

Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel der Skala "Jugendzentrismus"

Rost, Jürgen; 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:

Rost, J., & Georg, W. (1991). Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel der Skala "Jugendzentrismus". *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 28, 52-74. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-202475>

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.



Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel der Skala "Jugendzentrismus"

von Jürgen Rost und Werner Georg¹

1. Probabilistische Skalierungsmodelle

Im Bereich der Entwicklung sozialwissenschaftlicher und psychologischer Skalen zeigt sich nach wie vor eine Dominanz der klassischen Testtheorie (vgl. *Lord/Novick* 1968). So stellt *Schulze* (1990) fest, daß beispielsweise im *ZUMA-Handbuch* sozialwissenschaftlicher Skalen (1988) die mit Hilfe der klassischen Testtheorie entwickelten Skalen eine konkurrenzlose Position einnehmen.

Dieser Tatsache gegenüber steht die statistischen Modellbildung auf diesem Gebiet. Bereits Anfang der 60er Jahre entwickelte *Rasch* (1960) eine probabilistische Testtheorie, die sich zunächst auf dichotome Antwortvariablen bezog. Es ist ein sehr simples Modell, das - wie die meisten statistischen Modelle - eine additive Zerlegung der zu analysierenden Variablen anstrebt. Da diese Variablen, die Itemantworten, kategorial sind und ihre Wahrscheinlichkeiten auf das 0-1-Intervall beschränkt sind (und daher auch nicht additiv zerlegt werden können), wird die additive Zerlegung auf deren Logits bezogen:

$$(1) \quad \ln \left(\frac{p(x_{vi})}{1 - p(x_{vi})} \right) = \theta_v + \alpha_i$$

wobei $p(x_{vi})$ die Wahrscheinlichkeit von Person v ist, bei Item i die Antwortkategorie $x=1$ zu wählen, θ_v die Position der Person auf der (unbeobachteten) Eigenschaftsskala angibt und α_i die *Leichtigkeit* des Items, d.h. die Lage des Items auf der Eigenschaftsskala. *Rasch* hat gezeigt, daß eine derartige additive Zerlegung ("latent additivity") gewisse vorteilhafte Modelleigenschaften, wie die Separierbarkeit der Modellparameter gewährleistet. Dies ist nicht weiter überraschend, denn die meisten statistischen Modelle - von der Varianzanalyse bis zu log-linearen Modellen - formulieren ebenfalls (und nicht ohne Grund) ein

¹ PD Dr. Jürgen Rost ist wissenschaftlicher Mitarbeiter im Institut für die Pädagogik der Naturwissenschaften (IPN) an der Universität Kiel, Olshausenstr. 62, 2300 Kiel, Tel. 0431 880 3146.
Dr. Werner Georg ist wissenschaftlicher Mitarbeiter im Fachbereich 2 (Sozialpädagogik) an der Universität - Gesamthochschule Siegen, Adolf-Reichwein-Str. 2, 5900 Siegen, Tel. 0271 740 4493.

additives Zusammenwirken der an der Beobachtung einer manifesten Variable beteiligten Komponenten. Die o.g. Modellgleichung des Raschmodells läßt sich in die geläufigere Form

$$(2) \quad p(x_{ij}) = \frac{\exp(\theta_j + \alpha_i)}{1 + \exp(\theta_j + \alpha_i)}$$

bringen. Stellt man die Beziehung zwischen latenter Variable und Antwortwahrscheinlichkeit, wie sie in dieser Gleichung spezifiziert ist, graphisch dar (sog. Item Characteristic Curve, ICC), so zeigt sich eine "fast" lineare Beziehung (s. Abb.1).

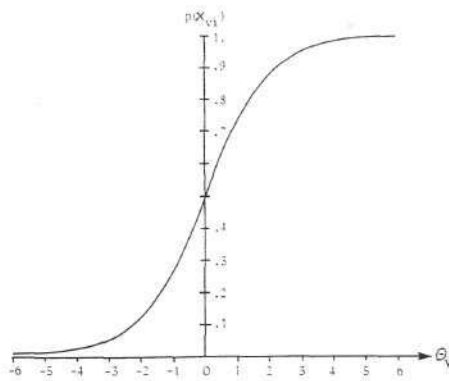


Abb. 1: Die Itemcharakteristik des Raschmodells

Die Linearität gilt speziell im mittleren Bereich der Antwortwahrscheinlichkeiten zwischen 0.2 und 0.8. An den "Enden" ist die Kurve zu Asymptoten "umgebogen", was auch notwendig ist, da die Antwortwahrscheinlichkeit in ihrem Wertebereich begrenzt ist, die latente Variable jedoch nicht. Das Raschmodell realisiert somit genau die Annahme des klassischen Modells essentiell tau-äquivalenter Messungen, nämlich die additive Aufspaltung in eine "Truescore"-Variable θ_j und die Itemleichtigkeit α_i . Der Unterschied besteht darin, daß diese Aufspaltung im Raschmodell in einer mathematisch sinnvollen Weise erfolgt (nämlich nach einer Logit-Transformation), während die klassische Testtheorie, sofern man sie auf dichotome Items anwendet, die Wahrscheinlichkeiten selbst zerlegt, was höchst problematisch ist.

Eine Konsequenz aus obiger Modellgleichung, welche oft als der "Kern" des Raschmodells angesehen wird, besteht darin, daß sich die ICCs aller Items eines Tests nicht überschneiden, ja sogar parallel verlaufen (vgl. Abb.2).

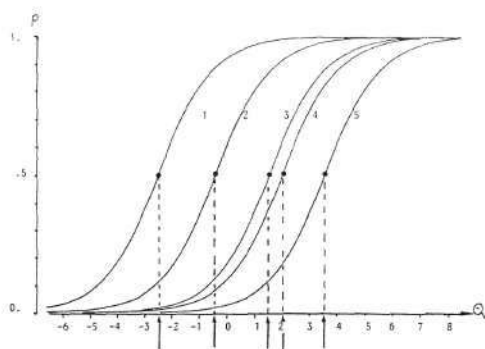


Abb. 2: Die ICCs mehrerer Items eines Tests

Dies wird deshalb oft als eine besonders bemerkenswerte und restriktive Eigenart des Raschmodells angesehen, weil man über eine unterschiedliche Steigung der ICCs auch unterschiedliche Itemtrennschärfen erfassen könnte. So hat das Item 1 in Abb. 3 eine größere Trennschärfe als Item 2, weil es aufgrund der größeren Steigung besser zwischen niedrigen und hohen Eigenschaftsausprägungen trennt.

Die philosophischen Grundlagen der Meßtheorie von *Rasch*, die mit dem Begriff der **spezifischen Objektivität** verbunden sind, verbieten in der Tat solche Überschneidungen der ICCs schon aus logischen Gründen: Die Skalierung mittels eines Fragebogens stellt - wie auch viele Meßvorgänge in der Physik - einen Meßvorgang dar, an dem zwei unterschiedliche Mengen von Meßobjekten beteiligt sind (die Personen und die Items), und bei dem sich die beobachteten Daten stets nur aus der Interaktion von je einem Element jeder Menge ergeben (die Reaktion einer Person auf ein Item). Die Idee spezifisch objektiver Messungen besagt nun, daß die Meßwerte der Elemente der einen Menge unabhängig davon sein müssen, welche Elemente der anderen Menge gerade für den Meßvorgang ausgewählt wurden. Das bedeutet, daß die Skalenwerte der Personen unabhängig davon sein müssen, ob eher leichte oder schwere Items zu einem Konstrukt im Test enthalten sind, und daß die Skalenwerte der Items unabhängig davon sein müssen, ob Personen mit eher niedriger oder eher hoher Eigenschaftsausprägung in der Stichprobe enthalten sind.

Die Forderung nach spezifisch objektiven Messungen ist eigentlich eine pure Selbstverständlichkeit, nicht nur in der Wissenschaft, sondern auch im Alltagsleben (denn wer würde schon akzeptieren, daß ein Päckchen auf der Waage im Postamt mehr wiegt als zu Hause auf der Küchenwaage?). Betrachtet man Abbildung 3, so stellt man fest, daß Item 1 für Person v leichter ist, das heißt eine höhere Antwortwahrscheinlichkeit hat, als Item 2, während Item 1 für Person w schwerer ist als Item 2. Je nachdem wie die Personenstichprobe zusammengesetzt ist, würde man also unterschiedliche Skalenwerte für die Items erhalten. Dies widerspricht der Forderung nach spezifisch objektiven Messungen und ist genau der Grund dafür, daß im Raschmodell alle ICCs parallel sind.

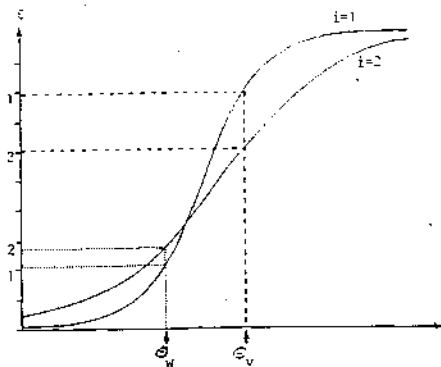


Abb. 3: Zwei ICCs mit unterschiedlicher Trennschärfe

Bereits zehn Jahre vor der ersten Publikation des Raschmodells hatte *Lazarsfeld* mit der "latent class analysis" (LCA) ebenfalls ein probabilistisches Modell entwickelt, das jedoch über keine quantitative latente Personenvariable verfügte, sondern die Versuchspersonen qualitativ in Klassen hinsichtlich ihrer unterschiedlichen Antwortwahrscheinlichkeiten einteilte (vgl. *Lazarsfeld* 1950, *Lazarsfeld und Henry* 1968). Die Antwortwahrscheinlichkeiten sollen hier nicht additiv zerlegt, sondern durch klassenspezifische Antwortwahrscheinlichkeiten erklärt werden:

$$(3) \quad P(*vi) = \sum_g j_{ig}^* \cdot j_g^*$$

wobei die Wahrscheinlichkeitsparameter j_g^* und j_{ig}^* die relative Größe (Klassenwahrscheinlichkeit) der latenten Klasse g , bzw. die Leichtigkeit (Antwortwahrscheinlichkeit) des Items i in Klasse g parametrisieren. Der Vorteil der LCA gegenüber dem Rasch-Modell besteht darin, daß man die Modellannahme der **Personenhomogenität** (für alle Personen

müssen dieselben Itemparameter gelten) durch die Einführung von Klassen mit unterschiedlichen Itemparametern lockern kann. Diese Annahme der Personenhomogenität ist relativ selten erfüllt. So fanden *Krebs* und *Schüssler* (1987) bei einem interkulturellen Skalenvergleich keine Skala, die den Modellannahmen des Raschmodells (auf 5% Wahrscheinlichkeitsniveau) entsprochen hätte.

In der LCA werden dagegen den einzelnen latenten Klassen keinerlei Restriktionen hinsichtlich der Itemparameter auferlegt, d.h. in jeder Klasse dürfen gänzlich unterschiedliche Itemschwierigkeiten gelten. Der "Preis" für diese Flexibilität des Modells liegt darin, daß alle Personen innerhalb einer latenten Klasse dieselben Lösungswahrscheinlichkeiten haben müssen, d.h. es gibt keine individuellen Abstufungen einer Eigenschaftsausprägung mehr, wie im Raschmodell. Das Ziel einer LCA besteht also darin, die Personen in (latente) Klassen einzuteilen, die dadurch definiert sind, daß die Personen innerhalb der Klassen dieselben, und zwischen den Klassen verschiedene Lösungswahrscheinlichkeiten haben.

Das Ergebnis einer LCA läßt sich dann anschaulich durch die "latenten Profile" darstellen, die für jede Klasse die Antwortwahrscheinlichkeiten für die Items angeben (vgl. Abb. 4).

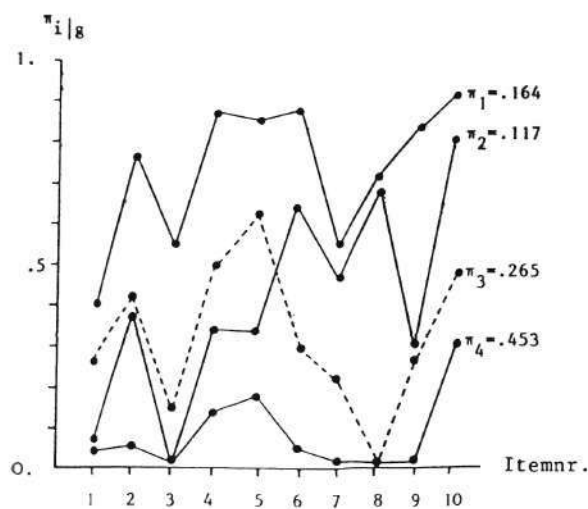


Abb. 4: Beispiel für die latenten Profile einer 4-Klassen-Lösung für 10 Items



Im Unterschied zu clusteranalytischen Verfahren handelt es sich um keine "manifeste Klassifikation", bei der jede Person genau einer oder mehreren Klassen zugeordnet wird, sondern jede Person gehört jeder Klasse mit einer bestimmten Wahrscheinlichkeit an (die gemessene Nominalvariable ist **latent**). Diese Wahrscheinlichkeit läßt sich aus o.g. Modellgleichung ableiten und lautet:

$$(4) \quad p(g|x) = \frac{\pi_g p(x|g)}{\sum_g \pi_g p(x|g)}$$

wobei die Patternwahrscheinlichkeit $p(x|g)$ von Klasse g gleich dem Produkt aller klassenspezifischen Antwortwahrscheinlichkeiten ist:

$$(5) \quad p(x|g) = \prod_i \pi_{ig}$$

Sollen Personen aufgrund einer LCA tatsächlich in Klassen eingeteilt werden, so ordnet man jede Person derjenigen Klasse zu, der sie am wahrscheinlichsten angehört (Modalwert). Die mittlere Höhe dieser Modalwerte kann gleichzeitig als Indikator für die Eindeutigkeit des Analyseergebnisses gewertet werden.

Das Raschmodell und die LCA sind sich recht ähnlich. Man kann sagen, daß sie sich nur im **Skalenniveau von der latenten Variable** unterscheiden, das einmal metrische Qualität (Raschmodell) und einmal Nominalqualität (LCA) hat. Die Ähnlichkeit erstreckt sich über mehrere Punkte:

1. In beiden Modellen wird direkt die Wahrscheinlichkeit einer Person, die Antwortkategorien jedes Items zu wählen, parametrisiert, d.h. es handelt sich um probabilistische Modelle, in denen die **Wahrscheinlichkeiten qualitativer Beobachtungen** Gegenstand der Modellbildung sind, und nicht (wie in der klassischen Testtheorie) "vorfindbare", intervallskalierte, wenn auch fehlerbehaftete Meßwerte.
2. Es handelt sich um Modelle, mit denen multivariate Abhängigkeiten in einer **mehrdimensionalen Kontingenztabelle** erklärt werden sollen, und zwar durch Einführung einer **latenten Variable**. Wenn man so will, handelt es sich (wie bei LISREL-Modellen) um latente Variablen Modelle, jedoch für Nominaldaten. Anders als LISREL, wird nicht von den bivariaten Assoziationen zwischen den manifesten



Variablen ausgegangen, sondern von den vollständigen Häufigkeiten der Kontingenztafel, in denen noch die Wechselwirkungen höchster Ordnung zwischen den manifesten Variablen enthalten sind. Dieser theoretische Vorteil wird jedoch mit praktischen Problemen erkauft, die die Anwendbarkeit (insbesondere der LCA) bei großen Itemmengen betreffen (Parameterschätzung und Modellkontrolle, s.u.).

3. In beiden Ansätzen wird die lokale stochastische Unabhängigkeit der Itemantworten zum Optimierungskriterium gemacht, d.h. sie wird nicht (ungeprüft) vorausgesetzt, sondern die latente Variable (nominal oder metrisch) wird so geschätzt, daß die manifesten Antwortvariablen bei konstant gehaltener latenter Variable (daher lokal) stochastisch unabhängig sind :

$$(6) \quad p(x_{vi} \text{ und } x_{vj} | \theta) = p(x_{vi} | \theta) \cdot p(x_{vj} | \theta)$$

wobei im Falle der LCA die latente Variable θ die Klassennummern als Werte annimmt: $\theta=1,2,\dots,h$. Mit anderen Worten, die beobachtbaren Kontingenzen zwischen den Itemantworten werden dadurch erklärt, daß eine latente Variable am Zustandekommen aller Itemantworten beteiligt ist. Hält man diesen Einfluß konstant, so besteht Unabhängigkeit zwischen den Items. Läßt sich keine metrische oder nominale latente Variable finden, unter der stochastische Unabhängigkeit gilt, so müßte sich dies in einer schlechten Modellanpassung niederschlagen. Insofern ist die lokale stochastische Unabhängigkeit keine lästige, weil zu restriktive a priori Annahme, sondern ein sinnvolles Optimierungskriterium.

4. Da sich beide Modellansätze auf dieselbe Datenstruktur, nämlich die Häufigkeiten der Antwortmuster, beziehen und beide Modelle Maximum-Likelihood-Schätzungen erlauben, sind auch Vergleiche der Modellanpassung zwischen beiden Modelltypen möglich. Das bedeutet, daß empirisch entschieden werden kann, ob für einen Test die Annahme einer quantitativen oder einer kategorialen latenten Variablen eher angemessen ist.

Hierfür wurden in den letzten Jahren verschiedene informationstheoretische Indices entwickelt. Neben dem "klassischen" AIC (Akaike's Information Criterion) scheint sich der BIC (Best Information Criterion, vgl. *Bozdogan* 1987) als der geeignetere zu erweisen. Er setzt die Likelihood der Daten zur Anzahl der Modellparameter in Beziehung:

$$(7) \quad \text{BIC} = -2\ln(L) + \ln(N)k$$

wobei L das Maximum der Likelihoodfunktion, N der Stichprobenumfang und k die Anzahl unabhängiger Modellparameter ist. Ein kleinerer BIC-Wert zeigt dabei die bessere Modellgeltung an. Anschaulich gesprochen, wird die größere Likelihood eines "besser" passenden Modells mit der Anzahl der Parameter verrechnet, mit der die höhere Likelihood "erkauf" wurde. Der "penalty term" enthält den Faktor $\ln(N)$, was bedeutet, daß für größere Stichproben auch der Anstieg der Likelihood für jeden zusätzlichen Parameter größer sein muß, damit ein Modell nach dem BIC besser paßt.

Inferenzstatistische Vergleiche zwischen beiden Modelltypen, etwa mittels Likelihoodquotiententests (LQT) sind dagegen aus zwei Gründen nicht möglich. Erstens stehen Raschmodell und LCA nicht direkt in einer hierarchischen Beziehung zueinander, was Voraussetzung für die Durchführung von LQT ist. Zweitens sind die Voraussetzungen für die asymptotischen Eigenschaften von LQT meistens nicht gegeben, da die Anzahl möglicher Antwortmuster bei realistischen Itemanzahlen so groß wird, daß deren erwarteten Häufigkeiten weit unter 1 liegen.

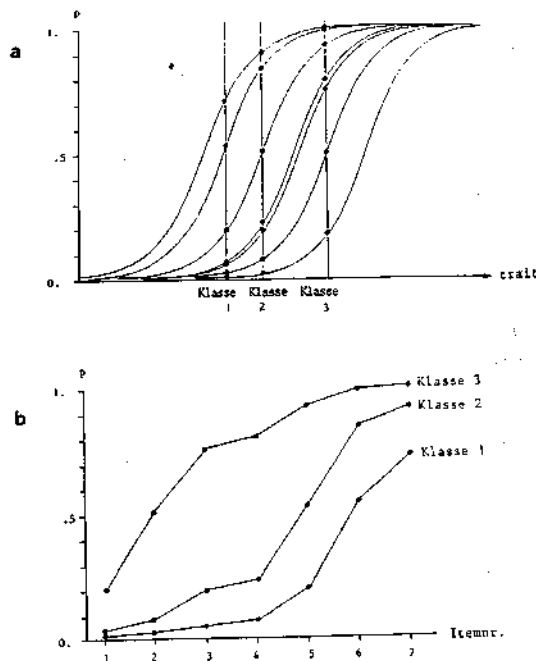


Abb. 5: Die latenten Profile für drei, auf einem Rasch-Kontinuum lokalisierte Klassen

Wenn beide Modelle in ihrer unrestringierten Form auch nicht in einer hierarchischen Relation stehen, so gibt es doch den Fall, daß beide ineinander übergehen, bzw. das eine ein Spezialfall vom anderen ist. Dieser Fall ist gegeben, wenn die latenten Profile einen Verlauf haben, der mit den ICCs des Raschmodells kompatibel ist. Anders ausgedrückt: wenn die latenten Klassen lediglich Punkte auf einem latenten Kontinuum markieren und die Items "in Wirklichkeit" eine Raschskala bilden. Dies ist in Abb. 5 veranschaulicht. Dort ist dargestellt, wie die latenten Profile von drei Klassen aussehen, die sich auf einem Rasch-Kontinuum anordnen. Sie ergeben sich durch die Schnittpunkte mit den ICCs. Es wird deutlich, daß man am Verlauf der latenten Profile erkennen kann, ob ein Test Rasch-skalierbar ist, jedoch ist die genaue Beziehung zu kompliziert um sie "mit dem Auge" erkennen zu können. Einfacher ist jedoch die folgende, graphisch ebenfalls leicht einsehbare, ein-eindeutige Beziehung: dann und nur dann wenn die latenten Profile überschneidungsfrei sind, gibt es ein latentes Kontinuum, bezüglich dessen alle Items monotone (wenn auch nicht überschneidungsfreie) ICCs haben. Die metrische Skalierbarkeit von Items läßt sich also am Ergebnis einer LCA ablesen: überschneiden sich die latenten Profile, so ist eine Skalierbarkeit ausgeschlossen. Von daher ist die LCA ein Auswertungsverfahren, daß auch vor metrischen Skalierungsversuchen gewinnbringend angewandt werden kann.

2. Geordnete Antwortkategorien

Da die meisten Daten im Bereich von Meinungen, Einstellungen, Werthaltungen, Interessen und Ähnlichem mit mehrstufigen Antwortskalen erhoben werden, ist erst die Verallgemeinerung dieser probabilistischen Modelle für ordinale Daten von besonderer Bedeutung für die sozialwissenschaftliche Forschung. Das dichotome Raschmodell wurde von *Rasch* selbst auch für mehrere nominale Antwortkategorien formuliert (*Rasch* 1961) und er hat Reduktionsbedingungen abgeleitet, die bei einer Anwendung auf ordinale Kategorien gelten müssen. Es ist jedoch erst den Arbeiten von *Andrich* (1978) und *Masters* (1982) zu verdanken, daß die Verallgemeinerung des Raschmodells für geordnete Kategorien konsequent aus einigen Grundannahmen abgeleitet wurde und damit auch eine eindeutige Interpretation der Modellparameter gegeben ist (s.a. *Rost* 1988a: 170ff.).

Die Logik dieser ordinalen Raschmodelle läßt sich am einfachsten graphisch, anhand der verallgemeinerten ICCs zeigen. Eine ICC stellt die Abhängigkeit der Antwortwahrscheinlichkeit eines (dichotomen) Items vom latenten Kontinuum dar. Im Fall von mehreren geordneten Antwortkategorien benötigt man für jede Kategorie eine solche Kurve (Category Characteristic Curve, CCC) und deren Verlauf für ein Item mit 6 Antwortkategorien sieht dann folgendermaßen aus (vgl. Abb. 6).

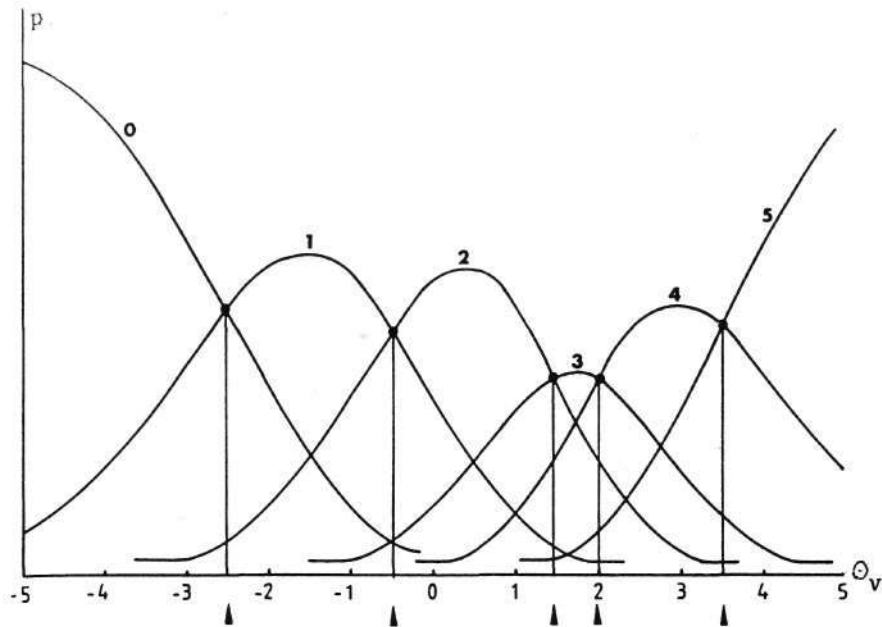


Abb. 6: Die CCCs eines Items mit 6 geordneten Antwortkategorien

Geht man das latente Kontinuum "hoch", d.h. von links nach rechts, so ist zunächst die Wahrscheinlichkeit für Kategorie 0 am höchsten, diese sinkt dann ab und die Wahrscheinlichkeit für Kategorie 1 steigt. Geht man weiter, so sinkt diese auch wieder und die Wahrscheinlichkeit für Kategorie 2 steigt, u.s.w. Diese Kurvenschar wird bei den ordinalen Raschmodellen (mit 6 Kategorien) durch 5 Parameter parametrisiert, nämlich genau durch die Abszissenwerte der Schnittpunkte je zweier benachbarter CCCs (s. Abb. 6). Liegen diese Schnittpunkte weiter auseinander, so weist die dazwischen liegende CCC einen höheren "Hügel", d.h. eine höhere Antwortwahrscheinlichkeit auf. Ein Schnittpunkt markiert genau jene Stelle, wo die Antwortwahrscheinlichkeiten der beiden beteiligten Kategorien gleich groß sind. Dieser Punkt wird die "Schwelle" zwischen den beiden Kategorien genannt.

Das Bemerkenswerte besteht nun darin, daß die Schwellen geordnet sind, d.h. die Schwelle zu Kategorie x liegt rechts von der Schwelle zu Kategorie x-1 (s. Abb. 6). Die Rang-



ordnung der Antwortkategorien schlägt sich in der Ordnung der Schwellen nieder. Anders ausgedrückt, nur wenn die Schwellen geordnet sind, handelt es sich bei der Antwortskala wirklich um eine Rangskala. Ob dies der Fall ist, läßt sich an den Schwellenparametern ablesen, denn dies sind die Parameter des ordinalen Raschmodells:

$$(8) \quad p(x_{ij}) = \frac{\exp(x \cdot \theta_j + \alpha_{ix})}{\sum_{x=0}^n \exp(x \cdot \theta_j + \alpha_{ix})}$$

wobei die α_{ix} die bis zur Schwelle x kumulierten Schwellenparameter sind. Das Schöne an dieser Verallgemeinerung ist, daß die Eigenschaft spezifisch objektiver Messung (s.o.) hierbei erhalten bleibt.

Das Besondere an dieser Art der Modellbildung für ordinale Daten liegt darin, daß man nicht einmal die Ordinalskalenqualität der Daten a priori annehmen muß, sondern man die Geltung dieser Annahme anhand der Modellparameter "ablesen" kann. Weiterhin wird über die Schwellendistanzen, also den Abstand zweier benachbarter Schwellen auf dem latenten Kontinuum, auch die "Größe" der jeweils dazwischen liegenden Antwortkategorie bestimmt: je größer eine Schwellendistanz, desto größer sind die Antwortwahrscheinlichkeiten der eingeschlossenen Kategorie. Dies ermöglicht es zum Beispiel, mit Hilfe von Restriktionen der Schwellendistanzen Annahmen über den Gebrauch der Antwortskala, wie die Äquidistanz der Schwellen oder konstante Itemdispersionen, zu formulieren und empirisch zu testen (vgl. Rost 1988b).

Diese Gründe waren ausschlaggebend, auch die LCA mittels desselben Schwellenansatzes für geordnete Kategorien zu spezifizieren (Rost 1988a). Das setzt zunächst eine logistische Reparametrisierung der LCA voraus, damit die klassenspezifischen Antwortwahrscheinlichkeiten auf Schwellenparameter zurückgeführt werden, die - wie die Raschparameter - auf einer logistischen Skala liegen. Die LCA für ordinale Antwortkategorien $x = 0, 1, \dots, m$ lautet dann in ihrer **unrestringierten** Form:

$$(9) \quad p(x_{ig}) = \sum_{f} \pi_f \cdot \frac{\exp(\alpha_{ixg})}{\sum_{x=0}^n \exp(\alpha_{ixg})}$$

wobei α_{ixg} die kumulierte Leichtigkeit der Schwellen bis zur Kategorie x bei Item i in Klasse g parametrisiert und als Normierungsbedingung gilt: $\alpha_{0ig} = 0$ für alle i und g . Un-

restringiert ist dieses Modell gegenüber der normalen LCA für (mehr als zwei) nominale Kategorien deshalb, weil es genauso viele Parameter enthält wie diese (nur statt Kategorienwahrscheinlichkeiten Schwellenleichtigkeiten) und sich beide Parameterarten direkt ineinander umrechnen lassen. Der Vorteil ist, daß man (wie beim ordinalen Raschmodell) an den Parametern ablesen kann, ob die beobachteten Variablen tatsächlich ordinalskaliert sind (sofern man die Ordnung der Schwellen als Kriterium für die Ordnung der Kategorien akzeptiert).

Ein weiterer Vorteil besteht darin, daß man die (dekumulierten) Schwellenparameter β_{ig} restringieren kann, um so Parameter zu sparen und spezifische Annahmen zu testen.² Die in dem Statistikprogramm LACORD (Rost 1990a) gewählte Art von Restriktionen definiert die mittlere Leichtigkeit aller Schwellen eines Items, μ_g als **Leichtigkeit des Items i in Klasse g** und hat dann nur noch die Schwellenabstände als davon unabhängige Parameter³. Diese Schwellenabstände können nun

- für die Kategorien unterschiedlich groß sein, aber für alle Items gleich groß sein (Modell 1),
- für alle Kategorien gleich groß (Äquidistanzannahme) sein, aber für die Items von unterschiedlicher Größe sein, was eine unterschiedliche Dispersion der Itemantworten bewirkt (Modell 2),
- für die Kategorien unterschiedlich groß sein (wie bei Modell 1), und (wie bei Modell 2) mit einem zusätzlichen Dispersionsparameter für die Items versehen sein (Modell 3),
- oder von unterschiedlicher Größe für jedes Item sowie jede Kategorie eines Items sein (Modell 4).

Bei dieser Auflistung ist davon ausgegangen worden, daß die so definierten Schwellendistanzen Eigenschaften des Items bzw. des Antwortformates, und somit für alle latenten Personenklassen identisch sind (**klassenunspezifische Schwellen**). Dies entspricht den Annahmen in ordinalen Raschmodellen, denn dort dürfen die Itemparameter auch nicht mit der Personenvariable interagieren. In der LCA ist es jedoch auch möglich, die Schwellendistanzen klassenspezifisch, also für jede Personenklasse unterschiedlich zu konzipieren. Das kann auch sehr sinnvoll sein, z.B. wenn in einzelnen Klassen bestimmte response sets

² Die dekumulierten Schwellenparameter β_{ig} sind in dem Modell fundamentale Parameter, da für die kumulierten Parameter α_{ig} gilt:

$$\alpha_{ig} = \sum_{r=0}^x \beta_{irg}$$

³ Vgl. auch die numerische Darstellung dieser Restriktion im Beitrag von *Tarnai & Rost* im selben Heft.

auftreten, die sich in unterschiedlichen Schwellendistanzen für die Personen niederschlagen. Nimmt man dieselben Annahmen, die den Modellen 1 bis 4 zugrunde liegen, läßt jedoch diese Restriktionen nur innerhalb jeder Klasse gelten (**klassenspezifische Schwellen**), so erhält man die Modelle 5 bis 8.

Die Berechnung der Parameteranzahlen für jedes Modell mag noch einmal die "Logik" dieser 8 Modelle verdeutlichen. Sei m die Anzahl der Schwellen ($m+1$ Antwortkategorien), k die Anzahl der Items und h die Anzahl der latenten Klassen, so hat **jedes** Modell zunächst einmal $h-1$ Klassengrößenparameter π_g und $h * k$ Itemleichtigkeiten μ_{ig} . Hinzu kommen für

- Modell 1: $m-1$ Schwellendistanzen,
- Modell 2: k itemspezifische Äquidistanzen,
- Modell 3: $(m-1)+k$ Grunddistanzen mit itemspezifischem Anteil,
- Modell 4: $(m-1)*k$ unrestringierte Distanzen,
- Modell 5: $h*(m-1)$ klassenspezifische Distanzen,
- Modell 6: $h*k$ klassen- und itemspezifische Äquidistanzen,
- Modell 7: $h*(m-1+k)$ klassenspezifische Distanzen im Sinne von Modell 3,
- Modell 8: $h*(m-1)*k$ klassenunspezifische, unrestringierte Distanzen.

Zählt man die Parameter von Modell 8 mit den eingangs genannten Itemleichtigkeiten zusammen, so sieht man, daß dieses Modell genauso viele Parameter enthält, wie die normale LCA, nämlich $h * m * k$ klassenspezifische Kategorienwahrscheinlichkeiten.

Mit diesem System von ordinalen LCA-Modellen ist die Möglichkeit gegeben, die verschiedensten Annahmen über das Zustandekommen von ordinalen Daten, über response sets, untypischen Gebrauch von Ratingskalen, itemspezifische Streuungen von Antworten, Personengruppen mit "zufälligen" Antworten und Ähnliches mehr zu prüfen. Die ersten vier Modelle sind zudem völlig parallel zu ihren "metrischen Vorbildern" aus der Rasch-Meßtheorie konstruiert: Modell 1 entspricht dem "rating scale model" von **Andrich** (1978), Modell 2 dem "dispersion model" von **Andrich** (1982), Modell 3 dem "successive interval model" von **Rost** (1988b) und Modell 4 dem "partial credit model" von **Masters** (1982). Alle Modelle können mit dem Programm LACORD berechnet werden, das vom Erstautor (**Rost** 1990a) zu beziehen ist.



3. Die Analyse der Skala "Jugendzentrismus" mit LACORD

- Die hier verwendete Skala "Jugendzentrismus" wurde in der Shell-Jugendstudie 1985 in dieser Form entwickelt. Sie beruht auf den jeweils zwei "besten" Items von 5 Unterskalen aus der 81er Jugendstudie (vgl. Jugendwerk der deutschen Shell 1981,1985; ZA-Archiv-Nrn.: 1201 und 1438/39). Die zehn Items dieser Skala sind bipolar aufgebaut, wobei die ersten beiden Kategorien Ablehnung und die restlichen beiden Zustimmung erfassen. Im einzelnen haben die Items folgende Formulierung:

Tabelle 1: Wortlaut der 10 Items zum Jugendzentrismus

- Item 1. Eigentlich behandeln alle Polizisten Jugendliche unfair.*
Item 2. In dieser Gesellschaft erlebst Du überall eine Feindseligkeit gegen uns Junge, die Dich total fertig macht.
Item 3. Unsere Gesellschaft tut eigentlich eine ganze Menge für die Jugendlichen
Item 4. Jugendliche sollten sich bei der Arbeit nicht alles gefallen lassen und ordentlich auf den Putz hauen, wenn's nötig ist.
Item 5. Eigentlich verdanke ich meinen Eltern sehr viel.
Item 6. Ich bemühe mich darum, meine Eltern zu verstehen, auch wenn's manchmal schwer fällt.
Item 7. Die wenigsten Erwachsenen verstehen die Probleme von Jugendlichen wirklich.
Item 8. Ich halte nicht viel von der Erfahrung der Erwachsenen. Ich verlasse mich lieber auf mich selbst.
Item 9. Von gleichaltrigen Freunden/Freundinnen lerne und erfahre ich mehr als von meinen Eltern.
Item 10. Eltern mischen sich dauernd in Sachen ein, die sie nichts angehen.

Zu jedem dieser Items waren vier Antwortmöglichkeiten mit den Extremen "stimmt nicht" (Antwortwert: 1) und "stimmt genau" (Antwortwert: 4) vorgegeben.

Im Band 5 der Shell Jugendstudie 1985 (97ff) sind diese Items einer eindimensionalen Skala zugeordnet (Cronbach's Alpha: .71, Spearman Brown: .77). Faktorenanalysen erbrachten jedoch weniger eindeutige Lösungen, bzw. eine drei-Faktor Lösung.

Zunächst gilt es zu klären, welche Klassenanzahl und welches der acht LACORD-Modelle für die Repräsentation der Items am angemessensten ist. Die Modelle 5 (klassenspezifische Schwellen, gleiche Kategorienabstände über alle Items) und 7 (klassenspezifische Schwellen, Grunddistanzen mit itemspezifischem Anteil) sind die Konkurrenten um die beste Modellanpassung, wobei die vier-Klassen-Lösung für alle Modelle zu bevorzugen ist. Nach Maßgabe des BIC fällt die Entscheidung für vier latente Klassen und Modell 5.

Tabelle 2: log-likelihood, Parameteranzahl (PA) und BIC für zwei bis vier latente Klassen

Anzahl der Klassen	Modell	log-likelihood	PA	BIC
2	5	-17431.592	25	35043.184
2	7	-17368.705	43	35047.010
3	5	-17288.393	38	34850.386
3	7	-17206.068	65	34880.140
4	1	-17294.916	45	34913.832
4	2	-17244.439	53	34870.478
4	3	-17242.527	54	34873.854
4	4	-17178.748	63	34811.093
4	5	-17196.205	51	34759.610
4	6	-17104.189	83	34805.978
4	7	-17097.238	87	34820.880
4	8	-17085.600	123	35056.800

Die in LACORD am einfachsten zu interpretierenden Parameter sind die Erwartungswerte der Antwortvariable. Sie geben den mittleren Antwortwert eines idealen Individuums in einer latenten Klasse wieder. Die Erwartungswerte in den vier latenten Klassen sind in Tabelle 3 wiedergegeben (da die Antwortvariable stets von 0 bis m variiert, liegen die Erwartungswerte im vorliegenden Fall zwischen 0 und 3).

Betrachtet man die Profile der Erwartungswerte der vier-Klassen-Lösung (vgl. Abb. 7), so fällt zunächst auf, daß drei geordnete Klassen (Klasse 1, 2 und 4) in dem Sinne existieren, daß sich die Erwartungswerte dieser Klassen nicht überschneiden. In Klasse 1 (Erwachsenen-zentrierte) ist durchgängig die stärkste Ablehnung jugendzentrierter Positionen festzustellen, in Klasse 2 läßt sich ebenso regelmäßig eine Zustimmung zu dieser Einstellung verzeichnen, während Klasse 4 (Durchschnittsgruppe) zwischen den beiden ersten Klassen anzusiedeln ist. Da es sich hier um geordnete Klassen handelt, könnten diese Klassen möglicherweise eindimensional Rasch-skalierbar sein. Die dritte Klasse jedoch zeigt kein so regelmäßiges Antwortmuster. Während sie bei den Items 1 und 2 (Polizisten behandeln Jugendliche unfair; die Gesellschaft ist feindselig gegenüber Jugendlichen) fast identisch mit der Durchschnittsgruppe ist, überschneidet sie diese bei Item 3 (die Gesellschaft tut viel für Jugendliche) in Richtung auf die jugendzentrierte, zweite Klasse, um sich bei den Items 4, 8 und 9 (man soll sich bei der Arbeit nicht alles gefallen lassen; ich verlasse mich auf mich selbst; von Gleichaltrigen lerne ich mehr) an die erste, erwachsenen-zentrierte Klasse anzunähern.

Tabelle 3: Erwartungswerte und Klassengröße für vier latente Klassen des Modells 5 (Items 3, 5 und 6 wurden umgepolt)

Item	Klasse 1 (29.4%)	Klasse 2 (17.7%)	Klasse 3 (13.3%)	Klasse 4 (39.6%)
1. Polizisten unfair	.558	1.907	1.221	1.112
2. Gesellschaft feindselig	.557	1.692	1.144	1.135
3. Gesellschaft tut viel	1.084	2.050	1.906	1.529
4. bei Arbeit nicht alles gefallen lassen	1.183	2.391	1.665	1.884
5. verdanke Eltern viel	.140	1.207	.701	.486
6. bemühe mich, Eltern zu verstehen	.348	.968	.769	.529
7. Erwachsene verstehen Probleme nicht	1.638	2.537	2.157	2.034
8. verlasse mich auf mich selbst	.983	2.462	1.505	1.735
9. lerne von Gleich- altrigen	1.091	2.463	1.329	1.839
10 Eltern mischen sich ein	1.159	2.447	1.912	1.734

Erwartungswerte der 4 latenten Klassen

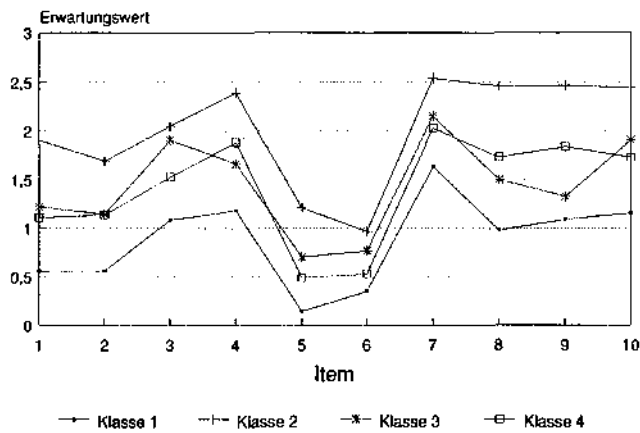


Abb. 7: Die latenten Profile der 4-Klassen-Lösung



Besonders bei den Items, bei denen es um eine Bejahung der eigenen Erfahrungen von Jugendlichen geht, nähert sich diese Klasse also der erwachsenenzentrierten Gruppe an, während sie mit den Jugendzentrierten die Auffassung teilt, die Gesellschaft tue nicht viel für die Jugendlichen.

Tabelle 4: Schwellenparameter (SP) für vier latente Klassen

	Klasse 1	Klasse 2	Klasse 3	Klasse 4
SP 1	1.06	1.10	-.18	1.58
SP 2	.18	-.04	.00	.01
SP 3	-1.24	-1.06	.18	-1.59

Ein dazu passendes Bild ergibt sich für die klassenspezifischen Schwellenparameter, bei denen negative Werte "schwierige" Schwellen anzeigen. Da in Modell 5 gleiche Schwellendistanzen für alle Items in einer Klasse unterstellt werden, ergeben sich nur drei Parameter pro Klasse. Die Schwellenparameter für die Klassen 1,2 und 4 sind hinsichtlich der Schwierigkeiten (die hier identisch mit Zustimmung zu den Items sind) gleichsinnig geordnet, und zwar in Richtung zunehmender Schwierigkeit von Schwelle zu Schwelle. In Klasse 3 hingegen sind die Schwellenparameter umgekehrt geordnet: die Schwierigkeit der Schwellen sinkt geringfügig mit der Zustimmung zu den Items. Derartige Unregelmäßigkeiten des Antwortverhaltens weisen ebenfalls darauf hin, daß die Personen in dieser Klasse die Items nicht im Sinne des latenten Kontinuums bearbeiten: Zumindest ordnen sich die Schwellen der Antwortskala nicht entlang einer solchen Dimension an (s.o. Kap. 2). Als Gewinn für die Analyse bleibt festzuhalten, daß LACORD in der Lage ist, Befragte mit einem abweichenden Antwortmuster zu identifizieren und in einer gesonderten Klasse zusammenzufassen.

4. Das Mischverteilungs-Raschmodell

Die LACORD-Ergebnisse zur Jugendzentrismusskala haben ein Resultat erbracht, das man in ähnlicher Form sehr oft bei der Auswertung von Einstellungsitems erhält: Es gibt einige geordnete latente Klassen, die auf eine eindimensionale Einstellungsvariable hinweisen, und es gibt eine (oder mehrere) Klassen, die aus dieser Ordnung "herausfallen". Im vorliegenden Fall sind es drei geordnete Klassen mit geordneten Schwellen, und eine, relativ kleine Klasse, die sich zudem durch gegensinnig geordnete Schwellenparameter auszeich-



net. Nach dem in Abschnitt 1 dargestellten Zusammenhang zwischen geordneten Klassen und (Rasch-) Skalierbarkeit legt ein solches Ergebnis nahe, daß der Test sehr wohl metrisch skalierbar ist, aber **nur für einen Teil** der untersuchten Stichprobe.

Es wäre also wünschenswert, ein Skalierungsmodell zu haben, das jenen Teil der Population, in der die Skala gilt, von jenem (möglichst kleinen) Anteil der Personen trennt, die im Sinne dieses Modells **nicht skalierbar** sind, oder nur mittels anderer Modellparameter skalierbar sind. Das ist genau die Idee des **Mischverteilungs-Raschmodells** ("MIRA", Rost 1990b). Dieses Modell verbindet die Rasch-Meßtheorie und die LCA in der Weise, daß jeweils für die Personen innerhalb einer latenten Klasse das Raschmodell gilt, aber mit unterschiedlichen Parametern von Klasse zu Klasse. D.h. die LCA wird dahin erweitert, daß innerhalb jeder Klasse nicht mehr Konstanz der Antwortwahrscheinlichkeiten über alle Personen herrscht, sondern zusätzlich eine quantitative Variation zwischen den Personen eingeführt wird. Ausgehend vom Raschmodell kann man auch sagen, das Raschmodell wird dahin erweitert, daß das Modell nicht mehr für alle Personen gelten muß, sondern es werden solche (latenten) Personengruppen gesucht, in denen das Modell gilt, mit jeweils anderen Parametern in jeder Gruppe.

Damit ist das MIRA-Modell das gemeinsame Obermodell von Raschmodell und LCA. Es lautet in seiner allgemeinen Form

$$(10) \quad p(x_w) = \sum_g \pi_g \cdot \frac{\exp(x\theta_{vg} + \alpha_{wg})}{\sum_{x=0}^m \exp(x\theta_{vg} + \alpha_{wg})}$$

und man sieht, daß es sich für nur eine latente Klasse g auf das ordinale Raschmodell (genauer: das partial credit model) reduziert (vgl. Gleichung 8). Sind dagegen alle Fähigkeitsparameter θ_{vg} in jeder Klasse konstant, so reduziert sich MIRA auf Modell 8 von LACORD, bzw. die normale LCA (vgl. Gleichung 9).

Die zusätzlichen quantitativen Eigenschaftsparameter θ_{vg} im latent-class Modell können natürlich einige Schwierigkeiten bei der Schätzung der Modellparameter bereiten. Hier erweist sich jedoch eine Eigenschaft des Raschmodells als sehr nützlich, die es erlaubt, die Item-, bzw. Schwellenparameter zu schätzen, ohne die Personenparameter gleichzeitig mitschätzen zu müssen (bedingte Inferenz). Man benötigt statt dessen "nur" die suffizienten Statistiken für die Personenparameter und das sind die Häufigkeiten der Summenscores. Während diese Scorehäufigkeiten beim "normalen" Raschmodell direkt beobachtbar sind (die Randsummen der Datenmatrix), sind es bei MIRA zunächst noch unbekannte



Größen, die sich erst im Laufe der Identifizierung der latenten Klassen ergeben. D.h., jede klassenspezifische Scorehäufigkeit ist ein unbekannter, aber zu schätzender Parameter und zählt auch entsprechend bei der Berechnung der Parameteranzahl (z.B. für die Berechnung des BIC, s.o.) mit.

Letzteres ist ein gewichtiger Grund, sich nach weiteren Möglichkeiten der Parameterreduktion umzusehen, denn es werden z.B. bei 10 vierkategorialen Items für jede Klasse allein 31-2=29 Parameter benötigt, um die (latente) Scoreverteilung zu parametrisieren. Das sind genauso viele Parameter, wie es in jeder Klasse Schwellenparameter gibt. Erweisen sich die Scores in jeder Klasse als annähernd normalverteilt, so würden 2 Parameter (μ und σ) ausreichen, die suffizienten Statistiken für die klassenspezifischen Personenparameter zu parametrisieren. Genau dies ist in der Normalverteilungs-(NV-)Variante von MIRA realisiert. Das Programm MIRA kann ebenfalls vom Erstautor bezogen werden.

5. Ergebnisse zum "Jugendzentrismus" mit MIRA

Wie bereits bei LACORD, kann der BIC auch beim Mischverteilungs-Raschmodell zur Prüfung der Angemessenheit unterschiedlicher Anzahlen von Klassen verwendet werden. Natürlich stellt der BIC nur einen empirischen Hinweis auf die Veträglichkeit der jeweiligen Lösung mit der Datenstruktur dar, der gegebenenfalls durch theoretische Erwägungen ergänzt werden sollte. Im vorliegenden Fall ergaben sich die in Tabelle 5 aufgeführten Anpassungsmaße.

Tabelle 5: log-likelihood, Parameteranzahl (PA) und BIC für 1-4 Mira-Klassen (Items 3, 5 und 6 invertiert, Normalverteilungsvariante)

Anzahl der Klassen	log-likelihood	PA	BIC
1	-17265.564	32	34760.400
2	-17104.301	63	34602.202
3	-17020.861	94	34716.522
4	-16939.619	125	34779.238

Die Skala "Jugendzentrismus" ist also für die Gesamtpopulation nicht eindimensional Rasch-skalierbar, vielmehr spricht der BIC für eine zwei-Klassen-Lösung. Aus Tabelle 6 können die entsprechenden Itemparameter beider Klassen entnommen werden.



Tabelle 6: Itemparameter und Klassengröße für zwei MIRA-Klassen

Item	Klasse 1 (20.2%)	Klasse 2 (79.7%)
1. Polizisten unfair	-.60	-1.34
2. Gesellschaft feindselig	-.82	-2.09
3. Gesellschaft tut viel	1.01	1.25
4. bei Arbeit nicht alles gefallen lassen	.56	2.17
5. verdanke Eltern viel	-1.75	-4.66
6. bemühe mich, Eltern zu verstehen	-1.76	-5.20
7. Erwachsene verstehen Probleme nicht	1.82	4.17
8. verlasse mich auf mich selbst	.24	1.65
9. lerne von Gleichaltrigen	.09	2.14
10. Eltern mischen sich ein	1.20	1.91

Beide Rasch-homogenen Klassen reproduzieren eine ähnliche Antwortstruktur: die "schwierigsten" Items, das heißt diejenigen mit der geringsten Zustimmung zu jugendzentrierten Positionen, beziehen sich auf die eigenen Eltern (Items 5 und 6), gefolgt von den Items, die eine gesellschaftliche Dimension erfassen (Items 1 und 2). Interessanterweise waren beide Itempaare in der Shell-Jugendstudie 1981 (vgl. Jugendwerk der Deutschen Shell 1981: 38ff) Elemente von Subskalen zum Jugendzentrismus (die Items 1 und 2 gehörten zur Skala "Gefühl der Diskriminierung der Jugendlichen durch gesellschaftliche Autoritäten" und die Items 5 und 6 zur Skala "persönliches Vertrauen und Dankbarkeit gegenüber den Eltern"). Zu den Items, die jugendzentrierte Positionen am stärksten ausdrücken, gehören in beiden Klassen die Items 7 (Erwachsene verstehen Probleme nicht) und 10 (Eltern mischen sich ein). Wichtige Unterschiede zwischen beiden Klassen bestehen - ähnlich wie bei LACORD zwischen der dritten, ungeordneten Klasse und der vierten "Durchschnittsklasse" - darin, daß in Klasse 1 das Aufbegehren bei der Arbeit (Item 4) und jugendspezifische Erfahrungen (Items 8 und 9) für weniger wichtig gehalten werden. Generell neigen die Befragten in Klasse 1 zu moderaterem Antwortverhalten, denn die Leichtigkeitparameter fallen betragsmäßig erheblich niedriger aus als in Klasse 2.



Ein Blick auf die Schwellenparameter, die hier aus Platzgründen nicht wiedergegeben werden, zeigt, daß in Klasse 1 ungeordnete Schwellenschwierigkeiten vorherrschen, d.h. die Anforderungen, die an eine Rating-Skala gestellt werden müssen, sind hier nicht erfüllt. Im Gegensatz hierzu sind alle Schwellen in Klasse 2, die mit 79.7% die große Mehrheit der Stichprobe umfaßt, gleichsinnig geordnet.

Natürlich ist es nun von Interesse, ob die ungeordnete, dritte Klasse der LACORD-Lösung zum Teil identisch mit der, hinsichtlich der Schwellenparameter ungeordneten, ersten Klasse des Mischverteilungs-Raschmodells ist. Wenn dies der Fall wäre, hätte man ein starkes Indiz dafür, daß zwar die große Mehrheit der Stichprobe eindimensional Raschskalierbar ist, jedoch eine kleine Gruppe ein unsystematisches Antwortverhalten zeigt. Wie aus Tab. 7 ersehen werden kann, werden tatsächlich 151 der 158 Mitglieder der dritten LACORD-Klasse in die ungeordnete MIRA-Klasse eingeordnet. Zusätzlich ist die Zuordnungswahrscheinlichkeit der sieben verbleibenden Klassenmitglieder in die zweite MIRA-Klasse mit .582 sehr niedrig, so daß eine sehr starke Überschneidung zwischen der Mitgliedschaft in der dritten LACORD-Klasse und der ersten Klasse des Mischverteilungs-Raschmodells besteht. Die Skala "Jugendzentrismus" ist somit für 80% (1194/1471) der Stichprobe eindimensional Raschskalierbar. Die verbleibenden 20% , für die die Antwortkategorien der Items offensichtlich eine abweichende Semantik haben, sind größtenteils identisch mit den Mitgliedern einer ungeordneten latenten Klasse, deren Schwellenparameter ebenfalls eine von der Mehrheit abweichende Struktur aufweisen.

Tabelle 7: Kreuztabelle der vier LACORD-Klassen und der beiden MIRA-Klassen (Häufigkeiten und *mittlere Zuordnungswahrscheinlichkeit*)

LACORD-Klassen	MIRA-Klassen		
	1	2	
1	70 .757	355 .859	425 .824
2	40 .733	218 .859	258 .818
3	151 .818	7 .582	158 .770
4	16 .554	614 .847	630 .774
	277 .808	1194 .911	1471

6. Schlußfolgerungen

Üblicherweise wird bei der Konstruktion von Skalen ein einheitliches Meßmodell - und dies gilt sowohl für die klassische Testtheorie als auch die Rasch-Skalierung - für die ganze Stichprobe zugrundegelegt. Bei diesem Vorgehen läuft man jedoch, wie im vorliegenden Fall gezeigt werden konnte, Gefahr, Ungleiches mit einer einheitlichen Meßlatte zu vergleichen. Zeigt sich im Verlauf der Analyse, daß ein eindimensionales Meßmodell "paßt", ist dies angenehm für den Forscher, denn es erleichtert die Interpretation. Die ungeprüfte Vereinheitlichung einer Skala für alle Befragten kann jedoch zu trügerischen Ergebnissen führen.

Das Verfahren der Skalenkonstruktion ist für die meisten Anwender inzwischen so stark routinisiert (explorative Faktorenanalysen, danach Itemanalysen), daß häufig eine Reflexion über die substanzwissenschaftlichen Implikationen dieses Vorgehens ausbleibt. Viele soziologische Theorieansätze, etwa zum Wertewandel oder zur Theorie sozialer Ungleichheit, gehen ja gerade von einer Auflösung einheitlicher Bezugssysteme aus (vgl. etwa *Inglehart* 1990; *Beck* 1983,1986; *Hradil* 1987,1990; *Berger und Hradil* 1990). Weshalb sollten dann Einstellungsitems für alle Befragten eine identische Semantik aufweisen? Viel wahrscheinlicher ist doch aus der Sicht derartiger bezugsgruppenspezifischer Theorienbildung eine unterschiedliche Dimensionalität in verschiedenen Subgruppen. In gewisser Weise kann man in der Konzeption von Mischverteilungsmodellen eine Transformierung substanzwissenschaftlicher Annahmen in statistische Modellbildung sehen.

Andererseits muß man sich verdeutlichen, welchen Preis man für die Aufgabe der Eindimensionalität zahlt. Allein diese Annahme ermöglicht es ja, unterschiedliche soziale Gruppen entlang eines einheitlichen Einstellungskontinuums zu verorten und im Sinne eines "mehr oder weniger" (... jugendzentriert, postmaterialistisch, autoritär etc.) miteinander in Beziehung zu setzen.

Literatur

Andrich, D. (1978): A rating formulation for ordered response categories. *Psychometrika* 43. 561-573

Andrich, D. (1982): An extension of the Rasch model for ratings providing both location and dispersion Parameters. *Psychometrika*, 47, 1. 105-113

Beck, U. (1983): Jenseits von Stand und Klasse: Soziale Ungleichheit, gesellschaftliche Individualisierungstendenzen und Entstehung neuer Formen und Identitäten. In: *Kreckel, R.* (Hrsg.): *Soziale Ungleichheiten*. Sonderband 2 der Sozialen Welt. Göttingen. 35-74

- Beck, U.** (1986): Risikogesellschaft. Auf dem Weg in eine andere Moderne. Frankfurt
- Berger, P. A. /Hradil, S.** (Hrsg.) (1990): Lebenslagen, Lebensläufe, Lebensstile. Sonderband 7 der Sozialen Welt Göttingen
- Bozdogan, H.** (1987): Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika 52,3. 345-370
- Hradil, S.** (1987): Sozialstrukturanalyse in einer fortgeschrittenen Gesellschaft. Opladen
- Hradil, S.** (1990): Postmoderne Sozialstruktur? Zur empirischen Relevanz einer "modernen" Theorie sozialen Wandels. In: Berger, P.A./Hradil, S. (Hrsg.) (1990)
- Inglehart, R.** (1990): Culture shift in advanced industrial society. Princeton
- Jugendwerk der Deutschen Shell** (1981): Jugend '81. Lebensentwürfe, Alltagskulturen, Zukunftsbilder. Bd 1. Hamburg
- Jugendwerk der Deutschen Shell** (1985): Jugendliche und Erwachsene '85. Generationen im Vergleich. Opladen
- Krebs, D. /Schuessler, K.F.** (1987): Soziale Empfindungen. Ein interkultureller Skalenvergleich bei Deutschen und Amerikanern. Frankfurt/Main und New York
- Lazarsfeld, P. F.** (1950): Logical and mathematical foundations of latent structure analysis. In: **Stouffer, S. A. / Guttman, L. / Suchman, E. A. / Lazarsfeld, P. F. / Star, S. A./ Clausen, J. A. (eds.):** Studies in social psychology in world war II. Vol IV. Princeton
- Lazarsfeld, P. F. /Henry, N. W.** (1968): Latent structure analysis. Boston
- Lord, F. M. /Novick, M. R.** (1968): Statistical theories of mental test scores. Reading / Mass.
- Masters, G. N.** (1982): A Rasch model for partial credit scoring. Psychometrika 47. 149-174
- Rasch, G.** (1960): Probabilistic models for some intelligence and attainment tests. Copenhagen
- Rasch, G.** (1961): On general laws and the meaning of measurement in psychology. University of California Press
- Rost, J.** (1988a): Quantitative und qualitative probabilistische Testtheorie. Bern
- Rost, J.** (1988 b): Measuring attitudes with a threshold model drawing on a traditional scaling concept. In: Applied Psychological Measurement, Vol 12, No. 4. 397-409
- Rost, J.** (1990 a): LACORD - latent class analysis for ordinal variables. Institut für die Pädagogik der Naturwissenschaften. Kiel
- Rost, J.** (1990 b): Rasch models in latent classes: An Integration of two approaches to item analysis. In: Applied Psychological Measurement, Vol. 14, No. 3
- Rost, J.** (1991): MIRA. Mischverteilungs-Raschmodell. Ein PC-Programm. Institut für die Pädagogik der Naturwissenschaften. Kiel
- Schulze, G.** (1990): Alltagsästhetik, Milieustruktur und Erlebnismarkt. Eine kultursoziologische Untersuchung der Bundesrepublik Deutschland. Projektbericht für die Deutsche Forschungsgemeinschaft. Anhang II: Meßtheoretische Überlegungen und Skalen. Bamberg
- Zuma - Handbuch Sozialwissenschaftlicher Skalen** (1988): Bonn: **IZ**