relationen der Items zur latenten Variablen ( $\lambda_j$ ). Die Items scheinen für die deutschen Befragten leichter (populärer) zu sein, was bedeutet, daß die Normalogiven (item characteristic curves ICCs) der Items auf dem latenten Kontinuum weiter links liegen und daß sie insgesamt flacher verlaufen als die ICC's der amerikanischen Befragten. Das Modell bietet in beiden Stichproben keine besonders gute Anpassung an die

Tabelle 8: COMB-Skala 2: Mißtrauen;  
Itemformulierungen und -kennwerte sowie Skalenkennwerte für die deutsche (GE: N=2003) und die amerikanische (AM: N=1522) Stichprobe

	Item- mittelwert		Item-Gesamt- Korrelation	
	GE	US	GE	AM
1. Es ist schwer herauszufinden, wem man heute noch trauen kann.	.74	.64	.43	.51
2. Genau genommen gibt es nur wenige Menschen auf der Welt, denen man trauen kann.	.63	.52	.48	.57
3. Fremden kann man im allgemeinen trauen.	.81	.65	.35	.37
4. Den meisten Menschen kann man trauen.	.58	.35	.51	.50
5. Die meisten Menschen kümmern sich im Grunde nicht darum, was aus ihren Mitmenschen wird.	.79	.48	.36	.49
6. Viele Leute sind nur deshalb freundlich, weil sie etwas von einem wollen.	.66	.47	.35	.43
Mittelwert		4.20	3.11	
Standardabweichung		1.68	1.94	
Eigenwert des 2. Faktors		0.95	0.87	
PEC		0.91	0.98	
Alpha		0.68	0.74	
Durchschnittlicher Itemmittelwert		0.70	0.52	

Diese Skala enthält vier Items mit negativem und zwei Items mit positivem Bedeutungsgehalt: Die vier negativen Items stellen die Vertrauenswürdigkeit anderer Menschen in Frage, die zwei positiven Items beinhalten die Überzeugung, daß man anderen Menschen durchaus vertrauen kann. Entsprechend der Polung der Items geht es hier nicht um eine Dimension mit den Endpunkten 'Vertrauen' einerseits und 'Mißtrauen' andererseits, sondern es geht um das Ausmaß der Überzeugung, inwiefern man die Mitmenschen als vertrauenswürdig ansehen kann. Die numerischen Werte dieser Skala reichen von 0 bis 6; hohe Werte indizieren starke Zweifel an der Vertrauenswürdigkeit, während Personen mit niedrigen Werten die Vertrauenswürdigkeit anderer weniger stark in Frage stellen.

Im Vergleich zwischen den beiden Stichproben heben sich die Personen der deutschen Stichprobe durch ein negatives Menschenbild gegenüber den Personen der amerikanischen Stichprobe ab. Die Item-Mittelwerte in der deutschen Stichprobe liegen durchgängig und konsistent über denen der amerikanischen Stichprobe. Der Gesamtmittelwert der Skala liegt in der deutschen Stichprobe deutlich höher als in der amerikanischen Stichprobe, und der durchschnittliche Item-Mittelwert indiziert, daß die Personen der deutschen Stichprobe sich bei 70% der Items für diejenige Antwort entscheiden, die die Vertrauenswürdigkeit der Mitmenschen in Frage stellt, während die Personen der amerikanischen Stichprobe dieses Antwortverhalten nur bei 52% der Items zeigen.

Tabelle 9: Ergebnisse des Rasch-, des Bartholomew (logit-probit)- und des Muthen-Modells für die deutsche (GE: N=2003) und die amerikanische (AM: N=1522) Stichprobe bei der COMB-Skala 2

Modell:	Rasch		Bartholomew logit-probit		Muthen	
Beobachtete Häufigkeiten	$2(p^2-p)$		$2^p$		$p(p+1)/2$	
Item- Nr.	GE	AM	GE	AM	GE	AM
1	1.15	1.55	1.54	0.98	-0.66	-0.36
2	0.74	1.00	0.83	0.15	-0.33	-0.05
3	1.52	1.65	1.83	0.79	-0.87	-0.39
4	0.63	0.53	0.59	-0.10	-0.23	0.04
5	1.42	0.89	1.65	-0.97	-0.82	0.40
6	0.85	0.84	0.83	-0.16	-0.43	0.07
Item- oder Summenwertgruppe	$s_i$		$\alpha_i$		$\lambda_i$	
1	0.42	0.41	0.93	1.22	0.71	0.81
2	0.68	0.67	1.07	1.45	0.76	0.85
3	1.00	1.00	0.74	0.65	0.62	0.85
4	1.46	1.48	1.20	1.02	0.79	0.72
5	2.41	2.46	0.66	0.92	0.59	0.79
6			0.58	0.74	0.55	0.64
$L_0^2$	1977	1900	1455	1663	585	998
DF	44	44	57	57	15	15
$L_1^2$	87	115	85	112	64	76
DF	20	20	43	41	9	9
PRE	0.90	0.87	0.92	0.92	0.82	0.87

Daten, wenn man das Verhältnis aus der statistischen Prüfgröße  $L^2$  und den entsprechenden Freiheitsgraden als Indikator für den Grad der Anpassung ansieht. Durch die Annahme einer latenten Variablen wird jedoch eine beträchtliche Reduktion der unerklärten Varianz erreicht, die in dem PRE- (proportionate reduction of error) Maß von 0.82 für die deutsche und 0.87 für die amerikanische Stichprobe zum Ausdruck kommt. Dieses Maß ist so zu interpretieren, daß die Residualvarianz durch Einführung einer latenten Variablen um 82% bzw. 84% reduziert wird.

Das Bartholomew-(logit-probit)Modell, dessen Ergebnisse für die COMB-Skala 2 in Spalte 2 der Tabelle 9 aufgeführt sind, unterscheidet sich nur geringfügig vom Muthén-Modell. Ebenso wie dort wird beim logit-probit-Modell angenommen, daß die latente Variable in der Population normal verteilt ist. Der Unterschied zum Muthén-Modell besteht darin, daß die Antwortwahrscheinlichkeiten nicht einer Normalogive, sondern einer logit-Ogive folgen. Das logit-probit-Modell beschreibt die Itempopularitäten als Achsenabschnitte ("intercepts"), symbolisiert durch  $\alpha_{i0}$  und den Grad der Abhängigkeit des Items von der latenten Variablen als Steigung ("slope"), symbolisiert durch  $\alpha_i$ . Je höher der numerische Wert von  $\alpha_{i0}$ , um so populärer (leichter) ist das Item, je höher der Wert von  $\alpha_i$ , um so stärker ist der Zusammenhang zwischen dem Item und der latenten Variablen. Anhand der "intercepts" können Items nach dem Ausmaß ihrer durchschnittlichen Popularität, anhand der Steigungen nach der Stärke des Zusammenhangs mit der latenten Variablen geordnet werden.

Die Ergebnisse des logit-probit-Modells entsprechen im wesentlichen denen des Muthén-Modells: die Itemogiven liegen in der deutschen Stichprobe weiter links auf dem latenten Kontinuum und steigen flacher von links nach rechts an als in der amerikanischen Stichprobe. Bei einer Modellspezifikation, die eine latente Variable (im Gegensatz zu keiner latenten Variablen) zuläßt, wird im logit-probit-Modell eine Reduktion der Restvarianz um 93% in der deutschen und um 92% in der amerikanischen Stichprobe erreicht. Die statistischen Prüfgrößen für diese beiden Modelle (aus denen sich auch der PRE-Wert errechnet) sind in Tabelle 9 als  $L_0^2$  (Unabhängigkeitsmodell, keine latente Variable) und  $L_1^2$  (einfaktorielles Modell) aufgeführt.

Das Rasch-Modell entspricht weitgehend dem logit-probit-Modell. Unterschiede bestehen in zweierlei Hinsicht: a) in der Annahme, daß alle logistischen Itemogiven dieselbe Steigung haben und b) darin, daß keine Annahme über die Verteilung der latenten Variablen in der Population gemacht wird. In dieser letztgenannten Hinsicht ist das Modell verteilungsfrei. Wenn das Modell zu einer guten Anpassung an die Daten führt, ist das als Anhaltspunkt dafür zu sehen, daß die Antwortunterschiede bei einzelnen Items ausschließlich durch unterschiedliche "latent-trait"-Werte für die verschiedenen Summenwertgruppen beeinflusst werden.



Die Lösung des Rasch-Modells findet sich in Spalte 1 der Tabelle 9. Schätzwerte für Itempopularitäten sind als  $I_j$  symbolisiert. Niedrige Werte signalisieren unpopuläre (schwere), hohe Werte signalisieren populäre (leichte) Items. Das Rasch-Modell liefert auch Schätzwerte für die Platzierung von Personen mit denselben Summenwerten auf der latenten Variablen. Hohe Werte indizieren ein Mehr, niedrige Werte ein Weniger des latenten Merkmales. Das Symbol für die Summenwertgruppen ist  $S_i$ . Die  $I_j$ -Werte entsprechen den Itemmittelwerten und die  $S_i$ -Werte haben dieselbe Rangordnung wie die Summenrohwerte. Auch das Rasch-Modell liefert keine gute Anpassung an die Daten, erreicht jedoch bei Annahme eines einfaktoriellen Modells eine Reduktion der Restvarianz von 90% in der deutschen und 87% in der amerikanischen Stichprobe, was impliziert, daß die Antworten durch die Existenz einer dominanten latenten Variablen beeinflußt werden.

Da alle drei Modelle - Bartholomew, Muthén und Rasch - zu keiner befriedigenden Anpassung an die Daten der COMB-Skala 2 führen, sollte man wahrscheinlich folgern, daß mit dieser Skala mehr als ein latentes Merkmal gemessen wird. Damit soll jedoch nicht angedeutet werden, daß mehr als eine Dimension des allgemeinen sozialen Lebensgefühls durch die COMB-Skala 2 erfaßt wird. Im Gegenteil: es ist durchaus möglich, daß die Antworten durch mechanisches Antwortverhalten wie z.B. die Zustimmungstendenz oder die Tendenz zu willkürlichen Antworten verzerrt werden. Solche latenten Tendenzen können die Beziehung zwischen den Itemantworten und dem latenten Lebensgefühl stören und Wertemuster produzieren, die dem einfaktoriellen Modell, das getestet wird, zuwiderlaufen. Das Ziel weiterer Analysen sollte u.E. der Versuch sein, Mittel und Wege zur Auspartialisierung von Antwortverzerrungen zu finden, da die so bereinigten Antworten vermutlich eher den Anforderungen strikter Eindimensionalität - wie sie von den probabilistischen Modellen angenommen wird - genügen.

### Schlußbemerkung

Die hier beschriebene Konstruktion von Skalen zur zuverlässigen Messung und zum Vergleich des allgemeinen sozialen Lebensgefühls bei Deutschen und Amerikanern erfolgte in mehreren Schritten. Zunächst wurde das Verhalten der 9 an der amerikanischen Stichprobe entwickelten Skalen in der deutschen Stichprobe untersucht. Diese Skalen wurden mit dem Ziel verändert, in beiden Stichproben die statistischen Kriterien der Eindimensionalität zu erfüllen. In einem nächsten Schritt wurden - basierend auf einer multivariaten Analyse der Antworten der deutschen Befragten - alternative Skalen konstruiert. Nach der Evaluation des Verhaltens dieser, an der deutschen Stichprobe entwickelten Skalen in der amerikanischen Stichprobe, wurden die Skalen wiederum so verändert, daß die statistischen Kriterien der Eindimensionalität in beiden Stichproben erfüllt waren. Die modifizierten Skalenversionen wurden sodann miteinander kombiniert und so lange durch Reduktion von Items modifiziert, bis die Skalenkriterien in beiden Stichproben annähernd realisiert waren. Die resultierenden Itembatterien wurden COMB-Skalen genannt, was darauf hinweist, daß sie aus einer Kombination der an der amerikanischen und der deutschen Stichprobe entwickelten und modifizierten Skalen hervorgegangen sind. Diese COMB-Skalen sind als standardisiert in dem Sinne anzusehen, daß sie die statistischen Kriterien der Eindimensionalität in beiden Stichproben erfüllen.

Die Eindimensionalität einer Anzahl von Items ist jedoch noch kein Beweis dafür, daß die Items tatsächlich das messen, was wir meinen, das sie messen sollten. Ein derartiger Beweis muß von außen kommen, von unabhängigen Indikatoren jener Lebensgefühl dimension, die die jeweilige Skala zu messen beabsichtigt.

Obwohl die COMB-Skalen in beiden Stichproben weitgehend ähnliche statistische Eigenschaften und Kennwerte aufweisen und in diesem Sinne als standardisiert anzusehen sind, kann man wahrscheinlich nie ganz sicher sein, daß sie bezüglich ihrer Validität identisch sind (näheres hierzu ist bei Krebs & Schuessler, 1987 ausgeführt). Dieser Punkt wird hier abschließend erwähnt, um in Erinnerung zu rufen und sich zu vergegenwärtigen, daß die Entwicklung von Skalen zur Messung des allgemeinen sozialen Lebensgefühls ein niemals endender Prozeß ist und daß ein Urteil darüber, ob eine Skala ihren Zweck in

einer vergleichenden Studie erfüllt, erst nach mehrfach wiederholter Anwendung in vergleichenden Untersuchungen möglich ist.

### Literatur

- Andersen, E., 1980: Discrete Statistical Models with Social Science Applications. Amsterdam: North Holland Publishing Co.
- Bartholomew, D.T., 1987: Latent Variable Models and Factor Analysis. London: Charles Griffin & Co.
- Jöreskog, K.G., 1979: Basic ideas of factor and component analysis. S. 5-20 in: Jöreskog, K.G./Sörbom, D., Advances in Factor Analysis and Structural Equation Models. Cambridge, Mass: Abt Books.
- Krebs, D., Schuessler, K.F., 1987: Soziale Empfindungen. Ein interkultureller Skalenvergleich bei Deutschen und Amerikanern. Frankfurt: Campus.
- Lord, F.M./Novick, M.R., 1968: Statistical Theories of Mental Test Scores. Reading, Mass: Addison-Wesley.
- Muthén, B.: Contributions to factor analysis of dichotomous variables. Psychometrika 43:551-560.
- Rasch, G., 1966: An individualistic approach to item analysis. In: Lazarsfeld, P.F./Henry, N.W. (Hrsg.), Readings in Mathematical Social Science. Chicago: Science Research Associates.
- Schuessler, K.F., 1982: Measuring Social Life Feelings. San Francisco: Jossey Bass.

## ZUMA-Arbeitsberichte

- 80/15 Gerhard Arminger, Willibald Nagl, Karl F. Schuessler  
Methoden der Analyse zeitbezogener Daten. Vortragsskripten der ZUMA-  
Arbeitstagung vom 25.09. - 05.10.79
- 81/07 Erika Brückner, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer, Peter  
Schmidt  
Methodenbericht zum "ALLBUS 1980"
- 81/19 Manfred Küchler, Thomas P. Wilson, Don H. Zimmerman  
Integration von qualitativen und quantitativen Forschungsansätzen
- 82/03 Gerhard Arminger, Horst Busse, Manfred Küchler  
Verallgemeinerte Lineare Modelle in der empirischen Sozialforschung
- 82/08 Glenn R. Carroll  
Dynamic analysis of discrete dependent variables: A didactic essay
- 82/09 Manfred Küchler  
Zur Messung der Stabilität von Wählerpotentialen
- 82/10 Manfred Küchler  
Zur Konstanz der Recallfrage
- 82/12 Rolf Porst  
"ALLBUS 1982" - Systematische Variablenübersicht und erste Ansätze zu  
einer Kritik des Fragenprogramms
- 82/13 Peter Ph. Mohler  
SAR - Simple AND Retrieval mit dem Siemens-EDT-Textmanipulations-  
programm
- 82/14 Cornelia Krauth  
Vergleichsstudien zum "ALLBUS 1980"
- 82/21 Werner Hagstotz, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer  
Methodenbericht zum "ALLBUS 1982"
- 83/09 Bernd Wegener  
Two approaches to the analysis of judgments of prestige: Interindi-  
vidual differences and the general scale
- 83/11 Rolf Porst  
Synopsis der ALLBUS-Variablen. Die Systematik des ALLBUS-Fragen-  
programms und ihre inhaltliche Ausgestaltung im ALLBUS 1980 und  
ALLBUS 1982
- 84/01 Manfred Küchler, Peter Ph. Mohler  
Qualshop (ZUMA-Arbeitstagung zum "Datenmanagement bei qualitativen  
Erhebungsverfahren") - Sammlung von Arbeitspapieren und -berichten,  
Teil I + II
- 84/02 Bernd Wegener  
Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-  
Prestige-Skala

- 84/03 Peter Prüfer, Margrit Rexroth  
Beschreibung eines Verfahrens zur Bewertung von Interviewerverhalten
- 84/04 Frank Faulbaum  
Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit  
von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der  
Sozialwissenschaften (ALLBUS) 1982
- 84/05 Jürgen Hoffmeyer-Zlotnik  
Wohnquartiersbeschreibung. Ein Instrument zur Bestimmung des sozialen  
Status von Zielhaushalten
- 84/07 Gabriele Hippler, Hans-Jürgen Hippler  
Reducing Refusal Rates in the Case of Threatening Questions: The  
"Door-in-the-Face" Technique
- 85/01 Hartmut Esser  
Befragtenverhalten als "rationales Handeln" - Zur Erklärung von  
Antwortverzerrungen in Interviews
- 85/02 Rolf Porst  
ALLBUS-Bibliographie (4. Fassung, Stand: 30.06.85)
- 85/03 Rolf Porst, Peter Prüfer, Michael Wiedenbeck, Klaus Zeifang  
Methodenbericht zum "ALLBUS 1984"



**Tabelle 1:** Skalenskennwerte der an der amerikanischen Stichprobe entwickelten USA-Skalen und deren Modifikation. USAM-Skalen nach Anwendung in der deutschen Stichprobe

Skalen-Nr.	Skalenname	Anzahl Items	USA-Skalen				USAM-Skalen			
			AM: N=1522		GE: N=2003		AM: N=1522		GE: N=2003	
			PEC <sup>1)</sup>	$\lambda_2$ <sup>2)</sup>	PEC	$\lambda_2$	PEC	$\lambda_2$	PEC	$\lambda_2$
1	Externe Kontrollorientierung	14	.94	1.02	.86	1.10	.90	.97	.94	.93
2	Mißtrauen	8	.86	.92	.72	1.14	.86	.98	.80	.90
3	Depression	10	.96	.99	.93	1.04	.97	.98	.98	.90
4	Arbeitszufriedenheit	9	.90	.97	.83	1.11	.92	.95	.80	.96
5	Politische Partizipation	10	.92	1.01	.83	1.17	.95	.99	.92	.96
6	Optimismus	11	.94	.97	.89	1.08	.94	.99	.79	.89
8	Politische Desillusionierung	9	.96	.96	.93	1.04	.96	.99	.95	.94
9	Zukunftsperepektiven	12	.96	.96	.90	1.03	.92	.98	.94	.96
10	Ökonomische Selbstbestimmung	5	.88	.93	.90	.94	.90	.98	.88	.94

1) Anteil der durch ein einfaktorielles Modell aufgeklärten Varianz (Korrelation)

2) Eigenwerte des 2. Faktors

**Tabelle 3:** Skalenskennwerte der an der deutschen Stichprobe entwickelten BRD-Skalen und deren Modifikation.  
BRDM-Skalen nach Anwendung in der deutschen Stichprobe

Skalen- Nr.	Skalen- name	Anzahl Items	BRD-Skalen				BRDM <sup>3)</sup> -Skalen				
			AM: N=1522		GE: N=2003		AM: N=1522		GE: N=2003		
			PEC <sup>1)</sup>	$\lambda_2$ <sup>2)</sup>	PEC	$\lambda_2$	PEC	$\lambda_2$	PEC	$\lambda_2$	
1	Externe Kontrollorientierung	9	.96	.91	.95	.97	9	.95	.97	.96	.91
2	Mißtrauen	6	.90	.87	.91	.95	6	.91	.95	.90	.87
3	Depression	9	.97	.91	.94	.92	9	.94	.92	.97	.91
4	Arbeitszufriedenheit	8	.80	.96	.92	.95	8	.92	.92	.80	.96
5	Politische Partizipation	5	.93	.90	.91	.93	5	.91	.93	.93	.90
6	Optimismus	10	.89	1.09	.95	.99	5	.95	.92	.90	.99
8	Politische Desillusionierung	10	.90	1.04	.98	.99	6	.98	.90	.97	.98
9	Zukunftsperepektiven	10	.95	.97	.94	.98	10	.94	.98	.95	.97
10	Ökonomische Selbstbestimmung	5	.88	.94	.90	.98	5	.90	.98	.88	.94

1) Anteil der durch ein einfaktorielles Modell aufgeklärten Varianz (Korrelation)

2) Eigenwerte des 2. Faktors

3) Bei den BRD-Skalen 1,2,3,5,9 und 10 war keine Modifikation erforderlich



Tabelle 7: COMB-Skala 9: Zukunftsperspektiven;  
Itemformulierungen und Kennwerte sowie Skalenkennwerte für die  
deutsche (GE: N=2003) und die amerikanische (AM: N=1522) Stich-  
probe

	Item- mittelwert		Item-Gesamt- Korrelation	
	GE	AM	GE	AM
1. In den nächsten Jahren wird es noch mehr Arbeitslose geben.	.82	.75	.34	.30
2. Obwohl sich ständig sehr viel ändert, weiß man im großen und ganzen doch, was man zu erwarten hat.	.28	.55	.21	.29
3. Die Zukunft dieses Landes ist ungewiß.	.59	.72	.49	.49
4. Die Zukunft sieht sehr trübe aus.	.52	.46	.56	.55
5. Meiner Meinung nach ist dieses Land kaputt.	.16	.43	.33	.38
6. Die Zukunft ist zu ungewiß, als daß man weit im voraus planen könnte.	.63	.50	.46	.45
7. Wir verlieren langsam unsere persönliche Freiheit, weil sich die Behörden immer weiter ausdehnen.	.54	.65	.37	.42
8. Das Leben des Normalbürgers wird schlechter, nicht besser.	.59	.53	.48	.45
9. So, wie die Zukunft aussieht, ist es unfair, Kinder in die Welt zu setzen.	.29	.35	.34	.37
10. Viele Grundsätze unserer Eltern gehen den Bach runter.	.72	.77	.24	.36
11. Ich habe zur Zeit wenig Vertrauen zur Regierung.	.62	.58	.44	.45
12. Ich finde, daß die Zukunft recht gut aussieht.	.70	.40	.45	.38
		Mittelwert	6.45	6.68
		Standardabweichung	2.88	3.06
		Eigenwert des 2. Faktors	0.00	0.00