BERICHT
aus dem
PSYCHOLOGISCHEN INSTITUT
DER UNIVERSITÄT HEIDELBERG

SKALENPROBLEME BEI PROBABILISTISCHEN MESSMODELLEN

Heinrich Wottawa

Diskussionspapier Nr. 3
März 1976

6900 HEIDELBERG 1, HAUPTSTR. 47-51, TEL. 06221/547346

SKALENPROBLEME BEI PROBABILISTISCHEN

MESSMODELLEN

Heinrich W o t t a w a

Diskussionspapier Nr. 3

März 1976

Zusammenfassung

Ausgehend von allgemeinen Überlegungen über die Beziehung zwischen der Modellgleichung und theoriebezogenem latenten Parameter werden Überlegungen bezüglich der Skaleneigenschaften von probabilistischen Meßmodellen angestellt und an drei Beispielen (Thurstone-V, BTL, logistisches Modell von Rasch) näher diskutiert. Es zeigt sich, daß alle diese Modelle auch bei vollständiger Erfüllung der Modellvoraussetzungen nur Ranginformation bezüglich des latenten Parameters liefern (bezüglich des Modell-Parameters können stärkere Skalen erhalten werden). Daraus ergibt sich etwa, daß das BTL-Verfahren und Thurstone-V stets identische Skalen des latenten Parameters ergeben. Der Wert solcher Modellansätze ist demnach nicht in einer Verbesserung der Skaleneigenschaft der Meßwerte, sondern insbesondere in der Möglichkeit der Abtestung der inhaltlichen Vorstellungen, die zu dem jeweiligen Modell geführt haben, zu sehen. Spezielle Probleme ergeben sich für das linear-logistische Modell, insbesondere im Hinblick auf die Invarianz und Interpretierbarkeit der Regressionskoeffizienten.

Es stellt sich die Frage, ob man nicht trachten sollte, die formale Struktur psychologischer Theorien ausschließlich auf Ordnungsrelationen aufzubauen.

1. Einleitung

Die inhaltlichen Theorien in der Psychologie beschäftigen sich in aller Regel mit Konstrukten, die sich nicht direkt beobachten lassen (latene Dimensionen). Diese hypothetischen Begriffe (Intelligenz, Extroversion u. dgl.) können hypothetisch zueinander in Beziehung gesetzt werden, es können Systeme von miteinander verbundenen theoretischen Aussagen entstehen. Möchte man solche Annahmen jedoch empirisch überprüfen, müssen die in der Theorie vorkommenden latenten Begriffe erfaßbar gemacht werden.

Ein häufiger Vorgehen ist das Festlegen der latenten Dimension durch beobachtbare Ereignisse in der Art der operationalen Definition. Dies führt zu Aussagen wie "Intelligenz ist, was der Intelligenztest XY erfaßt". Ein solches Vorgehen zeichnet sich durch große Einfachheit aus, hat allerdings den allgemein bekannten Nachteil, daß etwa jeder denkbare Intelligenztest (unendlich viele) zu einem neuen Begriff der Intelligenz führt. Es hängt dann vom "Zufall", jedenfalls von inzidentellen Bedingungen ab, ob die nach der gewählten Operationalisierung erhaltenen Meßwerte die von der Theorie vermuteten Zusammenhänge zeigen.

Es dürfte dem Gegenstandsbereich wesentlich angemessener sein, Meßinstrumente zu erstellen, die bezüglich der latenten Dimension <u>Indikatoren</u> darstellen. Bei diesem Vorgehen ist die beobachtete Meßwertreihe nicht mit der latenten Dimension der Theorie identisch, sie steht aber mit dem theoretischen Begriff in einem vermuteten, plausibel erscheinenden und nach Möglichkeit prüfbaren Zusammenhang. Eine Konstruktion entsprechender Instrumente ist entweder unter Verwendung von axiomatischen meßtheoretischen Ansätzen oder mit Hilfe von probabilistischen Meßmodellen möglich. Da sich in vielen praktischen Situationen die axiomatischen Ansätze nicht verwenden lassen, sind die probabilistischen Ansätze und insbesondere die Rasch-Modelle für diese Fragestellung besonders wertvoll. Andere meßtheoretische Vorstellungen, so insbesondere die klassische Testtheorie, eignen sich für dieses Vorgehen kaum (insbesondere aufgrund der mangelnden Überprüfungsmöglichkeit der Dimensionalität des jeweiligen Meßinstrumentes).

a) die Existenz eines Wertes zur ausreichenden Kennzeichnung des zu messenden Objektes angenommen wird (entweder wird dies explizit angenommen oder es folgt aus geforderten Eigenschaften von Vergleichen zwischen Objekten). Diese Dimension wird im folgenden mit lp("latenter Parameter") abgekürzt

b) eine monotone Funktion m existiert, die die latente Dimension lp in die Dimension mp ("Modellparameter") überführt

c) die Beziehung zwischen Wahrscheinlichkeiten von beobachtbaren Ereignissen und der latenten Dimension (die Modellgleichung) ausgehend von der latenten Dimension mp formuliert wird, so etwa

c1) Thurstone V: $p(a \text{ vor } b) = NV(mp(a)-mp(b))$ NV...Standardnormalverteilung

c2) BTL: $p(a \text{ vor } b) = \frac{mp(a)}{mp(a)+mp(b)}$

c3) Logistisches Modell nach Rasch: $p("+"/v,i) = \frac{mp_1(v) \cdot mp_2(i)}{1+mp_1(v) \cdot mp_2(i)}$

d) die Skaleneigenschaften der Dimension mp aufgrund der Modellgleichung bestimmt werden und die gewünschten "starken" Skalen zeigen. Diese Skaleneigenschaften gelten aber nicht für die eigentlich untersuchte Dimension lp.

Für vermutlich alle inhaltlichen Fragestellungen - ausgenommen sind vielleicht Überlegungen zu bestimmten Modelltests - ist die Skaleneigenschaft der Dimension mp uninteressant. Über die Beziehung von lp und mp kann nach den Modellkonstruktionen nur gesagt werden, daß sie monoton sind und damit mp eine Rangskala für die Dimension lp ergibt. Welche monotone Transformation m in der jeweiligen Untersuchung vorliegt, hängt von zufälligen Komponenten der Untersuchung ab (etwa Formulierung der Fragen u.dgl.). Für die Theorienbildung in der Psychologie ist daher auch mit diesen Modellen nur Ranginformation zu erhalten, falls man nicht eine bestimmte Situation bzw. einen bestimmten Fragensatz festhalten und für die operationale Definition der latenten Eigenschaft verwenden möchte. Bei einem solchen Vorgehen wird jedoch der Ansatz über ein probabilististellt die Zahl der Bevorzugungen eines Reizes eine Rangskala bezüglich der Erfassung seines latenten Parameters dar. Weitere Annahmen sind für das Erhalten dieser Rangskala nicht erforderlich.

Für das Thurstone-Modell-Case-V wird jetzt zusätzlich gefordert, daß die monotone Beziehung T1) für alle Reizpaare eine spezielle mathematische Form hat, und zwar bei entsprechender Normierung der Modellparameter die Gestalt der Summenkurve der Standardnormalverteilung (NV).

T2) $p(a > b) = NV(mp\ (a) - mp(b))$

Zu beachten ist, daß diese mathematische Beziehung für das Thurstone-Modell als Meßmodell nicht auf der Dimension der latenten Parameter, sondern auf der Dimension der Modellparameter definiert ist. Anders liegen die Verhältnisse, wenn man das Thurstone-Modell als inhaltliches Modell annimmt und die Meinung vertritt, daß aufgrund der menschlichen Gehirnstrukturen entsprechende Entscheidungs- bzw. Verwechslungsmechanismen empirisch vorhanden sind. In diesem Fall müßte die Dimension mp für verschiedene Fragestellungen identisch sein, es müßte möglich sein, mit verschiedenen Experimenten auf das gleiche - evtl. physiologische - "Ereignis" Zugriff zu nehmen. Als Meßmodell, also als formaler Ansatz zum Erhalt von Meßwerten mit wünschenswerten Eigenschaften, kann eine solche inhaltliche Forderung nicht aufgestellt werden. Durch die Forderung T2 wird verlangt, daß die zu vergleichenden Reize durch eine ganz spezielle mathematische Struktur charakterisiert sind. Es ist nicht möglich, durch entsprechende "Stauchung" oder "Streckung" der Dimension lp die Bedingung T2 stets zu erfüllen. Ausgehend von der Forderung T1 kann nur gesagt werden, daß mit zunehmender Ausprägung des latenten Parameters des Reizes in der 1. Position die Wahrscheinlichkeit diesen Reiz zu wählen für alle Vergleichsreize zunimmt. Welche Gestalt diese Zunahme hat, kann nicht gesagt werden (vgl. Diagramm 1).

Da mit Ausnahme dieser Monotonie nichts über die Dimension lp ausgesagt ist, kann für den Vergleich eines Reizes mit allen anderen jede gewünschte monotone Funktion durch eine entsprechende der Bedingung T2 voraus, die mit entsprechend trennscharfen Modelltests geprüft werden müßte. Da bei dem routinemäßigen Einsatz dieses Verfahrens zur Analyse von Paarvergleichen in aller Regel keine inhaltlichen Theorien zugrundeliegen, erscheint die zusätzliche Aufnahme von T2 eigentlich überflüssig. Der damit verbundene relativ hohe Rechenaufwand bringt keine essentielle Verbesserung der erhaltenen Skalenstruktur. Man erhält zwar eine Intervallskala für die Modellparameter, nicht aber bei dem untersuchten theoretischen Konstrukt.

3. BTL

Ausgehend vom Wahlaxiom (s. etwa Luce (1963)) erhält man die Modellgleichung

$$\text{BTL 1): } p(\text{a vor b}) = \frac{mp(a)}{mp(a)+mp(b)}$$

Die Skala mp ist aufgrund der Modellgleichung eindeutig bis auf eine multiplikative Konstante, also auf einer Rationalskala festgelegt.
Im Gegensatz zu Thurstone, dessen Ansätze sich auf inhaltliche Vorstellungen beziehen, entstand dieser Ansatz aus der Forderung nach einem guten Meßinstrument entsprechend dem choice-axiom. Ob ein vorliegender Datensatz eine in diesem Sinne "gute" Messung gestattet, ist wieder eine empirische Frage. Es muß nicht sein, daß eine entsprechende Stauchung der latenten Skala besteht, die zu dieser funktionalen Darstellung der Bevorzugungswahrscheinlichkeiten führt. Ebenso wie bei Thurstone kann für den Vergleich eines Reizes mit allen anderen die Erfüllung der Modellgleichung garantiert werden.
Auf den Unterschied zwischen der latenten Struktur und dem Modelparameter weisen bereits Suppes und Zinnes (1963) hin. So läßt sich etwa das Modell auch anschreiben als

$$p(\text{a vor b}) = \frac{e^{lp(a)}}{e^{lp(a)}+e^{lp(b)}}$$

Wenn man etwa setzt $mp(a) = e^{lp(a)}$.
Da die Dimension mp auf einer Rationalskala erfaßt wird, würde mit diesem Ansatz die Dimension lp auf einer Differenzskala erfaßt. Ebenso könnte man aber auch setzen

$$mp(a) = lp(a)^k + lp(a) \qquad (lp, k \text{ positiv})$$

nicht zu einer realitätsinadäquaten Interpretation von "starken" Skalen der erhaltenen Parameterschätzwerte führen.

4. Das logistische Modell von Rasch

Im Gegensatz zum BTL-Verfahren, bei dem die Objekte untereinander verglichen werden, erfolgt bei einem Vorgehen nach den Rasch-Modellen der Vergleich von Objekten aufgrund von sogenannten Agentien (meist Items). Das Kernstück eines solchen Modells ist stets die Annahme einer speziellen mathematischen Form der sogenannten Itemcharakteristik. Im Gegensatz zu den älteren Ansätzen von Lazarsfeld (1950) und Guttman (1946) werden bei Rasch die Itemcharakteristiken nicht aufgrund von formalen Kriterien festgelegt, sondern aus inhaltlichen Forderungen (im wesentlichen der sogen. spezifischen Objektivität) bestimmt. Aus der Forderung nach Eindimensionalität des latenten Parameters (wie sie etwa zwingend bei der Betrachtung von zweikategoriellen Antworten auftritt) und der Teilforderung der spezifischen Objektivität nach Unabhängigkeit des Vergleiches zwischen Personen von den dieselbe latente Dimension erfassenden zufällig gewählten Items folgt, daß die Itemcharakteristiken als eine monotone Funktion des eindimensionalen Parameters darstellbar sind (s. Diagramm 3).

Fordert man zusätzlich, daß der Vergleich der Items unabhängig von den zufällig herausgegriffenen Personen ist, darf ein Meßinstrument keine Items enthalten, deren Charakteristiken sich kreuzen (da der Vergleich der Items vor und nach der Schnittstelle eine andere Aussage über deren Schwierigkeit erbringen würde). Durch diese Forderung wird die Vielfalt der in einem Meßinstrument zugelassenen Itemcharakteristiken stark eingeschränkt. Verlangt man überdies noch, daß die Zahl der gelösten Aufgaben eine erschöpfende Statistik für die Ausprägungsgrade der Modellparameter sein soll, folgt (vgl. etwa Fischer 1968), daß sich die Itemcharakteristiken in der Form darstellen lassen:

R1) $$p(\text{"+"}/v,i) = \frac{o_v \cdot s_i}{1+o_v \cdot s_i}$$

Die üblicherweise verwendeten Parameterbezeichnungen beziehen sich auf die Dimension des Modellparameters, und nicht auf die

Items eliminiert), werden die neuen Ergebnisse für die erste Person wieder 10 zeigen, für die zweite Person evtl. 13 und für die letzte Person etwa 15. Die Veränderung der Meßwertreihen 10,15,19 zu 10,13,15 ist nicht linear, so daß die Annahme einer Intervallskala für die Zahl der gelösten Aufgaben nicht beibehalten werden kann. Im Gegensatz dazu sind die Schätzwerte der Modellparameter bei Anwendung des logistischen Modells zwar unabhängig von der zufälligen Zusammensetzung der Itemschwierigkeiten, die jeweilige Skalenform wird jedoch durch die ebenfalls zufällige Zusammensetzung der Itemcharakteristiken(verschiedene "Trennschärfen")und eine entsprechende Stauchung bedingt. Man vermeidet bei Berechnung der Modellparameter die eine Störquelle, muß dafür jedoch eine andere in Kauf nehmen.

Diese Schwierigkeiten bei der Interpretation der Skalenwerte im Rasch-Modell sind wohl seit langem bekannt. So findet man etwa standardmäßig bei den Programmausdrucken sowohl die oben dargestellte Modellform (R1) als auchden exponentiellen Ansatz. In der obigen Form werden für die Dimension mp Rationalskalen erhalten, bei dem exponentiellen Ansatz Differenzenskalen. Dies ist merkwürdig, da man stets den gleichen Datensatz mit zwei völlig gleichwertigen Modellvorstellungen untersucht. Trotzdem findet man häufig die Aussage, daß die entsprechenden Modellparameter für die inhaltlich relevante <u>latente</u> Dimension das Niveau einer Rationalskala zeigen. Man überzeugt sich - analog zum BTL - leicht, daß dies nicht der Fall sein kann. Alle inhaltlichen Forderungen der spezifischen Objektivität bzw. des entsprechenden Verlaufes der Itemcharakteristiken und auch die Forderung nach einer erschöpfenden Statistik bleiben voll bestehen, wenn man allgemein ansetzt.

$$o_v = m(lpo_v)$$

$$s_i = n(lps_i)$$

m,n monotone Funktionen

$$p("+"/v,i) = \frac{m(lpo_v)\cdot n(lps_i)}{1+m(lpo_v)\cdot n(lps_i)}$$

also etwa

$$p("+"/v,i)=\frac{((1lpo_v)^3+3(lpo_v)^2+\ln(1+lpo_v))\cdot((lps_i)^{1/2}+5lps_i)}{1+((lpo_v)^3+3(lpo_v)^2+\ln(1+lpo_v))\cdot((lps_i)^{1/2}+5lps_i)}$$

für stets positives lpo und lps.

nahme stärkerer Skalen, kann eine evtl. Übereinstimmung verdeckt werden. Da man meist nur an einer "guten" Messung der Personen interessiert ist, schränken solche Modelltests die praktische Anwendung dieses Verfahrens eher unnötig ein.

5. Probleme beim linear-logistischen Modell

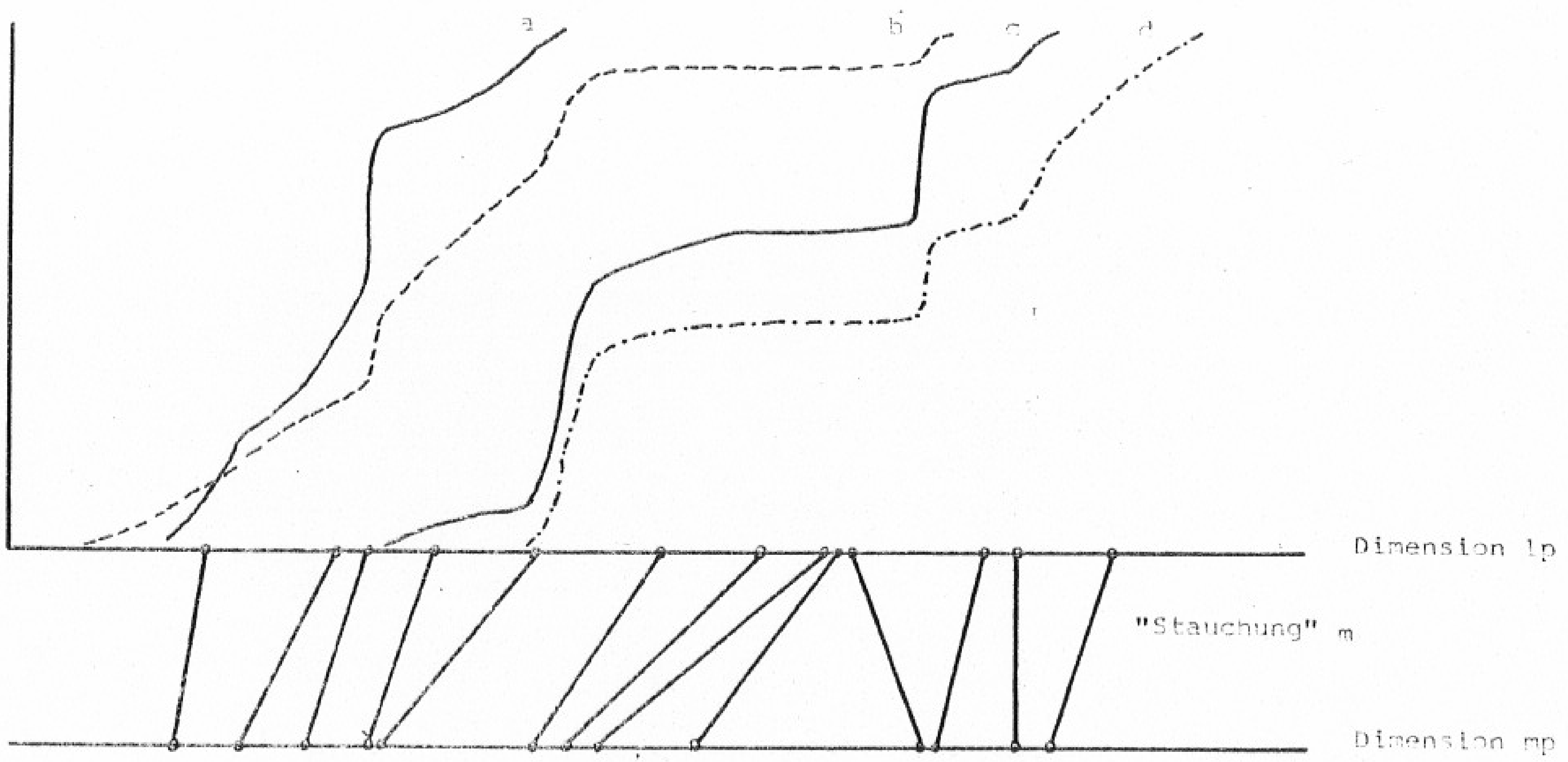
Die Grundvorstellung dieses zuerst vermutlich von Cox (1970) vorgestellten Ansatzes besteht darin, daß der "Itemparameter" oder der "Personenparameter" der in exponentieller Form angeschriebenen logistischen Funktion als Linearkombination von verschiedenen Prädikatoren dargestellt wird. Die Ausprägungsgrade der Prädiktoren werden als bekannt angenommen, die Gewichte, die den Prädiktoren bezüglich des zu beschreibenden Parameters zukommen, werden aus dem Datenmaterial geschätzt. Cox verwendete diesen Ansatz ursprünglich zur statistischen Analyse von zweikategoriellen Daten, er betrachtete nur die Darstellung der Antworten auf ein einziges Item. Die Verallgemeinerung dieses Ansatzes auf mehrere eindimensionale Items im Rahmen der Meßtheorie von Rasch findet sich etwa in Fischer (1974). Bei der üblichen Anwendung von Regressionsansätzen ist es in aller Regel fraglich, ob die abhängige Variable die für einen linearen Ansatz notwendige Intervallskaleneigenschaft aufweist. Meist wird dies nicht erfüllt sein.

Geht man entsprechend dem Abschnitt 4 von der Feststellung aus, daß die logistischen Modelle nur Ranginformation bezüglich der latenten Parameter liefern, so würde eine nachträgliche regressionsanalytische Behandlung der geschätzten Modellparameter zu den gleichen Problemen führen wie man sie etwa bei der Verwendung von Meßwerten, die im Rahmen der klassischen Testtheorie erhalten wurden, bezüglich der Interpretation von Regressionsansätzen kennt.

Bei der Verwendung eines linear-logistischen Modells geht man aber in der Regel nicht von einer nachträglichen Analyse der Modellparameter aus, sondern es wird der Regressionsansatz in die Modellstruktur aufgenommen. Daraus kann sich eine weitere Schwierigkeit ergeben. Da jede Methode der Parameterschätzung

Es soll ausdrücklich betont werden, daß mit diesen Ausführungen in keiner Weise der Nutzen und die große Bedeutung der probabilistischen Meßmodelle in Frage gestellt werden soll. Insbesondere für das BTL und die Rasch-Modelle sind keine Alternativen in Sicht, die auch nur annähernd gleichwertige Beiträge zum Problem der Eindimensionalität von Meßwerten erbringen können. Es stellt sich allerdings die Frage, ob der Hauptvorteil der Verwendung solcher Ansätze in der gelegentlich behaupteten Verbesserung der Skaleneigenschaft ("wie in den Naturwissenschaften") liegt. Es ist zu vermuten, daß eine solche starke Messung in der Psychologie erst dann erfolgen kann, wenn die Theorienbildung halbwegs das quantifizierbare Niveau der Naturwissenschaften erreicht hat. Dazu dürfte es aber ein weiter Schritt sein, falls überhaupt im sozialwissenschaftlichen Bereich eine solche Theorienbildung möglich sein sollte. Da die erwähnten Skalenprobleme nicht nur für die beispielhaft aufgeführten besonders bekannten Meßmodelle gelten, sondern für alle probabilistischen Modelle auf der Grundlage von latenten Parametern, stellt sich die Frage, ob es überhaupt möglich ist, mit Hilfe von probabilistischen Ansätzen zu "starken" Skalen zu kommen. Da als einzig vertretbare Alternative derzeit die axiomatischen Meßsysteme anzusehen sind und diese aufgrund ihrer restriktiven Forderungen in praktischen Situationen nur selten verwendet werden können, sollte man vielleicht den Anspruch an die Theorienbildung geringer halten. Evtl. sollte man anstreben, die den psychologischen Theorien zugrundeliegenden Eigenschafts- bzw. Dimensionsbegriffen eine ausschließlich auf Rangordnungen aufbauende formale Interpretation zugrunde zu legen. Die dabei auftretenden Probleme (so kann etwa eine Menge von Items (a, b, c) eindimensional sein, die Menge (a, b, d) ebenfalls, die Zusammenfassung (a, b, c, d) nicht) dürften allerdings nur schwer zu lösen sein. Auch wäre es erforderlich, daß man die der Theorienkonstruktion meist implizit zugrunde liegende Anschauung von "Dimensionen" und verwandten Begriffen (Eigenschaft etc.) auf dieser Grundlage verändert. Trotz dieser nicht unbeträchtlichen Schwierigkeiten wäre ein solches Vorgehen vielleicht dem psychologischen Gegenstandsbereich angemessener und überdies - zumindest nach dem heutigen Entwicklungsstand psychologischer Theorien - ehrliche

Diagramm 1



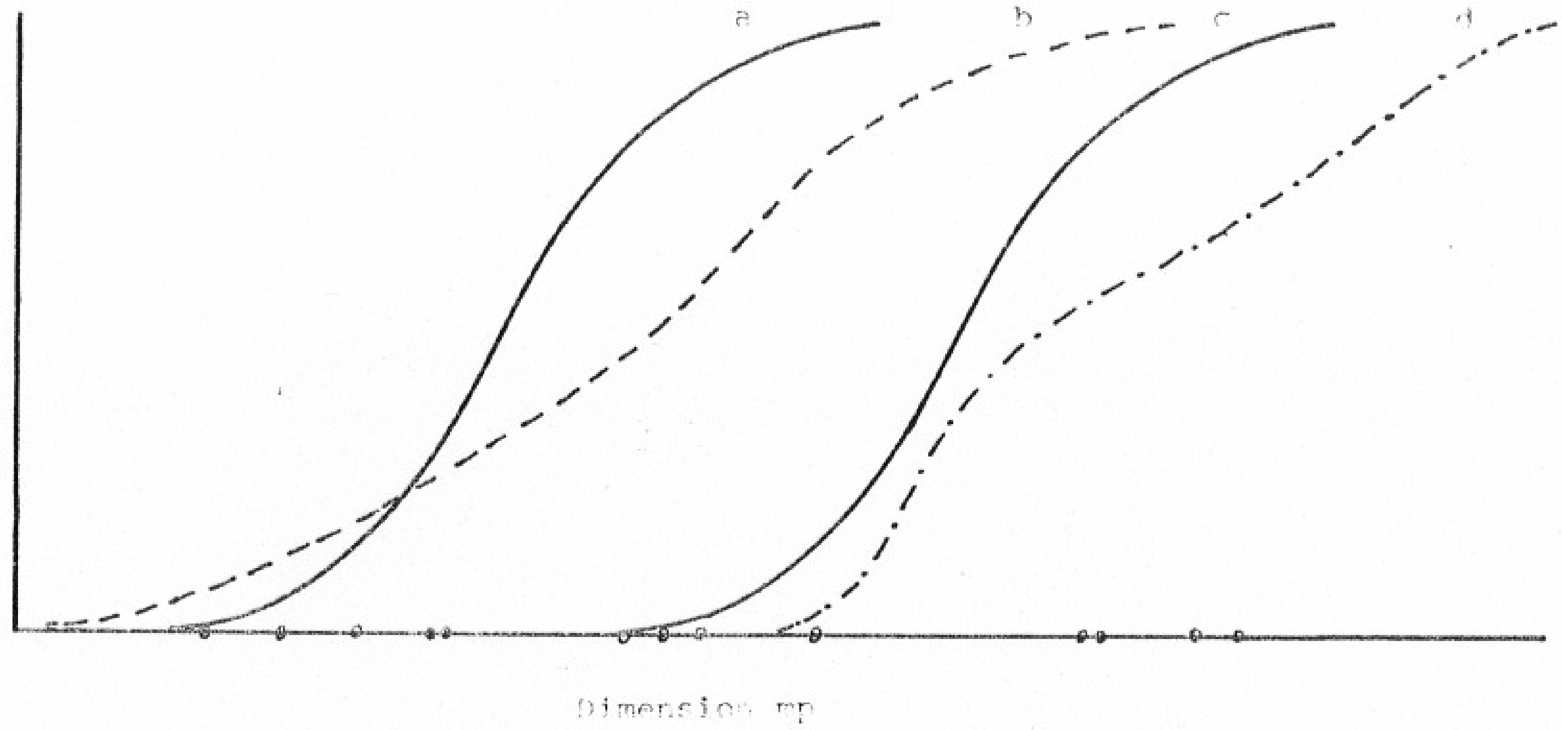

a
b
c
d
Dimension mp

Diagramm 3