

ANGEWANDTE PSYCHOLOGISCHE TESTTHEORIE UND DIAGNOSTISCHE GRUNDLAGENFORSCHUNG

Organisation: Kempf, Wilhelm (Konstanz)

In der Testtheorie wurden lange Zeit sehr restriktive Modelle favorisiert, die oft zu strenge Anforderungen an die Daten stellten und die psychologischen Prozesse der Testbearbeitung nur ungenügend abzubilden vermochten. Neuere Entwicklungen wie die Latent-Class-Analyse für ordinale Daten (Rost, 1988), das Mischverteilungs-Raschmodell (Rost, 1990) und die wiederaufgenommene Diskussion über dynamische Itemresponsemodelle (Andrich, 1985; Kempf, 1991) haben demgegenüber zu realistischeren Modellvarianten geführt.

Zielsetzung der Arbeitsgruppe ist es, die Angemessenheit solcher neuerer, testtheoretischer Modelle für die methodische Fundierung psychologischer Tests zu diskutieren und aus den Erfahrungen mit der praktischen Anwendung der Modelle konkrete Anforderungen an die Weiterentwicklung testtheoretischer Ansätze zu erarbeiten.

Andrich, D. (1985). A latent trait model for items with response dependencies: implications for test construction and analysis. In Embretson, S. (Ed.), *Test design: contributions from psychology, education and psychometrics*. New York: Academic Press.

Kempf, W. (1991). On dynamic Rasch-models. *Diskussionsbeiträge der Projektgruppe Friedensforschung Konstanz*, 11/91.

Rost, J. (1988). *Quantitative und qualitative probabilistische Testtheorie*. Bern: Huber.

Rost, J. (1990). Rasch-models in latent classes: an integration of two approaches to item analysis. *Applied Psychological Measurement*, 3, 271-282.

LATENT TRAIT UND LATENT CLASS ANALYSEN DES BECK-DEPRESSIONSINVENTARS

Keller, Ferdinand & Kempf, Wilhelm (Ravensburg)

Das Beck-Depressionsinventar (BDI) stellt eine der meistverwendeten Selbstbeurteilungsskalen zur Erfassung depressiver Symptomatik dar. Abgesehen von klassischen Gütekriterien (Hautzinger, 1991) ist über die psychometrischen Eigenschaften der Skala bisher jedoch nur wenig bekannt. Zwar haben Bouman und Kok (1987) eine dichotome Rasch-Analyse des BDI durchgeführt, in welcher drei homogene Subskalen identifiziert wurden: die Skalen "Guilt & Failure" (G&F; Items Nr. 3,5,7,8), "Mood & Inhibition" (M&I; Items Nr. 1,2,9,13,15) und eine "Somatic Scale" (SOM; Items Nr. 10,11,16-21). Da das BDI im Original über vier Antwortkategorien pro Item verfügt, ist dieses Ergebnis aber nur von beschränkter Aussagekraft. Gegenstand der vorliegenden Untersuchung sollte es daher sein, die dichotome Itemanalyse von Bouman und Kok zu replizieren und mittels politomer Itemanalyse die Ordinalität der Antwortkategorien zu untersuchen.

Als Datengrundlage dienten N=714 vollständig ausgefüllte Fragebögen von stationär behandlungsbedürftigen Patienten der Depressionsstationen der PLKs Weissenau (Ravensburg) und Reichenau (Konstanz).

Die Datenanalyse mittels des Mixed-Rasch-Modells zeigte, daß bei der 1-Klassen Lösung lediglich die fünf Items der Subskala M&I über geordnete Schwellenparameter verfügen, während die Items der Subskalen G&F und SOM keine geordneten Antwortkategorien besitzen (vgl. Tab.1). Auch die Subskala M&I erwies sich jedoch nicht als homogen. Der AIC-Index favorisierte für alle drei Skalen sowohl bei politomer als auch bei dichotomer Analyse eine Mehrklassenlösung, die sich (mit Ausnahme der dichotomen Analyse der Skala G&F) auch nach dem CLR-Test als signifikant erwies. Für die Skalen M&I und SOM mußte zudem die Annahme der lokalen stochastischen Unabhängigkeit der Antworten (LSU) zurückgewiesen werden. Die Daten können durch ein dynamisches Testmodell mit den Transferparametern Φ_0 und $\Phi_1 = \dots = \Phi_{k-1}$ signifikant besser beschrieben werden als durch das binäre RM. Obwohl sich die Homogenitätsvoraussetzungen des RM damit durchgehend als zu streng erwiesen haben, sprechen die mittels dichotomer Latent-Class-Analyse (LCA) identifizierten Klassen dennoch eher für eine quantitative Differenzierung der Symptomstärke denn für eine qualitative Differenzierung der erfaßten Symptomatik. Die Ergebnisse der polytomen LCA lassen jedoch Zweifel an der Skalierbarkeit der polytomen Antwortkategorien entstehen.

Hautzinger, M. (1991). Das Beck-Depressionsinventar (BDI) in der Klinik. *Der Nervenarzt*, 62, 689-696.

LOKALE SERIELLE ABHÄNGIGKEIT IN ITEM-BUNDLES

Kempf, Wilhelm (Konstanz)

Zwei Modellannahmen sind grundlegend für das binäre Rasch-Modell (RM): lokale stochastische Unabhängigkeit (LSU) der Antwortvariablen x_{vj} und Suffizienz des Summenscores. Während letztere Annahme ein technisches Erfordernis für die Parameterschät-

zung darstellt, bedeutet die LSU eine wesentliche Einschränkung des Anwendungsbereiches des Modells. Versuche zur Überwindung dieser Einschränkung führten zur Entwicklung des Dynamischen Testmodells (DTM), dessen Anwendung jedoch lange Zeit an numerischen Schwierigkeiten scheiterte, die erst nach einer Neuformulierung der Modellstruktur (Kempf, in press) gelöst werden konnten.

$$(1) \text{ Prob}(x_{vj}=1|r) = \{\exp(\theta_v - \delta_i) + \exp(\Phi_r - \delta_i)\} / \{1 + \exp(\theta_v - \delta_i)\}.$$

Grundlegend für das DTM ist das Konzept der lokalen seriellen Abhängigkeit (LSA), wonach die Lösungswahrscheinlichkeit eines Items i von der Anzahl r der zuvor gelösten Aufgaben abhängt. Ein anderer Weg zur Abschwächung der LSU geht auf Andrich (1985) zurück und besteht in der Anwendung des polytomen Rasch-Modells (PRM)

$$(2) \text{ Prob}(y_{vt}=y) = \exp(\theta_v - \delta_t y) / \left\{ \sum_{s=0}^k \exp(\theta_v - \delta_t s) \right\}$$

auf die Scores in sog. "Item-Bundles": Gruppen von k Items, zwischen welchen LSU besteht, während die Antworten innerhalb eines Bundles voneinander abhängig sein können. Andrich hat drei Modelle untersucht, die alle zu Spezialfällen des PRM für ordinale Daten mit $\theta_{vy} = y\theta_v$ führen: 1. Das Binomialmodell (konstante Itemschwierigkeiten und LSU innerhalb der Itembundles). 2. Das RM (variable Itemschwierigkeiten und LSU innerhalb der Itembundles). 3. Ein Modell mit konstanten Itemschwierigkeiten und lokaler Abhängigkeit innerhalb der Item-Bundles.

Anders als im DTM wird diese Abhängigkeit hier jedoch nicht über bedingte Lösungswahrscheinlichkeiten beschrieben, sondern über die Schwellenwahrscheinlichkeiten

$$(3) \text{ Prob}(y_{vt}=y | y-1 \leq y_{vt} \leq y) = \exp(\theta_v - \delta_t + c_y + \Phi_y) / \{1 + \exp(\theta_v - \delta_t + c_y + \Phi_y)\},$$

worin $c_y = \ln\{(k-y+1)/y\}$ (Kempf, 1991). Die Formulierung eines allgemeinen Modells mit lokaler Abhängigkeit bei variablen Itemschwierigkeiten ist innerhalb dieses Ansatzes nicht möglich, da es zu einer Überparametrisierung des PRM für ordinale Daten führt und die Modellparameter dann nicht identifizierbar sind. Soll beides gleichzeitig beschrieben werden - variable Itemschwierigkeiten und lokale Abhängigkeit - so führt die Anwendung des DTM auf die bedingten Lösungswahrscheinlichkeiten jedoch ebenfalls zur Geltung des PRM für die Scorewahrscheinlichkeiten der Bundle-Scores. Anders als die von Andrich diskutierten Modellvarianten führt das DTM jedoch nicht zu einer Variante des PRM für ordinale Daten, sondern zu einem PRM mit den Personenparametern

$$(4) \theta_{vy} = \sum_{r=0}^{y-1} \ln\{\exp(\theta_v) + \exp(\Phi_r)\}.$$

Kempf, W. (1991). On dynamic Rasch-models. Diskussionsbeiträge der Projektgruppe Friedensforschung Konstanz, 11/91.

INDIVIDUELLE UNTERSCHIEDE BEIM LÖSEN VON RAUMVORSTELLUNGS- AUFGABEN

Neufang, Michaela & Rost, Jürgen (Kiel)

Seit über 50 Jahren stellt das Konstrukt "Raumvorstellung" - allgemein anerkannt als ein Element menschlicher Fähigkeiten - einen Gegenstand vielfältiger, vor allem faktorenanalytischer Forschungsbemühungen dar. Die resultierenden Ergebnisse stellen allerdings kein einheitliches Bild dar (s. z.B. Michael et al., 1957). Läßt man die Pbn retrospektiv über die verwendeten Strategien berichten, so zeigen sich vor allem zwei Lösungsstrategien: erstens, mentale Rotationen der dargestellten Objekte (holistische Vorgehensweise) und zweitens nicht-transformale Prozesse, z.B. Merkmalsvergleichsprozesse (analytisches Vorgehen). Bestehen die individuellen Unterschiede primär in der Art der präferierten Lösungsstrategie, und lassen sich "im Prinzip" alle Raumvorstellungsaufgaben (wenn auch mit unterschiedlichen Itemschwierigkeiten) nach beiden Strategien lösen, so stellt die Faktorenanalyse kein adäquates Modell zur Abbildung individueller Differenzen dar. Vielmehr sind in einer Zufallsstichprobe von Personen Teilgruppen zu erwarten, die sich durch ihre präferierte Strategie auszeichnen, und die sich daher auch über unterschiedliche Itemschwierigkeiten identifizieren lassen. Eine solche Zuordnung von Personen zu Teilgruppen mit jeweils unterschiedlichen Profilen von Itemschwierigkeiten kann mit Hilfe des Mischverteilungs-Rasch-Modells erfolgen (Rost, 1990).

Es werden die Ergebnisse einer empirischen Untersuchung berichtet, in der an 264 erwachsenen Vpn sechs verschiedene Arten von Raumvorstellungsaufgaben analysiert wurden. Die Hypothese bestand in der Annahme einer Mischverteilung aus zwei Strategiepräferenzgruppen (analytisch und holistisch) und ggf. einer dritten Gruppe von "Wechslern". Die Hypothese wurde in Form von erwarteten Profilen der Itemschwierigkeiten innerhalb der Gruppen formuliert und ließ sich mit Hilfe von informationstheoretischen Maßen für Modellvergleiche (sog. AIC-index) stützen.

Michael, W.B., Guilford, J.P., Fruchter, B. & Zimmermann, W.S. (1957). The description of spatial visualization abilities. *Educational Psychological Measurement*, 17, 185-199.

Rost, J. (1990). Rasch-models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 3, 271-282.

DIE PRÜFUNG DER GÜLTIGKEIT DES RASCH-MODELLS MITTELS DES MIXED RASCH-MODELLS

Strauß, Bernd & Rost, Jürgen (Kiel)

Die meisten Modelltests für das Rasch-Modell basieren auf der Partitionierung der Personenstichprobe in verschiedene Teilstichproben, z.B. Rohwertgruppen, und dem Vergleich der Itemparameterschätzungen aus diesen Teilstichproben. Der bedingte Likelihood-Quotienten-Test (Andersen, 1973) ist der bekannteste Modelltest, der auf diesem Vergleich beruht.

Verschiedene Studien zeigen allerdings, daß unter bestimmten Bedingungen dieser Test Verletzungen der Dimensionalität nicht mit ausreichender Sensitivität entdecken kann. In einigen Studien wird die Art der Teilung der Personenstichprobe, wie z.B. nach der splitter-item-Technik, diskutiert, um eine höhere Sensitivität bzgl. der Entdeckung von Dimensionalitätsverletzungen zu erreichen.

Das Mixed Rasch-Modell (Rost, 1990) ist die Generalisierung des Rasch-Modells zu einem Mischverteilungsmodell und verbindet den Latent-Trait-Ansatz mit dem der Latent-Class-Analyse. Dieses Modell erlaubt die Identifikation von Teilstichproben, in denen das Rasch-Modell gültig ist und kann somit auch als Modelltest eingesetzt werden. Das Teilkriterium ist nicht manifest. In dieser Arbeit werden verschiedene Studien an realen und simulierten Daten vorgestellt. Es zeigt sich die Überlegenheit des Mixed-Rasch-Modells als Modelltest für das Rasch-Modell gegenüber dem Andersen-Test im Hinblick auf die Entdeckung von Dimensionalitätsverletzungen.

Andersen, E.B. (1973). A goodness of fit test for the Rasch-model. *Psychometrika*, 38, 123-140.

Rost, J. (1990). Rasch-models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 3, 271-282.

KLASSIFIKATION BERUFLICHER INTERESSEN MIT MODELLEN DER LATENT-CLASS-ANALYSE

Tarnai, Christian (Münster)

Die Mängel der Klassischen Testtheorie sind hinlänglich bekannt und werden immer wieder betont. Diese Erkenntnis bleibt allerdings ohne Konsequenzen für ihre weitere Anwendung. Ein Grund hierfür mag sein, daß viele Modelle der Probabilistischen Testtheorie (z.B. Rasch-Modelle) restriktive Annahmen implizieren, die in der Praxis kaum erfüllt sind. Eine Alternative innerhalb der Probabilistischen Testtheorie mit weniger restriktiven Annahmen ist die Latent Class Analyse (LCA). Sie beinhaltet Modelle, die auf der Grundlage nominaler oder ordinaler Variablen die Personen in geeigneter Weise auf einer latenten Dimension klassifizieren kann. Dies entspricht den in der Praxis meist gegebenen Voraussetzungen hinsichtlich des Meßniveaus der Daten und der diagnostischen Zielsetzung, Personen als Mitglieder einer Gruppe zu identifizieren, der sie aufgrund der gemessenen Merkmale am wahrscheinlichsten angehören.

Für eine Stichprobe von Studierenden und Berufstätigen verschiedener Fächer bzw. Berufsfelder (N=400) werden u.a. die beruflichen Interessen mit dem Allgemeinen-Interessen-Struktur-Test (A-I-S-T) von Bergmann und Eder erhoben. Der A-I-S-T erfaßt sechs Persönlichkeitsorientierungen nach Holland (1985) mit 60 Tätigkeiten, die mit einem fünfstufigen free-choice Antwortformat dem subjektiven Interesse entsprechend beurteilt werden. Die Auswertung der Daten mit der Latent Class Analyse für ordinale Daten (Rost, 1990) gestattet die dem Meßniveau des Antwortformats angemessene Klassifikation der Probanden, um zu überprüfen, inwieweit die angenommenen Typen bzw. Mischtypen von Orientierungen in der Stichprobe abbildbar sind. Für die Auswertung des A-I-S-T mit Hilfe der LCA ergeben sich Vorteile, die u.a. darin bestehen, daß für jede Person die Wahrscheinlichkeit ihrer Zugehörigkeit zu einem Orientierungstyp angegeben werden

kann. Hier ergibt sich die Möglichkeit, die Sicherheit individueller Diagnose abzuschätzen. Anhand der vorliegenden Daten kann demonstriert werden, daß die LCA für viele Fragen der psychologischen Diagnostik die passenden Antworten geben kann.

Holland, J.L. *Making vocational choices. A theory of vocational personalities and work environments.* Englewood Cliffs, NJ.: Prentice Hall.

Rost, J. (1990). LACORD. Latent class analysis for ordinal variables. A FORTRAN Program (2nd ed.). Kiel: Institut für die Pädagogik der Naturwissenschaften.

MISCHVERTEILUNGS-RASCH-ANALYSE DES BENTON-TESTS

Zweschper, Hans (Konstanz)

Der Benton-Test gibt vor, visuelle Merkfähigkeit zu prüfen. Eingesetzt wird er hauptsächlich in der neuropsychologischen Diagnostik, aber auch z.B. auf dem Gebiet der Schullaufbahnplanung. Anders als in bisherigen Arbeiten, wird der Test in der vorliegenden Studie (N=581) nicht mit Methoden der klassischen Testtheorie, sondern mit neueren, probabilistischen Testmodellen analysiert.

1. U.a. wurde eine Prüfung der Gesamtform (30 Items) auf Homogenität vorgenommen. Diese war sowohl in CML-Ratio-Tests mit den Kriterien 'diagnostische Zugehörigkeit' und 'Testleistung', als auch in Mischverteilungs-Rasch-Analysen inhomogen. Mit Hilfe des AIC Kennwertes ließen sich drei, raschhomogene Teilpopulationen identifizieren. Eine inferenzstatistische Prüfung der einzelnen Items über R.A. Fisher's Informationsfunktion ergab, daß die Inhomogenität des Tests auf die Items 2, 5, 8, 15, 18, 23 und 24 zurückzuführen ist.

2. Ein Vergleich nicht modellkonformer Items aus probabilistischer und klassischer testtheoretischer Analyse zeigt, daß nur drei von zehn nicht modellkonformen Items aus der klassischen Testanalyse im Rahmen des RM ebenfalls nicht modellkonform sind.

3. Eine Mischverteilungs-Rasch-Analyse der einzelnen Subtestserien C, D und E zeigte, daß die drei Serien unterschiedlich gut modellkonform sind. Während Subtest D ausgesprochen homogen scheint und Subtest C als 'noch' homogen zu betrachten ist, indiziert der AIC für Subtest E eine zwei Klassen raschhomogene Lösung. Die Charakterisierung der ermittelten latenten Klassen des Subtests E durch personale Variablen ergab, daß überzufällig häufig diagnostisch unauffällige, jüngere und gut gebildete Personen in Klasse 1 (25%) vertreten sind. Klasse 2 (65%) dagegen umfaßt einen überzufällig hohen Anteil schizophrener, depressiver, älterer und weniger gebildeter Personen $CC_{KORR}=19\%-26\%$. Die Charakteristik der latenten Klassen könnte darauf schließen lassen, daß die inhomogenen Items unterschiedliche Lösungsstrategien provozieren. Allerdings ist der Zusammenhang zwischen latenter Klasse und diagnostischer Zugehörigkeit nicht hoch genug, um darüber einen zusätzlichen diagnostischen Indikator zu erhalten.

4. Insgesamt scheint die Inhomogenität des Bentontests den Rahmen von Störeinflüssen nicht zu überschreiten. Eine Hauptkomponentenanalyse mit drei Faktoren, in welche die

drei Subtests C, D, und E eingingen, zeigte, daß alle drei hoch auf dem ersten Faktor laden ($r \geq .89$), der 80,4% der Varianz aufklärt.

5. Eine nonparametrische einfache Varianzanalyse ergab, daß inhomogene Subtests per se nicht schlechter zwischen nosologischen Gruppen differenzieren, als dies homogene Subtests tun. Somit zeigte sich, daß das Scheitern des Tests an der Raschhomogenität den Test nicht automatisch unbrauchbar für diagnostische Zwecke werden läßt.