

### Die Analyse kategorialer Panel-Daten mit Hilfe von log-linearen Kausalmodellen mit latenten Variablen: eine Anwendung am Beispiel der Skala "Jugendzentrismus"

Vermunt, Jeroen K.; Georg, Werner

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:

Vermunt, J. K., & Georg, W. (1995). Die Analyse kategorialer Panel-Daten mit Hilfe von log-linearen Kausalmodellen mit latenten Variablen: eine Anwendung am Beispiel der Skala "Jugendzentrismus". *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 36, 61-90. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-201159>

#### 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.

## Die Analyse kategorialer Panel-Daten mit Hilfe von log-linearen Kausalmodellen mit latenten Variablen: Eine Anwendung am Beispiel der Skala "Jugendzentrismus"

Jeroen K. Vermunt und Werner Georg<sup>1</sup>

### Zusammenfassung

In diesem Artikel wird ein allgemeiner Ansatz zur Analyse von kategorialen Panel-Daten vorgestellt, der ein log-lineares Pfadmodell für den strukturellen Teil mit einem Meßmodell für kategorialen Daten verbindet. Während das Strukturmodell aus einem System von Logit-Gleichungen besteht, die die kausalen Zusammenhänge zwischen den latenten Variablen spezifizieren, wird als Meßmodell für die kategorialen Indikatoren ein latent class Modell verwendet, das es erlaubt, wirklichen Wechsel von zufälligem, meßfehlerbedingtem Wechsel zu unterscheiden. Mit Hilfe von Restriktionen, die den Kategorien der Indikatoren oder der latenten Klassen auferlegt werden, können in diesem Rahmen diskretisierte Varianten der meisten latent trait Modelle (so des Rasch-Modell, des Lord-Birnbaum-Modell oder des partial credit Modells) als restringierte latent class Modelle formuliert werden. Log-lineare Pfadmodelle mit latenten Variablen können mit Hilfe des Programms *MEM* geschätzt werden. Am Beispiel der Skala "Jugendzentrismus", die über zwei Zeitpunkte gemessen wurde, werden mit Hilfe dieses Programms mehrere Meßmodelle getestet und in einem zweiten Schritt der Einfluß unterschiedlicher Kovariaten auf die Ausgangsposition und die latenten Übergänge der latenten Variablen "Jugendzentrismus" untersucht.

### Abstract

In this paper a general approach for the analysis of categorical panel-data is presented, in which a log-linear path model for the structural part is combined with a measurement model for categorical data. The structural model consists of a system of logit equations which specify the relationship between the latent variables. A latent class model, which allows to correct for measurement errors is used as measurement model for the categorical indicators. If linear restrictions are imposed on the categories of the indicators or the latent variables, most discretized variants of latent trait models (the Rasch-model, the Lord-Birnbaum-model or the partial credit model) can be formulated as restricted latent class models. Log-linear path models with latent variables can be estimated with the program *MEM*. To demonstrate this approach, several measurement models for the scale "Youth-Centrism", measured over two points in time, are tested. In a second step the influence of different covariates on the initial position and the transitions of the latent variable "Youth-Centrism" are investigated.

1 Jeroen K. Vermunt ist Universitätsdozent an der Abteilung für Methodologie, Fakultät Sozialwissenschaften, an der Katholischen Universität Tilburg. Dr. Werner Georg ist wissenschaftlicher Mitarbeiter an der Universität-GH Siegen, Projekt Bildungsmoratorium im Schwerpunktprogramm der DFG "Kinder und Jugendliche vor und nach der Vereinigung" und vertritt gegenwärtig eine Professur für empirische Sozialforschung an der Universität Marburg

## 1. Einleitung

Für die Untersuchung der Determinanten sozialen Wandels sind Panel- und Ereignisdaten besonders gut geeignet. In diesem Artikel werden wir einen allgemeinen Ansatz zur Analyse von kategorialen Panel-Daten vorstellen, der ursprünglich von *Hagenaars* (1990, 1993) vorgeschlagen wurde und der dem bekannten LISREL-Modell für kontinuierliche Variablen (*Jöreskog & Sörbom* 1988) ähnelt. Wie LISREL kombiniert dieser Ansatz ein Pfadmodell, in dem der strukturelle Zusammenhang zwischen den latenten Variablen spezifiziert wird, mit einem Meßmodell für die Verbindung zwischen den latenten Variablen und ihren Indikatoren. Wegen dieser Analogie mit dem LISREL-Modell nannte *Hagenaars* sein Modell einen "modifizierten LISREL-Ansatz".

Der strukturelle Teil des "modifizierten LISREL-Ansatzes" ist ein "modifiziertes Pfadmodell", wie es von *Goodman* (1973) vorgeschlagen wurde. Ein "modifiziertes Pfadmodell" ist ein log-lineares Modell, in dem einem Satz kategorialer Daten eine vorgegebene kausale Reihenfolge auferlegt wird. Ein derartiges log-lineares Kausalmodell besteht aus einem System von Logit-Gleichungen, in dem eine Variable, die in einer Gleichung als Response-Variable fungiert, in einer anderen als erklärende Variable für ihr kausal nachgeordnete Variablen verwendet wird.

Das Meßmodell des "modifizierten LISREL-Ansatzes" ist ein latent class Modell, ein "faktorenanalytisches Modell" für kategoriale Daten (*Lazarsfeld & Henry* 1968; *Goodman* 1974; *Haberman* 1979). Latent class Modelle können dazu verwendet werden, kategoriale beobachtete Variablen um ihren Meßfehler zu bereinigen oder, in anderen Worten, um Meßmodelle mit kategorialen Indikatoren zu konstruieren. Obwohl unrestringierte latent class Modelle die latente Variable und ihre Indikatoren als nominal skalierte Variablen behandeln, wurden von mehreren Autoren restringierte latent class Modelle entwickelt, die es erlauben, den Kategorien der latenten Variable oder der Indikatoren ordinale oder lineare Restriktionen aufzuerlegen (*Rost & Georg* 1991; *Formann* 1992; *Heinen* 1993). So zeigte *Heinen* (1993), daß die meisten latent trait Modelle als restringierte latent class Modelle dargestellt werden können, wenn die latente Verteilung diskretisiert wird.

*Hagenaars* (1990, 1993) demonstrierte, wie modifizierte Pfadmodelle und latent class Modelle kombiniert werden können. Auf diese Weise konnte der strukturelle Zusammenhang zwischen beobachteten und latenten Variablen mit Hilfe eines log-linearen Kausalmodells dargestellt werden. Bei der Analyse von Panel-Daten kann im Rahmen dieses Modells wirklicher Wechsel ("true change") von zufälligem, meßfehlerbedingtem Wechsel unterschieden werden (*Van de Pol & De Leeuw* 1986; *Hagenaars* 1992). Da eine allgemeinere Form des Logit-Modells verwendet wird, ist der hier dargestellte Ansatz etwas allgemeiner

formuliert als das ursprünglich von *Hagenaars* vorgeschlagene Modell (vgl. *Vermunt* 1995).

Das log-lineare Kausalmodell mit latenten Variablen ist eng mit dem latenten und gemischten *Markov* Modell verbunden. (*Wiggins* 1973; *Van de Pol & Langeheine* 1990). Tatsächlich sind diese Modelle der Panelanalyse Spezialfälle des hier dargestellten Ansatzes (*Vermunt* 1995). Wenn ein latentes oder gemischtes *Markov* Modell als log-lineares Kausalmodell formuliert wird, ist es jedoch möglich, den bedingten Wahrscheinlichkeiten Logit-Restriktionen aufzuerlegen. Für den strukturellen Teil des Modells bedeutet dies, daß sparsamere Logit-Modelle für die (latenten) Übergänge spezifiziert werden können. Im Meßmodell können Logit-Modelle beispielsweise dazu benutzt werden, diskrete Annäherungen an latent trait Modelle, wie etwa das semi-parametrische *Rasch*-Modell (*Lindsay, Clogg & Grego* 1991; *Heinen* 1993) zu spezifizieren.

In Abschnitt 2 wird das modifizierte Pfadmodell behandelt. Restringierte und unrestringierte latent class Modelle und das modifizierte LISREL-Modell werden in Abschnitt 3 diskutiert. Eine exemplarische Anwendung des hier beschriebenen Ansatzes für Panel-Daten anhand der in zwei Wellen erfaßten Skala "Jugendzentrismus" (Nachbefragung zur Shell-Jugendstudie 1992) wird in Abschnitt 4 dargestellt.

## 2. Modifizierte Pfadmodelle

In diesem Abschnitt werden wir eine pfadanalytische Erweiterung des klassischen Logit-Modells diskutieren. *Goodman* (1973) schlug ein log-lineares Modell vor, das vorgegebene Informationen über die kausale Reihenfolge der Variablen berücksichtigt. Er nannte sein Modell einen "modifizierten pfadanalytischen Ansatz". Wie weiter unten dargestellt wird, ist dieses Modell gut geeignet für die Analyse von kategorialen Panel-Daten. Hierbei ist anzumerken, daß der Ausdruck "kategorial" nicht gleichbedeutend mit "nominal" ist: ordinale Variablen und diskrete intervallskalierte Variablen sind in dieser Konnotation ebenfalls kategoriale Variablen (*Agresti* 1990).

### 2.1 Die Festlegung der Wahrscheinlichkeitsstruktur

In dem folgenden fiktiven Beispiel wird angenommen, daß im Rahmen einer Panel-Studie mit drei Wellen die individuellen Übergangswahrscheinlichkeiten einer spezifischen kategorialen Variablen erklärt werden sollen. Die abhängige Variable zum ersten, zweiten und dritten Meßzeitpunkt sei mit W, Y und Z bezeichnet. Weiterhin indizieren R, S und T drei kategoriale unabhängige Variablen, die dazu verwendet werden, die Werte von W sowie die Übergänge von W zu Y und von Y zu Z zu erklären. In diesem Beispiel sind R, S und T exogene Variablen, während W, Y und Z endogene Variablen repräsentieren, wobei angenommen wird, daß Y in der kausalen Reihenfolge W nachgelagert ist und Z auf Y folgt.

Für dieses Modell soll zunächst unterstellt werden, daß alle Variablen direkt beobachtbar sind.

Es sei  $\pi_{rstwyz}$  die Wahrscheinlichkeit daß  $R = r$ ,  $S = s$ ,  $T = t$ ,  $W = w$ ,  $Y = y$  und  $Z = z$ . Unter Verwendung der Information über die kausale Reihenfolge der Variablen, kann  $\pi_{rstwyz}$  dargestellt werden als (*Goodman* 1973):

$$\pi_{rstwyz} = \pi_{rst} \pi_{w|rst} \pi_{y|rstw} \pi_{z|rstwy}^2 \quad (1)$$

Nach Formel (1) wird die Information über die kausale Reihenfolge der Variablen benutzt, um die gemeinsame Wahrscheinlichkeit in ein Produkt aus Randwahrscheinlichkeiten und bedingten Wahrscheinlichkeiten zu zerlegen. Auf diese Weise wird ausgedrückt, daß der Wert einer spezifischen Variablen nur von den vorhergehenden Variablen, nicht jedoch von den nachfolgenden abhängen kann. So wird zum Beispiel angenommen, daß  $Y$  nur von den vorhergehenden Variablen  $R$ ,  $S$ ,  $T$  und  $W$  abhängen kann, nicht jedoch von der nachfolgenden Variable  $Z$ . Deshalb hängt die Wahrscheinlichkeit, daß  $Y = y$  nur von den Werten von  $R$ ,  $S$ ,  $T$  und  $W$  ab, nicht jedoch von dem Wert von  $Z$ . Es ist festzuhalten, daß das durch Gleichung 1 beschriebene Modell ein rekursives Modell ist. Obwohl nicht-rekursive Modelle für Kategorialdaten, wie sie von *Mare* und *Winship* (1991) vorgeschlagen wurden, auch innerhalb des hier dargestellten Ansatzes spezifiziert werden können, werden wir uns in den folgenden Ausführungen auf rekursive Modelle beschränken.

Die Zerlegung der gemeinsamen Wahrscheinlichkeit in einen Satz aus Randwahrscheinlichkeiten und bedingten Wahrscheinlichkeiten ist nur der erste Schritt für die Beschreibung der kausalen Beziehungen zwischen den untersuchten Variablen. Während in Gleichung 1 ein saturiertes Modell formuliert wurde, in dem angenommen wird, daß eine spezifische abhängige Variable von allen vorhergehenden Variablen einschließlich ihrer Interaktionen beeinflusst wird, soll im folgenden die Anzahl der zu schätzenden Parameter weiter reduziert werden.

Der einfachste Weg, sparsamere Modelle zu spezifizieren, liegt in der direkten Begrenzung der bedingten Wahrscheinlichkeiten, die in Gleichung 1 dargestellt werden. Im folgenden soll angenommen werden, daß  $W$  von  $R$  und  $S$  abhängt, aber nicht von  $T$ , daß  $Y$  von  $S$ ,  $T$  und  $W$  abhängt, aber nicht von  $R$  und, daß  $Z$  von  $S$ ,  $T$  und  $Y$  abhängt, aber nicht von  $R$ . Mit anderen Worten, die anfängliche Position hängt von  $R$  und  $S$  ab und die Übergangswahrscheinlichkeiten von  $S$  und  $T$ . Diese Einschränkungen können einfach in das Modell eingeführt werden, indem das unrestringierte Modell aus Gleichung 1 ersetzt wird durch:

$$\pi_{rstwyz} = \pi_{rst} \pi_{w|rs} \pi_{y|stw} \pi_{z|sty} \quad (2)$$

<sup>2</sup> In Formel 1 steht bspw. der Ausdruck  $\pi_{w|rst}$  für die Wahrscheinlichkeit, daß  $W=w$  ist unter der Bedingung, daß die unabhängigen Variablen  $r$ ,  $s$  und  $t$  spezifische Ausprägungen besitzen.

Die Einschränkung, daß die Wahrscheinlichkeit der Response-Variablen nur von einigen der vorhergehenden Variablen abhängt stellt die einfachste Möglichkeit der Spezifizierung eines sparsameren Modells dar. Auf ihr beruhen auch *Markov-Modelle* mit diskreter Zeitabhängigkeit (*Bishop, Fienberg & Holland* 1975; *Van de Pol & Langeheine* 1990). Bei dem in Gleichung 2 formulierten Modell handelt es sich ebenfalls um ein Markov-Modell.

Die oben beschriebene relativ einfache Möglichkeit zur Spezifizierung eines eingeschränkteren Modells hat jedoch einen wichtigen Nachteil: Die abhängige Variable wird immer in Beziehung gesetzt zur gemeinsamen Verteilung der unabhängigen Variablen. Wenn beispielsweise die Variable Y von der gemeinsamen Verteilung von STW abhängt, existiert für jede Kombination von S, T und W ein Parameter, der die Wahrscheinlichkeit beschreibt daß  $Y = y$ . Das heißt, wenn angenommen wird, daß eine spezifische Variable die jeweilige abhängige Variable beeinflusst, so werden automatisch alle Interaktionen mit den anderen unabhängigen Variablen in das Modell aufgenommen. Das Ergebnis wird im Allgemeinen sein, daß das Modell immer noch mehr Parameter enthält, als notwendig wären.

## 2.2 Die Logit-Parametrisierung der Wahrscheinlichkeiten

Mit Hilfe einer log-linearen oder Logit-Parametrisierung der Randwahrscheinlichkeiten und der bedingten Wahrscheinlichkeiten aus Gleichung 1 ist es möglich, sparsamere Kausalmodelle für kategoriale Daten zu spezifizieren. Diese Vorgehensweise führt zu dem von *Goodman* (1973) vorgeschlagenen "modifizierten pfadanalytischen Ansatz". Ein Kausalmodell für die Beziehungen zwischen den im Beispiel verwendeten Variablen würde aus vier sogenannten modifizierten Pfad-Schritten oder Sub-Modellen bestehen: Einem saturierten log-linearen Modell für die exogenen Variablen R, S und T und drei Logit-Modellen, in denen Y und Z als abhängige Variablen erscheinen.

Das Logit-Modell, in dem Y als Response-Variable verwendet wird, könnte z.B. folgende Form haben:

$$\pi_{y|rstw} = \frac{\exp(u_y^Y + u_{sy}^{SY} + u_{ty}^{TY} + u_{wy}^{WY} + u_{swy}^{SWY})}{\sum_y \exp(u_y^Y + u_{sy}^{SY} + u_{ty}^{TY} + u_{wy}^{WY} + u_{swy}^{SWY})}$$

In dieser Formel sind die u-Terme log-lineare Parameter, die den aus ANOVA-Modellen bekannten Einschränkungen unterworfen sind. Das Logit-Modell für Y enthält Interaktionen erster Ordnung für S, T, W und Y und die Interaktion zweiter Ordnung für S, W und Y. Die selbe Art von Logit-Modellen kann für die beiden bedingten Wahrscheinlichkeiten spezifiziert werden, die in Gleichung 1 thematisiert werden, nämlich für  $\pi_{w|rst}$  und  $\pi_{z|rstwy}$ . Somit wird deutlich, daß mit einem derartigen System von Logit-Gleichungen sparsamere Modelle formuliert werden können als mit Hilfe der in Gleichung 2 verwendeten Einschränkungen der bedingten Wahrscheinlichkeiten.

Die Spezifizierung eines Logit-Modells für einen Satz bedingter Wahrscheinlichkeiten ist gleichbedeutend mit der Spezifizierung eines log-linearen Modells für eine multivariate Kreuztabelle, bei der die Randverteilung der unabhängigen Variablen fixiert wird (*Goodman* 1972; *Haberman* 1979; *Agresti* 1990). So ist beispielsweise das Logit-Modell, das für  $\pi_{y|rstw}$  spezifiziert wurde äquivalent zu dem log-linearen Modell  $\{\text{RSTW}, \text{TY}, \text{SWY}\}$  für die (Rand) Häufigkeitstabelle RSTWY, oder

$$\log m_{rstwy} = \alpha_{rstw}^{RSTW} + u_y^Y + u_{sy}^{SY} + u_{ty}^{TY} + u_{wy}^{WY} + u_{swy}^{SWY} \quad (3)$$

wobei  $m_{rstwy}$  die erwartete Häufigkeit der Tabelle RSTWY bezeichnet und  $\alpha_{rstw}^{RSTW}$  den Effekt repräsentiert, der die Randverteilung der abhängigen Variablen fixiert.

Deshalb kann die Spezifizierung des log-linearen Kausalmodells dadurch erreicht werden, daß getrennte log-lineare Modelle für unterschiedliche Rand- oder Sub-Tabellen formuliert werden. Die Randtabellen werden dann aus den Variablen der vorhergehenden Randtabellen mit den entsprechenden abhängigen Variablen gebildet (*Goodman* 1973). In diesem Fall müssen log-lineare Modelle für die Tabellen RST, RSTW, RSTWY und RSTWYZ spezifiziert werden. Dabei können die Wahrscheinlichkeiten in Gleichung 1 aus den erwarteten Häufigkeiten der Sub-Tabellen durch (4) errechnet werden.

$$\pi_{y|rstw} = \frac{m_{rstwy}}{\sum_y m_{rstwy}} \quad (4)$$

### 2.3 Die Veränderung von Goodmans Ansatz

Weiter oben wurde gezeigt, daß die bedingten Wahrscheinlichkeiten eines modifizierten Pfadmodells entweder durch eine einfache Begrenzung der Wahrscheinlichkeiten oder eine Logit-Parametrisierung restringiert werden können. Tatsächlich ist es jedoch einfacher und für die Berechnung effizienter, diese beiden Arten der Einschränkung der bedingten Wahrscheinlichkeiten zu kombinieren. Genauer formuliert, kann man das Modell zuerst in der Weise einschränken, wie es in Gleichung 2 praktiziert wurde, und dann die in dieser Gleichung vorhandenen bedingten Wahrscheinlichkeiten mit Hilfe einer Logit-Parametrisierung restringieren. Dieses Vorgehen führt zu einer kleinen Veränderung der von *Goodman* (1973) vorgeschlagenen Vorgehensweise.

Betrachtet man das Modell für die abhängige Variable Y, so wird deutlich, daß man  $\pi_{y|rstw}$  durch  $\pi_{y|stwr}$  ersetzen kann, weil Y nicht von R abhängt. Deshalb kann man die log-linearen

Restriktionen, die zunächst auf  $\pi_{y|rstw}$  angewendet wurden, auch direkt auf  $\pi_{y|stw}$  beziehen.

Dies ist gleichbedeutend mit der Spezifizierung eines log-linearen Modells {STW, TY, SWY} für die Randtabelle STWY, d.h. Gleichung 3 kann ersetzt werden durch:

$$\log m_{stwy} = \alpha_{stw}^{STW} + u_y^Y + u_{sy}^{SY} + u_{ty}^{TY} + u_{wy}^{WY} + u_{swy}^{SWY}.$$

wobei  $m_{stwy}$  eine erwartete Häufigkeit in der Randverteilung STWY bezeichnen. Deshalb werden mit Hilfe von zwei schrittweisen Restriktionen log-lineare Parameter in der Randverteilung geschätzt, die lediglich diejenigen abhängigen Variablen enthalten, die auch wirklich benutzt werden. Diese Vorgehensweise verringert die Größe des Modells erheblich und ist deshalb bei der Berechnung effizienter. Zudem hat sie den Vorteil, daß sie "gefittete" Nullen verhindert, wenn die beobachtete Tabelle Nullen in der Randverteilung von RSTW enthält, nicht jedoch in derjenigen von STW.

Bis zu diesem Punkt wurden nur einfache hierarchische log-lineare Modelle benutzt, um bedingte Wahrscheinlichkeiten zu parametrisieren. Weiterhin wurde angenommen, daß alle in das modifizierte Pfadmodell aufgenommenen Variablen kategorial sind. Tatsächlich befinden sich diese beiden Merkmale in Übereinstimmung mit der Weise, in der **Goodman** sein modifiziertes Pfadmodell darstellte. Wenn man einen allgemeinen Typ von Logit-Modellen benutzt, um die bedingten Wahrscheinlichkeiten zu parametrisieren, ist es jedoch möglich, nicht-hierarchische log-lineare Modelle zu spezifizieren und kontinuierliche exogene Variablen in einem modifizierten Pfadmodell zu verwenden. Angenommen,  $k$  sei der Index der abhängigen Variable in einer spezifischen Logit-Gleichung und  $i$  bezeichne den Index der gemeinsamen Verteilung der unabhängigen Variablen. In diesem Fall besitzt das Logit-Modell die allgemeine Form:

$$\pi_{k|ij} = \frac{\exp\left(\sum_j u_j x_{ikj}\right)}{\sum_k \exp\left(\sum_j u_j x_{ikj}\right)}, \quad (5)$$

wobei  $u_j$  ein log-linearer Parameter und  $x_{ijk}$  ein Element der Design-Matrix ist. Dieses Logit-Modell ist äquivalent mit dem von **Habermann** (1979) vorgeschlagenen multi-nomialen Responsemodell. Wenn der Index  $i$  benutzt wird, um eine spezielle individuelle, statt die gemeinsamen unabhängigen Variablen zu bezeichnen, wird aus dem in Gleichung 5 wiedergegebenen Modell ein logistisches Regressionsmodell (**Agresti** 1990). In diesem Fall kann es mit kontinuierlichen unabhängigen Variablen verwendet werden, wobei  $x_{ijk}$  den Wert einer Person  $i$  auf der Variablen  $j$  für die Ausprägung  $k$  der Responsevariable bezeichnet.



## 2.4 Schätzung und Test

Maximum Likelihood-Schätzungen für die log-linearen Parameter und die erwarteten Häufigkeiten können mit Hilfe von Programmen für log-lineare Analysen durchgeführt werden. Dabei können die Gleichungen für die verschiedenen Untermodelle getrennt geschätzt werden und die geschätzten Zellen-Wahrscheinlichkeiten für das Gesamtmodell können mit Hilfe der Gleichungen (4) und (1) berechnet werden. Der Test des Modells kann beispielsweise mit der  $\text{Chi}^2$ -verteilten Likelihood-Ratio Teststatistik  $L^2$  durchgeführt werden. Die spezifischen  $L^2$ -Werte der Sub-Modelle und ihre Freiheitsgrade können eingefügt werden, um einen Test für die Anpassung des Gesamtmodells zu erhalten. Es muß jedoch betont werden, daß bei der Verwendung kontinuierlicher exogener Variablen die  $L^2$ -Statistik nicht mehr aussagekräftig ist. Wie in logistischen Regressionsmodellen kann man den Likelihood-Ratiotest dann nur benutzen, um hierarchisch geordnete Modelle gegeneinander zu testen.

Es wurde ein Programm namens **LEM** entwickelt, mit dem die in diesem Abschnitt diskutierten log-linearen Pfadmodelle geschätzt werden können, ohne die jeweiligen Randverteilungen getrennt schätzen zu müssen (*Vermunt* 1993). In **LEM** besteht die Standardprozedur der Modellierung einer beobachteten Häufigkeitstabelle in der Spezifizierung eines log-linearen Pfadmodells. Die allgemein verfügbare Version des Programms (**LEM 0.11**) verwendet die von *Goodman* (1973) vorgeschlagene Prozedur. Aber in der gegenwärtigen Testversion von **LEM** wurde die oben beschriebene effizientere schrittweise Prozedur der Restriktion von Wahrscheinlichkeiten implementiert.

Die in **LEM** implementierte Standardschätzprozedur, zur Berechnung hierarchischer log-linearer Modelle, ist der Iterative Proportional Fitting Algorithm (IPF). Jedoch kann **LEM** auch benutzt werden, um komplexere log-lineare Modelle zu schätzen, wie sie etwa in Gleichung 5 wiedergegeben sind. Dies wird durch die Möglichkeit der Spezifizierung von Design-Matrizen für einzelne log-lineare Effekte ermöglicht. In **LEM** ist es auch möglich, log-multiplikative Effekte, wie beispielsweise die von *Goodman* entwickelten "Type II Associations Modells" (*Goodman* 1979; *Clogg* 1982; *Xie* 1992) zu verwenden. Nicht-hierarchische log-lineare Modelle werden mit Hilfe einer Variante des eindimensionalen *Newton-Algorithmus*, wie er von *Goodman* (1979) vorgeschlagen wurde, geschätzt. Dieser Algorithmus unterscheidet sich von dem bekannten *Newton-Raphson-Algorithmus* dadurch, daß bei einer Iteration nur ein Parameter statt aller Parameter gleichzeitig verändert wird (*Vermunt* 1995).

Neue Optionen, die in der gegenwärtigen Arbeitsversion von **LEM** implementiert wurden, erlauben es, log-linearen Parametern, die in verschiedenen Logit-Gleichungen erscheinen,

Gleichheitsrestriktionen aufzuerlegen und die bedingten Wahrscheinlichkeiten gleichzusetzen oder ihnen feste Werte zuzuweisen.

### 3. Modelle mit latenten Variablen

Im letzten Abschnitt wurde angenommen, daß alle Variablen, die in einem log-linearen Kausalmodell verwendet wurden, direkt beobachtbar sind. Da in den Sozialwissenschaften jedoch viele Konzepte schwierig oder überhaupt nicht direkt zu messen sind, werden oft mehrere direkt beobachtbare Variablen, oder Indikatoren, als indirekte Maße für das Konzept benutzt, das wir messen wollen. Es wird dabei angenommen, daß die Werte der Indikatoren von den Werten einer unbeobachteten latenten Variablen beeinflußt werden. In latent structure Modellen wird dieses Prinzip statistisch durch die Annahme lokaler Unabhängigkeit eingeführt. Diese Annahme unterstellt, daß die Indikatoren, für einen spezifischen Wert der latenten Variablen, unabhängig voneinander sind. In anderen Worten, sie korrelieren miteinander nur wegen ihrer gemeinsamen Ursache.

Latent structure Modelle können hinsichtlich des Meßniveaus der latenten und manifesten Variablen klassifiziert werden (*Bartholomew* 1987; *Heinen* 1993). In der Faktorenanalyse werden kontinuierliche manifeste Variablen als Indikatoren für eine oder mehrere kontinuierliche latente Variablen benutzt. In latent trait Modellen wird normalerweise angenommen, daß eine kontinuierliche latente Variable einem Satz kategorialer Indikatorvariablen zugrunde liegt. Wenn sowohl die manifesten als auch die latenten Variablen als kategorial angenommen werden, handelt es sich um ein latent class Modell (*Lazarsfeld* und *Henri* 1968; *Goodman* 1974; *Habermann* 1979).

#### 3.1 Unrestringierte latent class-Modelle

Im folgenden Beispiel ist  $W$  mit Index  $w$  eine Variable und  $A, B, C$  sind drei Indikatoren mit Indizes  $a, b, c$ . Zusätzlich bezeichnet  $W^*$  die Anzahl der latenten Klassen. Die Grundgleichung des latent class Modells ist

$$\pi_{abc} = \sum_{w=1}^{W^*} \pi_{wabc} \quad (6)$$

mit

$$\pi_{wabc} = \pi_w \pi_{a|w} \pi_{b|w} \pi_{c|w}. \quad (7)$$

Hierbei bezeichnet  $\pi_{wabc}$  eine Wahrscheinlich in der gemeinsamen Verteilung, unter Ein-schluß der latenten Dimension  $W$ . Zusätzlich ist  $\pi_w$  der Anteil einer Population, die zur latenten Klasse  $W$  gehört. Die anderen  $\pi$ -Parameter sind die bedingten Antwortwahrschein-

lichkeiten. Z.B. ist  $\pi_{a|w}$  die Wahrscheinlichkeit, einen Wert von  $a$  auf  $A$  zu haben, wenn man zur latenten Klasse  $W$  gehört.

Aus Gleichung 6 kann man ersehen, daß die Population in  $W^*$  erschöpfende und sich ausschließende Klassen eingeteilt wird. Deshalb kann die gemeinsame Wahrscheinlichkeit der beobachteten Variablen durch eine Aufsummierung über die latenten Klassen erhalten werden. Die klassische Parametrisierung von latent class Modellen, wie sie von *Lazarsfeld* und *Henri* (1968) und *Goodman* (1974) vorgeschlagen werden, ist in Gleichung 7 wiedergegeben. Aus ihr kann ersehen werden, daß die beobachteten Variablen  $A$ ,  $B$  und  $C$  bezüglich eines spezifischen Wertes der latenten Variablen  $W$  wechselseitig unabhängig voneinander sind. Dabei ist festzustellen, daß Gleichung 7 hier dem modifizierten Pfadmodell ähnelt, wie es im vorhergehenden Abschnitt dargestellt wurde. Tatsächlich handelt es sich hier um ein modifiziertes Pfadmodell, in dem eine Variable unbeobachtet ist.

*Habermann* (1979) demonstrierte, daß latent class Modelle log-lineare Modelle mit latenten Variablen sind. So ist beispielsweise das latent class Modell, das in Gleichung 7 formuliert wurde, äquivalent einem hierarchischen log-linearen Modell  $\{WA, WB, WC\}$ . Deshalb kann sie auch dargestellt werden als

$$\log m_{wabc} = u + u_w^W + u_a^A + u_b^B + u_c^C + u_{wa}^{WA} + u_{wb}^{WB} + u_{wc}^{WC} \quad (8)$$

wobei  $m_{wabc} = \pi_{wabc} N$ . In dieser Formulierung eines latent class Modells wird die Annahme lokaler Unabhängigkeit durch die Einführung der rechten Parameter in das log-lineare Modell erfüllt. Gleichung 8 enthält außer dem Gesamtmittelwert und den Termen der einzelnen Variablen nur die Interaktionen erster Ordnung zwischen der latenten Variablen  $W$  und den manifesten Variablen. Aus der Tatsache, daß keine Interaktionseffekte zwischen den manifesten Variablen eingeführt werden, wird deutlich, daß diese bedingt unabhängig voneinander sind.

Die Beziehung zwischen den Parametern dieser beiden Formen der Parametrisierung des latent class Modells, d.h. zwischen den bedingten Wahrscheinlichkeiten aus Gleichung 7 und den log-linearen Parametern aus Gleichung 8 läßt sich aus folgender Formel ableiten:

$$\pi_{a|w} = \frac{\exp(u_a^A + u_{wa}^{WA})}{\sum_a \exp(u_a^A + u_{wa}^{WA})} \quad (9)$$

Hierbei fällt auf, daß dies die gleiche Art der Logit-Parametrisierung von bedingten Wahrscheinlichkeiten ist, wie sie in einem modifizierten Pfadmodell verwendet wird.

### 3.2 Restringierte latent class Modelle

Wenn wir den Parametern eines latent class Modells keine weiteren Restriktionen auferlegen sind beide Formulierungen von latent class Modellen vollkommen äquivalent. Wenn wir jedoch restringierte latent class Modelle spezifizieren, hängt es von der Art der Restriktion ab, welche Parametrisierung man wählen muß. Obwohl einige Restriktionen bei beiden Arten der Parametrisierung auferlegt werden können, sind andere spezifisch für eine der beide Parametrisierungen.

Wenn ein mit bedingten Wahrscheinlichkeiten dargestelltes latent class Modell in Kombination mit einer Logit-Parametrisierung dieser bedingten Wahrscheinlichkeiten formuliert wird, so ist diese Form der kombinierten Parametrisierung ähnlich der eines modifizierten Pfadmodells. Tatsächlich kombiniert *Formanns* linear-logistisches latent class Modell diese beiden Arten der oben diskutierten Parametrisierung (*Formann* 1992). *Formann* spezifiziert das latent class Modell in Form von latenten und bedingten Wahrscheinlichkeiten, wobei diese Wahrscheinlichkeiten in der Form parametrisiert werden, wie sie in Gleichung 9 dargestellt ist. Eine solche Formulierung macht es möglich, Restriktionen der Wahrscheinlichkeiten mit Restriktionen der log-linearen oder linear-logistischen Parameter zu verbinden

Kürzlich zeigte *Heinen* (1993), daß bei einer Diskretisierung der latenten Variablen die meisten latent trait Modelle als latent class Modelle mit Restriktionen auf den log-linearen Parametern formuliert werden können. Mit anderen Worten, die log-linearen latent class Modelle können verwendet werden, um diskretisierte Varianten der bekannten latent trait Modelle wie etwa das *Rasch*-Modell (*Rasch* 1960), das *Lord-Birnbaum-Modell* (*Lord* und *Nowick* 1968), das nominal response Modell (*Bock* 1972) und das partial credit-Modell (*Masters* 1982) zu formulieren.

Im folgenden soll angenommen werden, daß eine *Rasch*-Skala mit Hilfe dreier dichotomer Items ABC konstruiert wird. Die Grundannahme des *Rasch*-Modells ist, daß alle Items den gleichen Trennschärfe-Koeffizienten besitzen, oder in anderen Worten, daß die Item Characteristic-Kurven parallel verlaufen (*Rasch* 1960; *Rost* und *Georg* 1991). Es wird vorausgesetzt, daß die Wahrscheinlichkeit einer "richtigen" Antwort nur von der Fähigkeit einer Person und der Schwierigkeit des jeweiligen Items abhängt. Nach *Heinen* (1993) erhält man eine diskrete Variante des *Rasch*-Modells, indem man den Interaktionsternen erster Ordnung, die in den bedingten Antwortwahrscheinlichkeiten des latent class Modells auftauchen, spezifische Restriktionen (vgl. Gleichung 9) auferlegt. Der einzige Unterschied zu dem normalen *Rasch*-Modell besteht darin, daß im diskreten *Rasch*-Modell die latente Verteilung der Fähigkeit diskretisiert wird. Genauer gesagt, die Anzahl der unterschiedlichen

Fähigkeitsausprägungen ist gleich der Anzahl der latenten Klassen. Eine diskrete Variante des Rasch-Modells kann man erhalten, indem man folgende Restriktion durchführt.

$$\begin{aligned} u_{wa}^{WA} &= \alpha \theta_w x_a, \\ u_{wb}^{WB} &= \alpha \theta_w x_b, \\ u_{wc}^{WC} &= \alpha \theta_w x_c. \end{aligned} \quad (10)$$

Der Parameter  $\alpha$  ist der Trennschärfekoeffizient, von dem angenommen wird, daß er für alle Items gleich ist. Weiterhin sind  $x_a$ ,  $x_b$  und  $x_c$  die Werte der Kategorien der Items und  $\theta_w$  bezeichnet den Wert der Kategorie  $w$  von  $W$ . Die Kategorien der Items sind hierbei festgelegte Werte. Üblicherweise wird der Wert 0 für eine falsche Antwort und 1 für eine richtige Antwort vergeben. Sollen jedoch die ANOVA-Modellen üblichen Restriktionen erhalten bleiben, so kann man den beiden Kategorien von A, B und C die Werte -1 und +1 zuordnen. Die Werte der latenten Variablen, die auch latent nodes genannt werden, können entweder feste Größen oder Zufallsrealisationen darstellen. Wenn die latent nodes als zufällig angenommen werden, so erhält man ein semi-parametrisches *Rasch*-Modell, das von *Lindsay*, *Clogg* und *Grego* (1991) vorgeschlagen wurde.

Ein diskretisiertes *Lord-Birnbaum-Modell* erhält man durch folgende Art der Restringierung:

$$\begin{aligned} u_{wa}^{WA} &= \alpha^A \theta_w x_a, \\ u_{wb}^{WB} &= \alpha^B \theta_w x_b, \\ u_{wc}^{WC} &= \alpha^C \theta_w x_c. \end{aligned} \quad (11)$$

Der einzige Unterschied zu dem *Rasch*-Modell ist der, daß die Trennschärfekoeffizienten als itemspezifisch angenommen werden. Mit anderen Worten enthält das *Lord-Birnbaum-Modell* außer den Schwierigkeitsparametern, die hauptsächlich von den Haupteffekten abhängen, einen Parameter für jedes Item, der die Stärke der Verbindung zwischen der latenten Variablen und dem entsprechenden Item wiedergibt.

Das allgemeinste Item Response-Modell für polytome Items ist das nominal response Modell. Wenn A, B und C polytom sind, kann eine diskrete Variante des nominal response Modells erhalten werden durch:

$$\begin{aligned} u_{wa}^{WA} &= \alpha_a^A \theta_w, \\ u_{wb}^{WB} &= \alpha_b^B \theta_w, \\ u_{wc}^{WC} &= \alpha_c^C \theta_w. \end{aligned} \quad (12)$$

Hier wird deutlich, daß das nominal response Modell einen Verbindungsparameter für jede Itemkategorie enthält. Weiterhin kann man aus dieser Gleichung ersehen, daß die latente Variable als intervallskalierte Variable angenommen wird, während von den Items vermutet wird, daß sie nominalskaliert sind. Unter der Verwendung der von *Goodman* (1979) eingeführten Terminologie besitzen die latenten Variablen und die Items die Struktur eines column association Modells, wobei die latente Variable eine Zeilenvariable ist.

Ein restriktiveres Modell für polytome Items ist das partial credit Modell. Wie das *Rasch*-Modell erhält man das partial credit Modell, indem man den Interaktionstermen zweiter Ordnung, die in Gleichung 10 bezeichneten Restriktionen auferlegt, wobei nun aber die Items mehr als zwei Kategorien besitzen. Oder in log-linearer Terminologie: Wenn die latent nodes fixiert sind, und sowohl von den latenten als auch den manifesten Kategorien angenommen wird, daß sie äquidistant sind, dann erhält man ein linear-by-linear oder uniforme association Modell (*Habermann* 1979; *Goodman* 1979; *Clogg* 1982) für jede Variableninteraktion. Wenn die Items polytom sind, ist es auch möglich, ein Meßmodell mit den in Gleichung 11 formulierten Restriktionen zu formulieren. Dann erhält man ein partial credit Modell mit itemspezifischen Trennschärfe-Koeffizienten.

### 3.3 Modifizierte LISREL-Modelle

Es wurden einige Erweiterungen der klassischen latent class Modelle vorgeschlagen, wie etwa Modelle mit mehr als einer latenten Variablen (*Goodman* 1974; *Habermann* 1979), Modelle mit sogenannten externen Variablen (*Clogg* 1981), Modelle für den Mehrgruppenvergleich (*Clogg* und *Goodman* 1984; *McCutcheon* 1987) und lokale Abhängigkeitsmodelle (*Hagenaars* 1988). Eine Einschränkung dieser Erweiterung ist jedoch, daß sie alle innerhalb des Rahmens der klassischen oder der log-linearen latent class Modelle entwickelt werden. Deshalb ist es nicht immer möglich, die unterstellte kausale Reihenfolge der in das Modell eingeführten strukturellen Variablen zu postulieren. In diesem Abschnitt wird eine allgemeine Formulierung des modifizierten Pfadmodells mit latenten Variablen, wie es von *Hagenaars* (1990, 1993) vorgeschlagen wurde, entwickelt (vgl. *Vermunt* 1995). Dieses Modell kombiniert ein Strukturmodell mit einem Meßmodell für latente Variablen, d.h. ein modifiziertes Pfadmodell mit einem latent class Modell. Wegen der Analogie mit dem LIS-

REL-Modell für kontinuierliche Variablen (*Jöreskog* und *Sörbom* 1988) nannte *Hagenaars* dieses Modell "einen modifizierten LISREL-Ansatz".

Angenommen, die endogenen Variablen in dem Kausalmodell aus Gleichung 1, W, Y und Z sind latente Variablen und jede von ihnen wird indirekt unter Verwendung dreier beobachteter Variablen gemessen. Die Indikatoren für W werden A, B und C benannt, für Y, D, E und F und für Z, G, H und I. Wie bereits oben dargestellt wurde, können die latenten Variablen auf ihre Indikatoren mit Hilfe des latent class Modells bezogen werden. Die Spezifizierung des gleichen Modells wie in Gleichung 1, aber mit der Annahme, daß die endogenen Variablen latent sind, führt zu folgendem modifiziertem Pfadmodell:

$$\pi_{abcdefghirstwyz} = \pi_{rst} \pi_{w|rstw} \pi_{y|rstw} \pi_{z|rstw} \pi_{a|w} \pi_{b|w} \pi_{c|w} \pi_{d|y} \pi_{e|y} \pi_{f|y} \pi_{g|z} \pi_{h|z} \pi_{i|z} . \quad (13)$$

Wie in dem modifizierten Pfadmodell können alle bedingten Wahrscheinlichkeiten, die in dieser Gleichung auftauchen, mit einer Logit-Parametrisierung restringiert werden. Obwohl in Gleichung 13 implizit angenommen wird, daß das Meßmodell für W, Y und Z nicht von R, S und T abhängt, ist es kein Problem, diese Annahme zu lockern. Wenn wir beispielsweise testen wollen, ob R die Beziehung zwischen Y und B beeinflusst, müssen wir nur  $\pi_{d|y}$  durch  $\pi_{d|y,r}$  ersetzen.

D.h. die Einführung latenter Variablen in ein modifiziertes Pfadmodell setzt hier die Spezifizierung einer Anzahl von zusätzlichen Pfadschritten oder Submodellen voraus, in denen die Beziehung zwischen den latenten Variablen und ihren Indikatoren bezeichnet werden. Diese Submodelle haben die gleiche Struktur wie das latent class Modell, das in Gleichung 7 dargestellt wird.

In dem Modell, das benutzt wurde, um den modifizierten LISREL-Ansatz zu erläutern, wurden die latenten Variablen als indirektes Maß für alle diejenigen Variablen verwendet, die nicht direkt beobachtbar sind. Tatsächlich wurde eine Art latentes *Markov*-Modell mit Kovariaten spezifiziert. Jedoch kann die Möglichkeit, latente Variablen in das Modell zu integrieren, auch dazu benutzt werden, um verschiedene Arten von finite mixture Modellen (*Titterton; Smith* und *Makov* 1985), gemischte *Markov*-Modelle (*van de Pool* und *Langeheine* 1990), gemischte Rasch-Modelle (*Rost* 1990) und gemischte Logit-Modelle (*Formann* 1992) zu spezifizieren.

### 3.4 Schätzverfahren

Die Maximum-Likelihood-Schätzung der Parameter eines log-linearen Modells mit latenten Variablen, ist etwas komplizierter als bei log-linearen Modellen, die nur beobachtbare Variablen enthalten. Maximum-Likelihood-Schätzungen können z.B. mit Hilfe des EM-Algorithmus (*Dempster; Laird; Rubin* 1977) durchgeführt werden. Der EM-Algorithmus ist ein allgemeiner iterativer Algorithmus zur Schätzung von Modellen mit fehlenden

Daten. Er besteht aus zwei unterschiedlichen Schritten pro Iterations-Zyklus, einem E(xpectation) und einem M(aximization) Schritt.

Im E-Schritt des EM-Algorithmus werden die fehlenden Daten geschätzt. In unserem Fall müssen wir Schätzungen für die beobachteten Häufigkeiten der kompletten Tabelle ABCDEFGHIRSTWYZ erhalten,  $\hat{n}_{abcdefghirstwyz}$ , die nach Maßgabe der beobachteten Daten und Parameterschätzung der letzten EM-Iteration geschätzt werden. Dies wird erreicht unter Verwendung der beobachteten unvollständigen Daten und der Parameterschätzung der letzten Iteration durch:

$$\hat{n}_{abcdefghirstwyz} = n_{abcdefghirst} \hat{\pi}_{wyz|abcdefghirst} . \quad (14)$$

Hier stellt  $n_{abcdefghirst}$  eine beobachtete Häufigkeit dar und  $\hat{\pi}_{wyz|abcdefghirst}$  die Wahrscheinlichkeit, daß  $W = w$ ,  $Y = y$  und  $Z = z$ , unter der Bedingung der beobachteten Variablen und unter Verwendung der gegenwärtigen Parameterschätzungen.

Im M-Schritt können Standardschätzprozeduren für log-lineare Modelle wie IPF oder *Newton-Raphson* benutzt werden, um die verbesserte Parameterschätzung zu erhalten unter Verwendung der vervollständigten Daten als seien sie beobachtete Daten. In der Tat wird die Likelihood-Funktion in der  $\hat{n}_{abcdefghirstwyz}$  als Daten erscheinen, die sogenannte Likelihood der vollständigen Daten, maximiert. Die verbesserten Parameterschätzungen werden im E-Schritt wiederum verwendet, um neue Schätzungen für vervollständigte Tabellen zu erhalten usw. Die EM-Iterationen fahren so lange fort, bis irgendein Konvergenzkriterium erfüllt ist, z.B. ein minimaler Anstieg in der Likelihood-Funktion.

Das Programm *EM* (*Vermunt* 1993) ist speziell für die Schätzung modifizierter Pfadmodelle mit latenten Variablen entwickelt worden. Tatsächlich werden in den Modellspezifikationen latente und beobachtete Variablen in genau der gleichen Weise behandelt. Mit Hilfe von *EM* ist es auch möglich, nur teilweise beobachtete Daten in der Analyse zu verwenden (*Vermunt* 1994/1995).

Der in *EM* verwendete Algorithmus ist eine veränderte Version des echten EM-Algorithmus, wobei der M-Schritt immer nur aus einer Iteration besteht. Allgemein gesprochen wird die Likelihood-Funktion der kompletten Daten nicht maximiert, sondern nur im einzelnen M-Schritt verbessert. Dies ist ein Spezialfall des sogenannten GEM-Algorithmus, der unterstellt, daß jede Verbesserung der Likelihood der vollständigen Daten auch zu einer Verbesserung der unvollständigen Daten führt, die maximiert werden soll (*Dempster* und *Rubin* 1977; *Little* und *Rubin* 1987). In der Tat ist der in *EM* verwendete Algorithmus eine Version des ECM-Algorithmus (*Meng* und *Rubin* 1993). Im ECM-Algorithmus wird der M-Schritt ersetzt durch einen bedingten Maximierungs(CM)-Schritt. Bedingte Maxi-



mierung bedeutet, daß anstelle der simultanen Verbesserung aller Parameter, Untermengen der Parameter aktualisiert werden, wobei die anderen auf ihren vorherigen Werten fixiert bleiben. Dies ist die gleiche Vorgehensweise wie die des IPF und des eindimensionalen-Newton-Algorithmus. *Meng* und *Rubin* 1993 zeigen, daß solche einfachen und stabilen linearen Konvergenzmethoden oft sinnvoller für den M-(oder CM)Schritt des EM (oder ECM)-Algorithmus zu verwenden sind, als superlinear konvergierende, aber weniger stabile Algorithmen wie beispielsweise *Newton-Raphson*. Dieser GEM-oder ECM-Algorithmus konvergiert in fast der gleichen Anzahl an Iterationen wie der richtige EM-Algorithmus. Dies macht ihn praktikabler als den echten EM-Algorithmus, speziell bei Anwendung, in denen der wirkliche M-Schritt viele Iterationen zur Konvergenz benötigen würde.

#### 4. Anwendung auf den Wandel in jugendzentrierten Einstellungen

Der weiter oben angestellte Ansatz zur Analyse von Panel-Daten wird im folgenden mit Hilfe einer Anwendung auf den Wandel im Hinblick auf eine jugendzentrierte Einstellung illustriert. Die Daten stammen von der Shell-Jugendstudie 1992 (n = 4005), die im Sommer 1991 durchgeführt wurde und einer zweiten Welle, die zwischen Juli und September 1993 erhoben wurde. In der Wiederbefragung, die Teil eines von der DFG finanzierten Forschungsprojektes ist, wurde eine Unterstichprobe von 292 Personen zusammen mit ihren Eltern interviewt.

##### 4.1 Daten

###### 4.1.1 Der Hintergrund der Shell-Studien

Die erste Shell-Studie wurde 1953 mit dem Ziel durchgeführt, eine jährliche Berichterstattung über die Situation der deutschen Jugend vorzulegen (vgl. *Zinneker* 1985, S. 409 ff.). Die ersten von der deutschen Shell finanzierten Jugendstudien zeichneten ein Bild der westdeutschen Nachkriegsjugend (vgl. etwa *Schelsky* 1957). Seit 1980 wurden drei Shell-Studien mit unterschiedlichen Schwerpunktsetzungen durchgeführt. In der 1981er Studie lag der Schwerpunkt auf den Zukunftsperspektiven und biographischen Orientierungen westdeutscher Jugendlicher. Die Shell-Studie 1985, die 1984 durchgeführt wurde, bezog sich auf einen Vergleich der zeitgenössischen Jugend mit ihrer Elterngeneration. Als Folge der deutschen Einigung lag das Schwergewicht der Shell-Jugendstudie 1992 auf einem Vergleich zwischen ost- und westdeutscher Jugend. Diese Studie wurde mit Hilfe eines Quotensamples mit den Quotierungsmerkmalen Gemeindegrößenklasse, Alter, Schulabschluß und Geschlecht erhoben.

Ein wichtiges gemeinsames theoretisches Konstrukt der letzten drei Shell-Studien ist das Konzept "Jugendzentrismus". Diese Einstellungsdimension "beschreibt eine Haltung der

Jugendlichen, ihre eigene Welt und Lehensauffassung gegen diejenige der Erwachsenen abzugrenzen, gesellschaftlichen Autoritäten zu mißtrauen, da sie von Erwachsenen bestimmt werden und auf ihrem Recht zu beharren, eigene Erfahrungen zu machen, da die Erfahrungen der Erwachsenen zur Lösung der eigenen Probleme nicht tauglich sind. Gekoppelt wird diese Auffassung von Gefühlen von Machtlosigkeit und Aggressivität gegenüber der 'Erwachsenenwelt' (Jugendwerk der Deutschen Shell 1981, 38). Insgesamt wurden für das Konstrukt "Jugendzentrismus" fünf Unterskalen entwickelt, die folgende Dimensionen erfassen: Gefühl der Diskriminierung der Jugendlichen durch gesellschaftliche Autoritäten, Akzeptieren des Erfahrungsvorsprungs und der Vorrechte der Erwachsenen, persönliches Vertrauen und Dankbarkeit gegenüber Erwachsenen und Eltern, Fremdheitsgefühle und Unabhängigkeitsforderungen gegenüber Erwachsenen, Erwachsene / Eltern als verständnislos und übermächtig.

In der 84er Studie wurden aus jeder dieser fünf Unterskalen die zwei "besten" Items nach Maßgabe der Trennschärfe ausgewählt, so daß eine Super-Skala "Jugendzentrismus" aus den fünf Unterdimensionen konstruiert wurde. Diese Skala mit zehn Items wurde in gleicher Weise in der Shell-Studie 1992 verwendet, wobei die Einzelitems vierstufig kodiert wurden (1 = stimmt nicht, 2 = stimmt kaum, 3 = stimmt vielleicht, 4 = stimmt genau).

Die Eigenschaften dieser aus zehn Items bestehenden Skala wurden bereits in verschiedener Hinsicht analysiert. Im Rahmen einer ordinalen latent class Analyse fanden **Rost** und **Georg** (1991) für die 1985er Shell-Studie vier latente Klassen mit klassenspezifischen Schwellenwerten für die Item-Kategorien. Weil nur drei dieser vier Klassen geordnet waren, wurde im folgenden ein Mischverteilungs-**Rasch**-Modell angewendet. Im Rahmen dieser Analyse konnten zwei **Rasch**-skalierbare latente Klassen, bestehend aus 80 bzw. 20% der Stichprobe identifiziert werden.

**Georg** (1992) benutzte die letzten vier Items der 10-Item-Skala in einem Kulturvergleich zwischen Ost- und Westdeutschland und einer Zeitreihenanalyse zwischen 1984 und 1991. Zu diesem Zweck wurde mit Hilfe von LISREL ein Mehrgruppenvergleich unter Verwendung von latenten Mittelwerten berechnet. Für die vier Untergruppen der ost- und westdeutschen Jungen und Mädchen im Jahre 1991 und für den Vergleich zwischen 1984 und 1991 konnte ein gemeinsames kongenerisches Meßmodell konstruiert werden. Zudem konnte festgestellt werden, daß die latenten Mittelwerte der Variable Jugendzentrismus zwischen 1984 und 1991 abnahmen.

#### 4.1.2 Hypothesen

Es wurde bereits erwähnt, daß die Daten für die nun folgende Anwendung aus einer Unterstichprobe der Shell-Studie 1992 (n = 292), die im Jahr 1993 wiederbefragt wurde, stammen. Diese Daten enthalten nur Informationen aus den zwei jüngsten Altersgruppen (15 - 17 Jahre, 18-20 Jahre). Die hohe Panel-Mortalität ist auf die Tatsache zurückzuführen, daß zwischen beiden Wellen das Feldinstitut wechselte und in der zweiten Welle die Eltern der befragten Jugendlichen ebenfalls interviewt wurden. Für die folgende Analyse wurden die vier letzten Items der Skala Jugendzentrismus ausgewählt, die im einzelnen folgenden Wortlaut haben:

1. Die wenigsten Erwachsenen verstehen die Probleme von Jugendlichen wirklich.
2. Ich halte nicht viel von den Erfahrungen der Erwachsenen, ich verlasse mich lieber auf mich selbst.
3. Von gleichaltrigen Freunden/Freundinnen lerne und erfahre ich mehr als von meinen Eltern.
4. Eltern mischen sich dauernd in Sachen ein, die sie nichts angehen.

Als dichotome Kovariaten wurden in dem Modell die Variablen Ost-, Westdeutschland, Geschlecht und Alter verwendet.

In Ostdeutschland fand zwischen 1991 und 1993 ein beträchtliches Maß an sozialem Wandel statt. Der Zusammenbruch traditioneller Institutionen und Normen in Verbindung mit einem Deindustrialisierungsprozeß führte zu einer Situation von Identitätsbedrohung und Arbeitslosigkeit für Jugendliche. Es kann deshalb erwartet werden, daß dieser Tatbestand zu einer Zunahme jugendzentrierter Einstellungen in Ostdeutschland zwischen 1991 und 1993 führte. Diese Veränderung muß jedoch als Periodeneffekt und nicht als Alterseffekt interpretiert werden.

Für Geschlechtsunterschiede in bezug auf eine jugendzentrierte Einstellung können keine Hypothesen formuliert werden. Ein Alterseffekt im Hinblick auf den Wandel einer jugendzentrierten Einstellung hat keine substantielle Bedeutung aus sich selbst heraus, sondern nur in Verbindung mit Statuspassagen, die den Übergang von der Jugend zur Adoleszenz markieren. Das Alter ist stark mit dieser Art von Statuspassagen, von denen man annehmen kann, daß sie vermindernd auf eine jugendzentrierte Einstellung wirken, verknüpft. Zusammenfassend läßt sich feststellen, daß aufgrund der Hypothesen für Jungen und Mädchen keine Unterschiede im Wechsel einer jugendzentrierten Einstellung unterstellt werden, daß in Ostdeutschland eine Zunahme des Jugendzentrismus zwischen 1991 und 1993 als Ergebnis eines bedeutsamen sozialen Wandels vermutet wird, und daß der Jugendzen-

trismus, bedingt durch graduelle Übergänge von der Jugend zum Erwachsenenstatus mit steigendem Alter abnimmt.

**4.2 Ergebnisse**

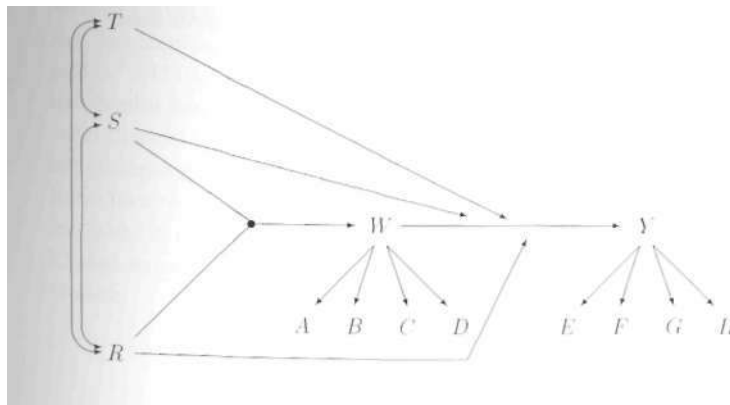
Die Hintergrundvariablen Ost-West, Geschlecht und Alter werden mit R, S und T bezeichnet. Desweiteren werden die zeitspezifischen latenten Variablen Jugendzentrismus mit W und Y benannt, die Indikatoren von W mit A, B, C und D und die Indikatoren von Y mit E, F, G und H. Um die Analyse zu vereinfachen und die beobachteten Häufigkeiten in den Zellen der zu analysierenden multivariaten Kreuztabellen nicht zu stark zu reduzieren, werden die Items der Skala Jugendzentrismus dichotomisiert, wobei 1 Ablehnung und 2 Zustimmung bedeutet. Obwohl diese Vorgehensweise zu einem Verlust von Informationen führt, ist dieser Informationsverlust nicht größer als bei der Durchführung einer LISREL-Analyse, in deren Rahmen die beobachteten Beziehungen zwischen den Variablen vollkommen auf bivariate Zusammenhangsmaße zurückgeführt werden. Die Daten, die für die Schätzung der Modellparameter verwendet werden, bestehen aus der multivariaten Häufigkeitstabelle mit den beobachteten Zellhäufigkeiten  $n_{abcdefghirst}$ .

In seiner allgemeinsten Form ist das modifizierte Pfadmodell mit den latenten Variablen für die oben beschriebenen Variablen gegeben durch

$$\pi_{abcdefghirstwy} = \pi_{rst} \pi_{w|rst} \pi_{y|rstw} \pi_{a|w} \pi_{b|w} \pi_{c|w} \pi_{d|w} \pi_{e|y} \pi_{f|y} \pi_{g|y} \pi_{h|y}.$$

Dieses Modell ist ziemlich komplex, insbesondere dann, wenn man verschiedene Arten von Logit-Restriktionen auf die bedingten Wahrscheinlichkeiten anwenden will, die das Struktur- und Meßmodell beschreiben. Abbildung 1 gibt ein Beispiel für die restringierte Form eines modifizierten LISREL-Modells wieder, das im Rahmen der im Modell verwendeten Variablen formuliert werden kann.

**Abb. 1:** Modifiziertes LISREL-Modell



In diesem Modell sind W und Y latente Variablen mit jeweils vier Indikatoren. Zusätzlich wird angenommen, daß R und S die anfängliche Position (W) beeinflussen, wobei zwischen diesen Variablen ein Interaktionseffekt unterstellt wird. Weiterhin wird postuliert, daß die latenten Übergänge von  $W = w$  zu  $Y = y$  von R, S und T beeinflußt werden. Tatsächlich ist das in Abbildung 1 wiedergegebene Modell dem Endmodell ähnlich, das wir aus der Analyse der Shell-Paneldaten erhielten.

Wegen der Komplexität dieses Modells wurde ein schrittweises Vorgehen in der Analyse befolgt. Zuerst wurden zeitspezifische Meßmodelle untersucht, d.h..

$$\begin{aligned}\pi_{abcd|w} &= \pi_w \pi_{a|w} \pi_{b|w} \pi_{c|w} \pi_{d|w}, \\ \pi_{efgh|y} &= \pi_y \pi_{e|y} \pi_{f|y} \pi_{g|y} \pi_{h|y}.\end{aligned}$$

Danach wurde die Stabilität des Meßmodells getestet, indem eine Analyse unter Verwendung beider Zeitpunkte, aber ohne die Einbeziehung der Hintergrundvariablen, durchgeführt wurde. In anderen Worten, wir schätzen ein Modell folgender Form:

$$\pi_{abcdefgh|wy} = \pi_w \pi_y \pi_{a|w} \pi_{b|w} \pi_{c|w} \pi_{d|w} \pi_{e|y} \pi_{f|y} \pi_{g|y} \pi_{h|y}.$$

Dies Modell stellt uns auch Informationen über den latenten Wechsel zwischen beiden Zeitpunkten zur Verfügung. In einem letzten Schritt wurden schließlich die Beziehungen zwischen den strukturellen Variablen untersucht. Genauer gesagt wird untersucht, ob R (Ost-Westdeutschland), S (Geschlecht) und T (Alter) die Intensität einer jugendzentristischen Einstellung zum ersten Zeitpunkt und die Übergangswahrscheinlichkeiten zwischen dem ersten und zweiten Zeitpunkt beeinflussen.

#### 4.2.1 Getrennte Meßmodelle

Tabelle 1 gibt die Modellanpassungsmaße für die geschätzten Meßmodelle für zwei getrennte Zeitpunkte wieder. Wie man aus ihr ersehen kann, wurden unrestringierte latent class Modelle, *Rasch-Modelle* und *Lord-Birnbaum-Modelle* getestet.

Tabelle 1: Testergebnisse für die geschätzten Modelle

Modell	$L^2$	df	$p(L^2)$	npar	BIC
getrennte Modelle zum Zeitpunkt 1					
1a. 2 Klassen	12.05	6	.061	10	1493
1b. 2 Klassen Rasch	19.03	9	.025	7	1483
1c. 3 Klassen (nicht geordnet)	1.13	1	.288	15	1510
1d. 3 Klassen Rasch (random Nodes)	19.00	7	.008	9	1494
1e. 3 Klassen Rasch	19.00	8	.015	8	1488
1f. 3 Klassen Lord-Birnbaum	11.32	5	.045	11	1498
getrennte Modelle zum Zeitpunkt 2					
2a. 2 Klassen	9.40	6	.152	10	1552
2b. 2 Klassen Rasch	16.29	9	.061	7	1542
2c. 3 Klassen (geordnet)	2.71	1	.100	15	1574
2d. 3 Klassen Rasch (random Nodes)	12.98	7	.073	9	1550
2e. 3 Klassen Rasch	15.49	8	.050	8	1547
2f. 3 Klassen Lord-Birnbaum	6.02	5	.304	11	1555
simultane Modelle zu den Zeitpunkten 1 und 2					
3a. 2 Klassen homogen	294.36	244	.015	12	2998
3b. 2 Klassen heterogen	278.27	236	.031	20	3027
3c. 2 Klassen $\pi_{a w} < > \pi_{e y}$	292.07	242	.015	14	3007
3d. 2 Klassen $\pi_{b w} < > \pi_{f y}$	291.21	242	.017	14	3006
3e. 2 Klassen $\pi_{c w} < > \pi_{g y}$	294.10	242	.012	14	3009
3f. 2 Klassen $\pi_{d w} < > \pi_{h y}$	284.10	242	.033	14	2999
3g. 2 Klassen Rasch homogen	298.62	247	.014	9	2985
3h. 3 Klassen homogen (nicht geordnet)	271.30	235	.052	21	3026
3i. 3 Klassen Rasch homogen (random Nodes)	281.39	241	.038	15	3002
3j. 3 Klassen Rasch homogen	281.91	242	.040	14	2997
3k. 3 Klassen Lord-Birnbaum homogen	277.24	239	.045	17	3009

Wenden wir uns zunächst dem Modell mit zwei latenten Klassen zu. Für beide Zeitpunkte sind die unrestringierten Zwei-Klassen-Modelle (Modelle 1a und 2a) ziemlich gut angepaßt ( $L^2 = 12.05$ ,  $df=6$ ,  $p=.061$  und  $L^2 = 9.40$ ,  $df=6$ ,  $p = .152$ ). Die aus zwei Klassen bestehenden *Rasch-Modelle* (Modelle 1b und 2b) erhält man, indem man die Interaktionen erster Ordnung zwischen den Items gleichsetzt. Dabei ist zu beachten, daß es bei zwei Klassen gleichgültig ist, ob die nodes fixiert sind oder zufällig sind. Die bedingten Likelihood-Ratio-Tests zwischen den Modellen 1a und 1b ( $L^2 = 6.98$ ,  $df = 3$ ,  $p = .073$ ) und den Modellen 2a und 2b ( $L^2 = 6.89$ ,  $df = 3$ ,  $p = .075$ ) besagen, daß für keinen Zeitpunkt die Zwei-Klassen-*Rasch-Modelle* schlechter angepaßt sind als die unrestringierten latent class Modelle.

Die unrestringierten Drei-Klassen-Modelle (Modelle 1c und 2c) sind für alle Zeitpunkte fast perfekt angepaßt. Es ist in diesem Zusammenhang zu beachten, daß als Resultat von Schwierigkeiten, die mit den Grenzen des Parameterraums zusammenhängen (*Titterton; Smith und Makov* 1985) die Zwei- und Drei-Klassenmodelle nicht mit Hilfe der  $L^2$ -Statistik gegeneinander getestet werden können. Es ist weiterhin bedeutsam, daß obwohl die Drei-Klassen-Modelle zum ersten Zeitpunkt gut angepaßt sind, die Klassen nicht in der Weise geordnet werden können, daß die Item-Characteristic-Kurven (ICC) monoton steigend sind (*Croon* 1990). Dies bedeutet, daß in diesem Fall die latente Variable wirklich eine nominale Variable ist.

Im folgenden wurden mehrere Drei-Klassen-Modelle geschätzt, nämlich *Rasch*-Modelle mit latent nodes (Modelle 1d und 2d), *Rasch*-Modelle mit fixierten latent nodes (Modelle 1e und 2e) und *Lord-Birnbaum-Modelle* mit fixierten latent nodes (Modelle 1f und 2f). Zum ersten Zeitpunkt sind beide Drei-Klassen-*Rasch*-Modelle sehr schlecht angepaßt. Die Modelle 1d und 1e sind viel schlechter angepaßt als die unrestringierten Modelle mit drei latenten Klassen und sie haben zudem fast den gleichen  $L^2$ -Wert wie die *Rasch*-Modelle mit zwei Klassen. Das *Lord-Birnbaum-Modell* ist nicht signifikant besser angepaßt als das *Rasch-Modell* ( $L^2 = 6.68$ ,  $df = 3$ ,  $p = 0.083$ ) und zudem ist es signifikant schlechter angepaßt als das unrestringierte Drei-Klassen-Modell ( $L^2 = 1.019$ ,  $df = 4$ ,  $p = .037$ ). Es bleibt festzustellen, daß für den ersten Zeitpunkt das am besten angepaßte Drei-Klassen-Modell das unrestringierte Modell (Modell 1c) ist.

Keines der restringierten Drei-Klassen-Modelle für den zweiten Zeitpunkt ist schlechter angepaßt als das unrestringierte Modell. Dies kann man aus dem bedingten Test zwischen Modell 2c und 2d ( $L^2 = 12.87$ ,  $df = 7$ ,  $p = 0.75$ ), zwischen den Modellen 2c und 2e ( $L^2 = 10.27$ ,  $df = 6$ ,  $p = .114$ ) und zwischen den Modellen 2c und 2e ( $L^2 = 3.31$ ,  $df = 4$ ,  $p = .507$ ) ersehen. Zudem ist das *Rasch*-Modell mit fixierten nodes nicht schlechter angepaßt als das *Rasch-Modell* mit random nodes ( $L^2 = 2.51$ ,  $df = 1$ ,  $p = .113$ ). Und schließlich ist das *Lord-Birnbaum-Modell* besser angepaßt als das *Rasch-Modell* mit fixierten nodes ( $L^2 = 9.47$ ,  $df = 3$ ,  $p = .024$ ). Somit kann für den zweiten Zeitpunkt festgestellt werden, daß das *Lord-Birnbaum-Modell* das am besten angepaßte Drei-Klassen-Modell ist.

Die letzte Spalte von Tabelle 1 berichtet die BIC-Werte für alle Modelle, wobei der BIC definiert ist als  $-2\log\text{-Likelihood} + \log(N) \cdot \text{np}$ . Bei Verwendung des so definierten BIC ist das Modell mit dem niedrigsten BIC-Wert das zu bevorzugende. Aus der Tabelle kann ersehen werden, daß nach dem BIC-Kriterium das Zwei-Klassen-*Rasch*-Modell für beide Zeitpunkte bevorzugt werden sollte.

#### 4.2.2 Simultane Meßmodelle

Der zweite Teil von Tabelle 1 gibt die Testergebnisse für simultane Meßmodelle für zwei Zeitpunkte wieder. Der hauptsächliche Grund für diese Analyse war es zu testen, ob das Meßmodell zwischen zwei Zeitpunkten stabil bleibt. Es ist zu berücksichtigen, daß Schlußfolgerungen über den latenten Wechsel bedeutungslos sind, wenn das Meßmodell nicht als stabil angenommen werden kann.

Die Stabilität des Meßmodells zwischen zwei Zeitpunkten wurde mit Hilfe unrestringierter Zwei-Klassen-Modelle getestet (Modelle 3a bis 3f). Der Vergleich der  $L^2$ -Werte des vollkommen gleichen Zwei-Klassen-Modells (Modell 3a) und des vollkommen unterschiedlichen Zwei-Klassenmodells (3b) zeigt, daß die Unterschiede zwischen den zwei Zeitpunkten gerade signifikant sind ( $L^2 = 16.09$ ,  $df = 8$ ,  $p = .041$ ). Um zu sehen, welche Items für die Unterschiede der Meßmodelle zu beiden Zeitpunkten verantwortlich sind, wurden vier Modelle spezifiziert, in denen jedesmal eine der bedingten Antwortwahrscheinlichkeiten für ein Item variieren durfte (Modelle 3c bis 3f). Die Ergebnisse der Modellanpassungstests zeigen, daß hier das vierte Item (D und H) sich zu beiden Meßzeitpunkte verschieden verhält ( $L^2 = 10.26$ ,  $df = 2$ ,  $p = .006$ ). Jedoch sind die Übergangswahrscheinlichkeiten, wie weiter unten gezeigt werden wird, nicht stark von der Annahme der Homogenität für alle Items beeinflusst. Deshalb unterstellten wir für die weitere Analyse die Homogenität des Meßmodells.

Wie in den getrennten Meßmodellen, war auch hier das Zwei-Klassen-*Rasch*-Modell (3g) nicht schlechter angepaßt als das unrestringierte latent class Modell ( $L^2 = 4.26$ ,  $df = 3$ ,  $p = .235$ ). Des weiteren wurden einige Modelle mit drei latenten Klassen geschätzt. Von diesen Drei-Klassen-Modellen (Modelle 3h bis 3k) verfügt das *Rasch*-Modell mit fixierten latent nodes über die beste Modellanpassung. Es ist nicht schlechter angepaßt als das unrestringierte Modell ( $L^2 = 10.61$ ,  $df = 7$ ,  $p = .157$ ) noch als das *Lord-Birnbaum-Modell* ( $L^2 = 4,67$ ,  $df = 3$ ,  $p = .198$ ). Weil die *Rasch*-Modelle mit zwei und drei Klassen nicht mit Hilfe der Likelihood Statistik verglichen werden können, müssen wir für diesen Vergleich das BIC-Kriterium heranziehen. Nach diesem Kriterium sollte das *Rasch*-Modell mit zwei Klassen bevorzugt werden.

Tabelle 2 gibt die Parameterschätzungen des homogenen Zwei-Klassen-*Rasch*-Modells wieder (Modell 3f). Da das Meßmodell für Y mit dem Meßmodell für W identisch ist, werden diese Parameter in Tabelle 2 nicht ausgewiesen. Die bedingten Antwortwahrscheinlichkeiten weisen daraufhin, daß die erste Klasse die nicht-jugendzentrierte Klasse ist, während die Klasse 2 über eine jugendzentrierte Einstellung verfügt: Personen, die zu Klasse 1 gehören, haben eine viel höhere Wahrscheinlichkeit mit den jugendzentrierten Items nicht übereinzustimmen als Personen, die zu Klasse 2 gehören. So haben beispielsweise Perso-



nen, die zu Klasse 1 gehören, eine Wahrscheinlichkeit von .7338 mit dem Item D nicht übereinzustimmen, während die Wahrscheinlichkeit für Klasse 2 nur bei .2364 liegt. Aus Tabelle 2 kann ersehen werden, daß die Items B, C und D nahezu die gleiche Schwierigkeit besitzen, während Item A viel einfacher ist als die drei anderen Items: Sogar Personen, die zu der nicht jugendzentrierten Klasse gehören, haben eine Wahrscheinlichkeit von .5167 mit dem Item A übereinzustimmen.

Tabelle 2: Parameterschätzungen für das homogene 2-Klassen-Rasch-Modell

$\pi_w$		
W = 1	0.4699	
W = 2	0.5301	
$\pi_{y w}$		
	Y = 1	Y = 2
W = 1	0.9596	0.0404
W = 2	0.5786	0.4214
$\pi_{a w}$		
	A = 1	A = 2
W = 1	0.4833	0.5167
W = 2	0.0951	0.9049
$\pi_{b w}$		
	B = 1	B = 2
W = 1	0.6911	0.3086
W = 2	0.2008	0.7992
$\pi_{c w}$		
	C = 1	C = 2
W = 1	0.7229	0.2771
W = 2	0.2267	0.7733
$\pi_{d w}$		
	D = 1	D = 2
W = 1	0.7338	0.2662
W = 2	0.2364	0.7636

Die anfängliche Verteilung der latenten Variable Jugendzentrismus,  $\pi_w$ , zeigt, daß die beiden latenten Klassen zum ersten Meßzeitpunkt nahezu gleich groß sind. Die Übergangswahrscheinlichkeiten,  $\pi_{y|w}$ , weisen daraufhin, daß sich zwischen den beiden Meßzeitpunkten ein beachtlicher Wechsel vollzieht, besonders bei Personen, die zum ersten Zeitpunkt zu der jugendzentrierten Klasse gehören. Nahezu alle Befragten, die zum ersten Meßzeitpunkt nicht-jugendzentriert waren, bleiben zum zweiten Zeitpunkt in diesem Zustand. Im Gegensatz dazu haben Befragte, die zum ersten Zeitpunkt jugendzentriert waren, eine Wahrscheinlichkeit von .5786, auf eine nicht-jugendzentrierten Position zu wechseln. Als Ergebnis wird die Population weniger jugendzentriert: Zum zweiten Meßzeitpunkt beträgt die Wahrscheinlichkeit, nicht-jugendzentriert zu sein .7626 (= .4699 x .9596 + .5301 x .5786).

Das homogene nicht restringierte Zwei-Klassenmodell (Modell 3a) verfügt nahezu über die gleichen Übergangswahrscheinlichkeiten wie diejenigen, die in Tabelle 2 wiedergegeben werden. Das Modell, in dem das letzte Item, das den Homogenitätstest nicht besteht (D/H), über beide Meßzeitpunkte verschieden sein durfte (Modell 3g), verfügt über eine niedrigere Übergangswahrscheinlichkeit von  $W = 2$  zu  $Y = 1$ , nämlich .4981 anstatt .5786. Deshalb führt die Nichtberücksichtigung der Tatsache, daß das vierte Item zwischen beiden Meßzeitpunkten in seiner Reliabilität variiert, zu einer Überschätzung des tatsächlichen Wechsels.

### 4.2.3 Strukturmodelle

Tabelle 3 gibt die Testergebnisse der geschätzten Strukturmodelle wieder. In allen diesen Modellen wurde ein Zwei-Klassen-Rasch-Modell für die Beziehung zwischen den latenten Klassen  $W$  und  $Y$  und ihren Indikatoren unterstellt. Zunächst wurden einige Referenzmodelle geschätzt. In Modell 1 wurde angenommen, daß sowohl der Übergang von  $W$  zu  $Y$  als auch das Anfangsniveau unabhängig von  $R$ ,  $S$  und  $T$  ist, während Modell 2 das saturierte Strukturmodell darstellt. Mit anderen Worten, geben die Modelle 1 und 2 die untere bzw. obere Grenze des  $L^2$ -Wertes wieder, die man durch die Einführung von Kovariaten in das Modell erreichen kann. In Modell 3 und 4 wurde ein saturiertes Modell für  $\pi_{W|rst}$  und  $\pi_{Y|rstw}$  postuliert, das von der Annahme ausgeht, daß  $W$  und  $Y$  für den anderen Teil des Modells unabhängig von  $R$ ,  $S$  und  $T$  sind. Dieses Modell gibt den unteren  $L^2$ -Wert an, den man erhalten kann, wenn man unterstellt, daß  $W$  und der Übergang von  $W$  zu  $Y$  jeweils getrennt von  $R$ ,  $S$  und  $T$  beeinflußt wird. Hierbei ist zu beachten, daß wegen der geringen Zellenbesetzung die in Tabelle 3 wiedergegebene  $L^2$ -Statistik nicht  $\chi^2$  verteilt ist. Deshalb sind die wiedergegebenen p-Werte bedeutungslos. Jedoch kann der  $L^2$ -Wert immer noch benutzt werden, um hierarchisch geordnete Modelle gegeneinander zu testen.

**Tabelle 3:** Testergebnis für die geschätzten Strukturmodelle unter der Annahme eines homogenen zwei Klassen Rasch Meßmodells

Modell		$L^2$	$df$	$p(L^2)$	npar	BIC
1.	{W} {WY}	963,76	2032	1,000	16	4125
2.	{RSTW} {RSTWY}	938,79	2011	1,000	37	4219
3.	{RSTW} {WY}	953,93	2025	1,000	23	4155
4.	{W} {RSTWY}	946,49	2018	1,000	30	4187
5.	{RSW} {WY}R {RSW <sup>1</sup> } {WY}	957,67	2029	1,000	19	4136
6.	{W} {RWY, SWY, TWY}	958,24	2031	1,000	17	4125
7.	{W} {RWY <sup>2</sup> , SWY <sup>3</sup> , TWY <sup>4</sup> }	949,82	2026	1,000	22	4145
8.	{RSW <sup>1</sup> } {RWY <sup>2</sup> , SWY <sup>3</sup> , TWY <sup>4</sup> }	952,74	2029	1,000	19	4131
9.	{RSW <sup>1</sup> }, $\pi_{w 211} < > \pi_{w rst}$	948,76	2028	1,000	20	4133
10.		948,76	2030	1,000	18	4121

1: Effekt von  $S$  auf  $W$  nur für  $R = 2$   
 2,3 und 4 : Effekt von  $R, S, T$  nur auf  $\pi_{Y|rst}$

Anschließend testeten wir mehrere hierarchische Logit-Modelle für  $\pi_{W|RS}$ . Der einzige Interaktionseffekt zwischen zwei Variablen, der die Modellanpassung signifikant verbesserte, war der Effekt von S auf W. Obwohl die Interaktion erster Ordnung RW nicht mehr signifikant war, verbesserte sich die Modellanpassung wiederum, wenn man den Interaktionsterm zweiter Ordnung  $\mu_{RSW}^{RSW}$  in das Modell einführte. Modell 5 beinhaltet diese 3 Interaktionseffekte. Eine Inspektion der Parameter von Modell 5 zeigt, daß dieser Interaktionseffekt zweiter Ordnung dafür benötigt wurde, die Tatsache zu beschreiben, daß das Geschlecht (S), W nur bei Personen, die in Ostdeutschland leben ( $R = 2$ ) beeinflusst. Modell 6 ist ein nichthierarchisches Modell, daß nur diesen Effekt einschließt. Es ist nicht signifikant schlechter angepaßt als das Modell 3 ( $L^2 = 4.31$ ,  $df = 6$ ,  $p = .635$ ), und ist zudem besser angepaßt als Modell 1 ( $L^2 = 5.52$ ,  $df = 1$ ,  $p = .014$ ). Hieraus wird deutlich, daß es nur in Ostdeutschland einen Geschlechtsunterschied in einer jugendzentrierten Einstellung gibt. Dies war der einzige signifikante Kovariateneffekt auf W.

Als nächstes testeten wir einige hierarchische Modelle für die Effekte von R, S und T auf den Übergang von W zu Y. Modell 7, daß die Interaktion zweiter Ordnung zwischen RWY, SWY und TWY beinhaltet, war das am besten angepaßte Modell. Wie für  $\pi_{W|RS}$ , wurde versucht, sparsamere Modelle zu formulieren, indem zugelassen wurde, daß das Logit-Modell für Y nicht hierarchisch war. Dieses Vorgehen führt zu Modell 8, in dem Ost/Westdeutschland, das Geschlecht und das Alter den Übergang von  $W = 1$  zu  $Y = y$  beeinflusste, aber nicht den Übergang von  $W = 2$  zu  $Y = y$ . Dieses Modell ist nicht schlechter angepaßt als Modell 4 ( $L^2 = 6.25$ ,  $df = 11$ ,  $p = .856$ ) und es ist zudem besser angepaßt als Modell 1 ( $L^2 = 9.02$ ,  $df = 3$ ,  $p = .029$ ). Deshalb hängt der Übergang aus dem Zustand "nicht-jugendzentriert" von den Werten der Kovariaten ab, während der Übergang aus dem Zustand "jugendzentriert" dies nicht tut.

Modell 9 ist eine Kombination aus dem Modell 6 und 8. Es beschreibt einen großen Teil der Variation der abhängigen Variablen W und Y, der durch die unabhängigen Variablen erklärt werden kann. Der Unterschied zwischen den  $L^2$ -Werten vom Modell 1 und 9 beträgt 15.00 unter Verwendung von 4 zusätzlichen Parametern, während der größte  $L^2$ -Zuwachs durch die Einführung der Kovariaten 24.97 mit 21 zusätzlichen Parametern betrug.

Der Effekt des Geschlechts auf W in Ostdeutschland beträgt -0.2444. Dieser Tatbestand deutet darauf hin, daß in Ostdeutschland Jungen ( $S = 1$ ) eine niedrigere Wahrscheinlichkeit besitzen, nicht-jugendzentriert zu sein ( $W = 1$ ) als Mädchen, d.h. Jungen sind jugendzentrierter als Mädchen. Die Effekte von R, S und T auf dem Übergang von  $W = 1$  zu  $Y = y$  sind 1.9646, -2.3430 und -2.0496. Die Werte dieser Parameter sind ziemlich extrem. Diese Parameterschätzungen deuten darauf hin, daß Personen mit  $R = 1$ ,  $S = 2$  und  $T = 2$ , d.h. westdeutsche ältere Mädchen über die höchste Wahrscheinlichkeit verfügen, in einer nicht jugendzentrierten Position zu verharren, während ostdeutsche jüngere männliche Befragte

die höchste Wahrscheinlichkeit besitzen, von einer nicht-jugendzentrierten Position ( $W = 1$ ) zu einer jugendzentrierten ( $Y = 2$ ) zu wechseln.

Schließlich wurde ein Modell geschätzt, in dem  $\pi_{y|rstw}$  nicht mit Hilfe eines Logit-Modells parametrisiert wurde, sondern mit Gleichheitsrestriktionen, die den Wahrscheinlichkeiten auferlegt wurden. Der Grund für dieses Vorgehen lag darin, daß eine Inspektion der bedingten Wahrscheinlichkeiten aus Modell 9 für alle Kombinationen von R, S und T eine Wahrscheinlichkeit von nahe 1 für das Verharren im Zustand  $W = 1$  ergab, außer für ostdeutsche Jugendliche der jüngsten Altersgruppe. Diese Gruppe besitzt eine Wahrscheinlichkeit von nahezu 50%, von einer nicht-jugendzentrierten in eine jugendzentrierte Position zu wechseln. Offensichtlich beschreiben die extremen Logitparameter aus Modell 9 diese "Ausreißer". Diese Vermutung wird bestätigt durch die Testergebnisse von Modell 10, in dem die Wahrscheinlichkeit, den Zustand  $W = 1$  zu verlassen für alle Ausprägungen von R, S und T als gleich angenommen wurde, außer für die Gruppe  $R = 2, S = 1$  und  $T = 1$ . Modell 10 verfügt exakt über den gleichen  $L^2$ -Wert wie Modell 9. Nach dem BIC-Kriterium ist Modell 10 das am besten angepaßte Modell. Für eine Beschreibung der Kovariateneffekte auf W und auf den Übergang von W zu Y werden im Vergleich zu Modell 1 nur zwei zusätzliche Parameter benötigt, nämlich ein log-linearer Parameter, der die Tatsache beschreibt, daß ostdeutsche Jungen jugendzentrierter eingestellt sind als ostdeutsche Mädchen und eine bedingte Wahrscheinlichkeit, die berücksichtigt, daß ostdeutsche jüngere männlich Befragte über eine höhere Wahrscheinlichkeit verfügen, von einer nicht-jugendzentrierten zu einer jugendzentrierten Einstellung zu wechseln als andere Personen.

## 5. Diskussion

In diesem Artikel wurde ein allgemeiner Ansatz für die Analyse von kategorialen Paneldaten dargestellt. Am Beispiel der Einstellung "Jugendzentrismus" konnte der Vorteil log-linearer Kausalmodelle mit latenten Variablen deutlich demonstriert werden. Ein restringiertes latent class Modell wurde benutzt, um für die latente Variable Jugendzentrismus ein semi-parametrisches *Rasch*-Modell zu konstruieren, das zu zwei Zeitpunkten gemessen wurde. Der Wechsel in einer jugendzentrierten Einstellung zwischen den beiden Wellen wurde mit Hilfe einer latenten Veränderungstabelle dargestellt. Zusätzlich wurde sowohl für den anfänglichen Zustand als auch für die Übergangswahrscheinlichkeiten zwischen den beiden Zeitpunkten mit Hilfe eines Logit-Modells der Einfluß von Kovariaten untersucht.

Es ist interessant, die in diesem Ansatz dargestellte Information über latenten Wechsel mit den Ergebnissen zu vergleichen, die mit Hilfe einer LISREL-Analyse erhalten worden wären. Es kann erwartet werden, daß innerhalb eines LISREL-Modells eine geringe Stabilität der jugendzentristischen Einstellung gefunden worden wäre. Außerdem wäre es unter Ver-

wendung von latenten Mittelwerten kein Problem gewesen, die Tendenz zu einer Abnahme jugendzentrister Einstellung herauszufinden. Diese Vermutung wurde im Rahmen von LISREL-Mehrgruppenvergleichen bestätigt. Jedoch liefert die Beschreibung des Veränderungsprozesses mit Hilfe von log-linearen Kausalmodellen mit latenten Variablen mehr Information, da die latente Veränderungstabelle exakte Informationen über die Richtung des Wechsels für Personen mit einer spezifischen Anfangsposition gibt. In dieser Anwendung konnte festgestellt werden, daß, außer für ostdeutsche jüngere männliche Befragte, Personen mit einer nicht jugendzentrierten Einstellung zumeist in ihrer Anfangsposition verharrten, während viele jugendzentrierte Personen sich in Richtung auf eine nicht-jugendzentrierte Position bewegten.

Natürlich hat der modifizierte LISREL-Ansatz, verglichen mit dem LISREL-Modell auch einige Nachteile. Obwohl die gegenwärtige Arbeitsversion des Programms **LEM** wesentlich effizienter ist als die vorhergehende Version, ist die Anzahl der innerhalb eines Problems bearbeitbaren Variablen immer noch erheblich kleiner als in LISREL. Deshalb ist die Größe des Problems, das bearbeitet werden kann, begrenzt. Ein anderes Problem der log-linearen Analyse liegt in der Tatsache, daß bei geringen Zellenbesetzungen die theoretische  $\chi^2$ -Annäherung des Likelihood-Ratio-Tests nicht zufriedenstellend ist. Obwohl in einer solchen Situation die Signifikanz der Parameter immer noch mit Hilfe des Likelihood-Ratio-Tests überprüft werden kann, kann die Modellanpassung nicht mehr getestet werden (Agresti 1990).

#### Literatur:

**Agresti, A., 1990:**  
Categorical data analysis.  
New York: Wiley.

**Bartholomew, D.J., 1987:**  
Latent variables models and factor analysis.  
London: Griffin.

**Bishop, R.J.; Fienberg, S.E.; Holland, P.W., 1975:**  
Discrete multivariate analysis: theory and practice.  
Cambridge, Mass.: MIT Press.

**Clogg, C.C., 1981:**  
New developments in latent structure analysis. In: *D.J. Jackson; E.F. Borgotta* (Hrsg.), Factor analysis and measurement in sociological research, S. 215-246.  
Beverly Hills: Sage Publications.

**Clogg, C.C., 1982:**  
Some models for the analysis of association in multiway cross-classifications having ordered categories.  
In: Journal of the American Statistical Association, 77, S. 803-815.

**Clogg, C.C.; Goodman, L.A., 1984:**  
Latent structure analysis of a set of multi-dimensional contingency tables.  
In: Journal of the American Statistical Association, 79, S. 762-771.

**Croon, M., 1990:**  
Latent class analysis with ordered latent classes.  
In: British Journal of Mathematical and Statistical Psychology, 43, S. 171-192.

**Dempster A.P.; Laird N.M.; Rubin, D.B.**, 1977:

Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion)

In Journal of the Royal Statistical Society, Ser. B., 39, S. 1-38.

**Formann, A.K.**, 1992:

Linear logistic latent class analysis for polytomous data

In: Journal of the American Statistical Association, 87, S. 4/6-486.

**Georg, W.**, 1992:

Die Skala Jugendzentrismus im Zeitreihen- und Kulturvergleich

In Jugendwerk der deutschen Shell (Hrsg.), Jugend '92. Lebenslagen, Orientierungen und Entwicklungsperspektiven im vereinten Deutschland. Band 4: Methodenberichte, Tabellen, Fragebogen. S. 15-16.

Opladen: Leske und Budrich.

**Goodman, L.A.**, 1972:

A modified multiple regression approach for the analysis of dichotomous variables.

In: American Sociological Review, 37, S. 28-46.

**Goodman, L.A.**, 1973:

The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach.

In: Biometrika 60, S. 179-192.

**Goodman, L.A.**, 1974:

Exploratory latent structure analysis using both identifiable and unidentifiable models.

In: Biometrika, 61, S. 215-231.

**Goodman, L.A.**, 1979:

Simple models for the analysis of association in cross-classifications having ordered categories.

In: Journal of the American Statistical Association, 74, S. 537-552.

**Haberman, S.J.**, 1979:

Analysis of qualitative data. Vol. 2, New developments. New York: Academic Press.

**Hagenaars, J.A.**, 1988:

Latent structure models with direct effects between indicators: local dependence models.

In: Sociological Methods and Research, 16, S. 379-405.

**Hagenaars, J.A.**, 1990:

Categorical longitudinal data - loglinear analysis of panel, trend and cohort data.

Newbury Park: Sage.

**Hagenaars, J.A.**, 1992:

Exemplifying longitudinal loglinear analysis with latent variables. In: *P.G.M. Van der Heijden; W. Jansen, B.*

*Francis, G.U.H. Seeber* (Hrsg.): Statistical modelling, S. 105-120.

Amsterdam: Elsevier Science Publishers B.V.

**Hagenaars, J.A.**, 1993:

Loglinear models with latent variables

Newbury Park: CA: Sage.

**Heinen, A.**, 1993:

Discrete latent variables models.

Tilburg: Tilburg University Press.

**Joreskog, K.G.; Sorbom, D.**, 1988:

Lisrel 7: a guide to the program and applications.

**Jugendwerk der deutschen Shell**, 1995:

Jugend '81 Lebensentwürfe, Alltagskulturen, Zukunftsbilder. Band 1.

Hamburg: Jugendwerk der deutschen Shell.

**Lazarsfeld, P.F.; Henry, N.W.**, 1968:

Latent structure analysis. Boston: Houghton Mill

**Lindsay, B.; Clogg, C.C.; Grego, J.**, 1991:

Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis.

In: Journal of the American Statistical Association, 86, S. 96-107.

**Little, R.J.; Rubin, D.B.**, 1987:

Statistical analysis with missing data.

New York: Wiley.

- Mare, R.D.; Winship, C.**, 1991:  
Loglinear models for reciprocal and other simultaneous effects. In: *C.C. Clogg* (Hrsg.): *Sociological Methodology* 1991, S. 199-234.  
Oxford: Basil Blackwell.
- Masters, G. N.**, 1982:  
A Rasch model for partial credit scoring.  
In: *Psychometrika*, 47, S. 149-174.
- McCutcheon, A.L.**, 1988:  
Sexual morality, pro-life values and attitudes toward abortion.  
In: *Sociological Methods and Research*, 16, S. 256-275.
- Meng, X.L.; Rubin, D.B.**, 1993:  
Maximum likelihood estimation via the ECM algorithm: A general framework.  
In: *Biometrika*, 80, S. 267-278.
- Van de Pol, E; De Leeuw, J.**, 1986:  
A latent Markov model to correct for measurement error.  
In: *Sociological Methods and Research*, 15, S. 118-141.
- Van de Pol, E; Langeheine, R.**, 1990:  
Mixed Markov latent class models. In: *C.C. Clogg* (Hrsg.): *Sociological Methodology* 1990.  
Oxford: Basil Blackwell.
- Schelsky, H.**, 1957:  
Die skeptische Generation. Eine Soziologie der deutschen Jugend.  
Düsseldorf, Köln: Eugen Diederichs Verlag.
- Rasch, G.**, 1960:  
Probabilistic models for some intelligence and attainment tests.  
Copenhagen: Nielson and Lydicke.
- Rost, J.**, 1990:  
Rasch models in latent classes: an integration of two approaches to item analysis.  
In: *Journal of Applied Psychological Measurement*, 14, S. 271-282.
- Rost, J.; Georg, W.**, 1991:  
Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel zur Skala "Jugendzentrismus".  
In: *ZA-Information*, 28, S. 52-74.
- Titterton, D.M.; Smith, A.E; Makov, U.E.**, 1985:  
Statistical analysis of finite mixture distributions.  
Chichester: John Wiley & Sons.
- Vermunt, J.K.**, 1993:  
LEM: log-linear and event history analysis with missing data using the EM algorithm.  
WORC PAPER 93.09.015/7.  
Tilburg: University.
- Vermunt, J.K.**, 1994:  
Causal log-linear modeling with missing data. WORC PAPER 94.05.021/7.  
Tilburg: University.
- Vermunt, J.K.**, 1995:  
Log-linear event history analysis; a general approach with missing data, latent variables, and unobserved heterogeneity.  
Tilburg: Tilburg University Press.
- Wiggins, L.M.**, 1973:  
Panel analysis.  
Amsterdam: Elsevier.
- Xie, Yu**, 1992:  
The log-multiplicative layer effects model for comparing mobility tables.  
In: *American Sociological Review*, 57, S. 380-395.
- Zinnecker, J.**, 1985:  
Die Jugendstudien von EMNID/Shell 1953-1955. Zur Archäologie repräsentativer Jugendforschung im Nachkriegsdeutschland und zugleich zu einigen Schwierigkeiten der Wiederholung solcher Studien. In: *Jugendwerk der deutschen Shell* (Hrsg.): *Jugendliche und Erwachsene '85. Generation im Vergleich*, Band 3: *Jugend der fünfziger Jahre und heute*, S. 408-480.  
Opladen: Leske und Budrich.

## Die Variable "Beruf" in der empirischen Haushalts- und Familienforschung

### Zur Ausschöpfung relevanter Informationsanteile aus standardisierten Berufsklassifikationssystemen

von Ralf Bohrhardt und Wolfgang Voges<sup>1</sup>

#### Zusammenfassung

*Der vorliegende Beitrag erläutert zunächst die Bedeutsamkeit der Variable "Beruf für die empirische Haushalts- und Familienforschung. Dies geschieht vor dem Hintergrund der theoretischen Konzepte der Lebensstilforschung und der Theorie soziokultureller Ungleichheit. Nach einer Erläuterung der Interpretations- und Komparabilitätsproblematik bestehender Berufsklassifikationssysteme wird der Versuch unternommen, entlang der Analysedimension "kulturelles Kapital" berufsbezogene Angaben aus sich bedeutsam unterscheidenden Klassifikationssystemen für die (international vergleichende) Haushalts- und Familienforschung empirisch nutzbar zu machen. Die Konstruktion und Überprüfung einer Skala berufsgebundenen kulturellen Kapitals erfolgt am Beispiel des Familiensurveys des Deutschen Jugendinstitutes von 1988 und des National Survey of Families and Households des Center for Demography and Ecology an der University of Wisconsin-Madison von 1987/88<sup>2</sup>*

#### Abstract

*This article comments firstly on the importance of the variable "occupation" for household- and family research within the theoretical framework of life style research and the theory of socio-cultural inequality. Secondly, it explains the problems of interpreting and comparing existing occupational classification schemes. Thirdly, a scale is introduced for making the values of different occupational classification schemes comparable and interpretable in a relevant way for household- and family research using "cultural capital" as the main dimension of a reclassification. The construction and testing of the scale is done with data from the 1988 Familiensurvey des Deutschen Jugendinstitutes (Family Survey conducted by the German Youth Institute) and the 1987/88 National Survey of Families and Households conducted by the Center for Demography and Ecology of the University of Wisconsin-Madison.*

1 **Ralf Bohrhardt**, cand. rer. soc., ist Projektmitarbeiter am Zentrum für Sozialpolitik der Universität Bremen, Parkallee 39, D-28209 Bremen.

PD Dr. **Wolfgang Voges**- ist wissenschaftlicher Angestellter in der Abteilung "Institutionen und Geschichte des Wohlfahrtsstaates" am Zentrum für Sozialpolitik und Mitglied des DFG-Sonderforschungsbereiches 186 der Universität Bremen: "Statuspassagen und Risikolagen im Lebensverlauf".

2 Für Anregungen und Hinweise danken wir **Greg Duncan**, **Walter Heinz** und **Wolfgang Teckenberg**.