



Skalenkonstruktion nach Mokken für mehrdimensionale Variablenstrukturen

Ein Anwendungsbeispiel mit SPSS

Leila Akremi und Markus Ziegler

Bamberger Beiträge zur empirischen Sozialforschung

Herausgeber:

Gerhard Schulze und Leila Akremi

Nr. 14, 2007

ISSN 1611-583X

Bamberger Beiträge zur empirischen Sozialforschung

1	Einführung in die Methoden der empirischen Sozialforschung (2., überarbeitete Auflage)	Gerhard Schulze	2004
2	Einführung in die Methoden der empirischen Sozialforschung. Übungsaufgaben und Lösungen (2., überarbeitete Auflage)	Gerhard Schulze	2004
3	Die biographische Methode. Ein Verfahren zur qualitativen Analyse individueller Verlaufsmuster in den Sozialwissenschaften (2., unveränderte Auflage)	Nina Baur	2003
4-1	Sozialwissenschaftliche Datenanalyse am PC für Fortgeschrittene. Ein Arbeitsbuch mit SPSS für Windows	Nina Baur	2003
5	Takeoff der Auswertung. Zur Vorbereitung statistischer Analysen	Nina Baur	2003
8	Univariate Statistik. Eine Einführung in den Umgang mit eindimensionalen Häufigkeitsverteilungen mit SPSS für Windows	Nina Baur	2003
9	Bivariate Statistik, Drittvariablenkontrolle und das Ordinalskalensproblem. Eine Einführung in die Kausalanalyse und in den Umgang mit zweidimensionalen Häufigkeitsverteilungen mit SPSS für Windows (2., korrigierte Auflage)	Nina Baur	2003
10	Multiple lineare Regressionsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	2003
11	Binäre logistische Regressionsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	2005
12	Faktorenanalyse und Reliabilitätsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	2003
13	Wie kommt man von den Ergebnissen der Faktorenanalyse zu Dimensionsvariablen? Eine Einführung in die Dimensionsbildung mit SPSS für Windows (2., korrigierte Auflage)	Nina Baur	2003
14	Skalenkonstruktion nach Mokken für mehrdimensionale Variablenstrukturen – Ein Anwendungsbeispiel mit SPSS	Leila Akremi / Markus Ziegler	2007
19-1	Quantitative Analyse zeitlicher Veränderung Band 1: Überblick und theoretische Grundlage	Nina Baur (Hg.)	2003
20	Präsentation statistischer Daten Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Simone Zdrojewski / Jan D. Engelhardt	2003
21	Datenaufbereitung. Arbeitsschritte zwischen Erhebung und Auswertung quantitativer Daten	Detlev Lück	2003
22	Arbeiten mit MAXqda. Kurze Einführung in die computergestützte Analyse qualitativer Daten (2., korrigierte Auflage)	Britta Wagner	2003
23	Wo liegen die Grenzen quantitativer Längsschnittsanalysen?	Nina Baur	2004

ISSN: 1611-583X

Herausgegeben von Gerhard Schulze und Leila Akremi
Professur für Methoden der empirischen Sozialforschung
Otto-Friedrich-Universität Bamberg
96045 Bamberg

http://www.uni-bamberg.de/fakultaeten/sowi/fachgebiete/soziologie/profes-sur_fuer_methoden_der_empirischen_sozialforschung_und_wissenschaftstheorie/leistungen/forschung/

Inhalt

1	Vorüberlegungen zu Skalierungsverfahren.....	5
1.1	Skalierung und Messmodelle	5
1.2	Mögliche Formen des Zusammenhangs zwischen der Antwortfunktion von Items und der latenten Dimension – Verschiedene Messmodelle	5
1.3	Einführung des Variablenbeispiels	8
2	Erläuterung der wesentlichen Konzepte des Skalierungsverfahrens nach Mokken.....	8
2.1	Anwendungsvoraussetzungen	9
2.2	Skalierungsprozedur mit dem Variablenset	12
2.3	Abschluss des Skalierungsprozesses: Bildung der Skalenwerte am Beispiel.....	19

1 Vorüberlegungen zu Skalierungsverfahren

Latente Variablen spielen in der empirischen Sozialforschung eine zentrale Rolle. Durch sie können komplexe Eigenschaftsbündel in einem Begriff zusammengefasst und von der individuellen auf die kollektive Ebene abstrahiert werden. Diese Variable werden deshalb als „latent“ bezeichnet, weil sie nicht direkt beobachtbar sind. Die konkrete Ausprägung dieser Variablen beim Individuum kann daher auch nicht erfragt werden.

Wir müssen darum manifeste Sachverhalte finden, die mit dieser latenten Dimension in Beziehung stehen und die wir auch erheben können. Diese so erhobenen Werte stehen zunächst nur für sich. Mit Hilfe von Skalierungsverfahren wird ein Weg gesucht, die Items zu verdichten und in einer einzigen (latenten) Variable zusammenzuführen. Dabei stellen sich grundsätzlich folgende Fragen:

- Repräsentieren die erhobenen Items die Dimension, die unseren theoretischen Überlegungen entspricht?
- Sind alle Items zur Dimensionsbestimmung geeignet?
- Lässt sich auf Basis des Itempools nur eine Dimension bestimmen oder liegen mehrdimensionale Strukturen vor?
- Welche Ausprägung besitzen die Befragten auf der latenten Dimension (Skalenwerte)?

1.1 Skalierung und Messmodelle

Die verschiedenen Skalierungsverfahren stellen unterschiedliche Möglichkeiten der Beantwortung dieser Fragen dar. Sie basieren auf Messmodellen, die letztlich nichts anderes sind, als Hypothesen über die Form des Zusammenhangs von Items und latenter Dimension.

Für die Mokkenskalisierung sind vor allem die Messmodelle nach Guttman und Rasch von Bedeutung. Sie sind für monotone dichotome Items entwickelt worden. Da diese Skalierungsverfahren zuerst in der Psychologie Anwendung fanden, lässt sich die Logik, die diesen Modellen zugrunde liegt, vorab am anschaulichsten mit einem Beispiel aus psychologischen Tests erläutern. Die hierfür interessierende latente Variable soll Intelligenz sein¹. Mittels eines Intelligenztests werden verschiedene Aufgaben entwickelt, die Aussagen über die Ausprägung der Intelligenz einer bestimmten Person zulassen. Die gestellten Aufgaben entsprechen Intelligenzitems, die gelöst („bejaht“) oder nicht gelöst („verneint“) werden. Es ist plausibel anzunehmen, dass eine Person um so intelligenter ist, je mehr Aufgaben sie lösen (bejahen) kann². Ebenso lässt sich nachvollziehen, dass die Lösung der Aufgaben auch von deren Schwierigkeit abhängt. Dies sind bereits die zwei wichtigsten Sachverhalte für die oben erwähnten Skalierungsverfahren. Über die Wahrscheinlichkeit, eine Aufgabe zu lösen, wird ein Zusammenhang zur latenten Dimension hergestellt. Je größer die Wahrscheinlichkeit, desto stärker muss die Dimension ausgeprägt sein. Die Wahrscheinlichkeit, eine Aufgabe zu lösen, hängt also von den zuvor beschriebenen Bedingungen ab:

- Personenparameter: Ausprägung der latenten Variable bei einer bestimmten Person (z.B. wie intelligent ist sie?)
- Itemparameter: Ausprägung der latenten Variable beim Item (wie schwer ist die gestellte Aufgabe?)

Dies impliziert, dass sich sowohl die Items nach ihrem Ausprägungsgrad auf der Dimension in eine Rangfolge bringen lassen (z.B. leichte, mittlere und schwere Aufgaben) als auch die befragten Personen (z.B. wenig intelligente, durchschnittlich intelligente und hoch intelligente Personen). Der Unterschied zwischen den drei Skalierungsverfahren besteht nun in den Vorannahmen über den möglichen Zusammenhang zwischen der Antwortfunktion (Bejahungswahrscheinlichkeit der verschiedenen Items in Abhängigkeit von Personen- und Itemparameter) und der Ausprägung auf der latenten Dimension.

1.2 Mögliche Formen des Zusammenhangs zwischen der Antwortfunktion von Items und der latenten Dimension – Verschiedene Messmodelle

a) Guttmanskalierung

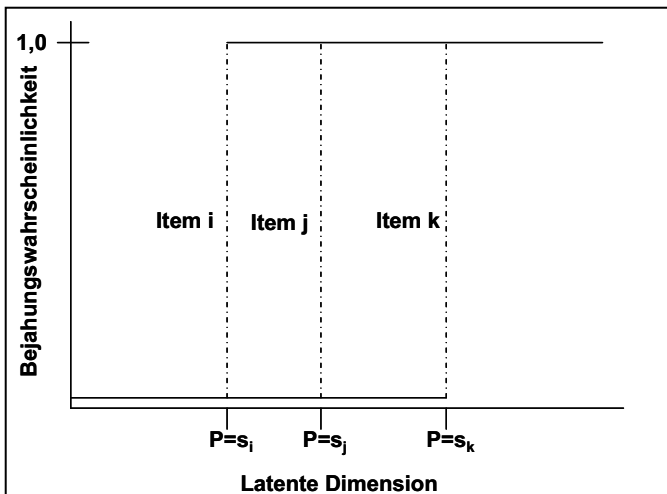
Guttman geht von einem deterministischen Zusammenhang aus (vgl. Meyer 2004: 46). Solange die gestellten Intelligenzaufgaben zu schwer für die Fähigkeiten einer Person (Ausprägung dieser Person auf der latenten Dimension Intelligenz) sind, ist die Bejahungswahrscheinlichkeit (Lösung einer Aufgabe) gleich Null. An der Stelle, an der die Aufgabe genau soviel Intelligenz erfordert, wie es der Fähigkeit einer befragten Person entspricht (Itemparameter = Personenparameter → beide besitzen denselben Ausprägungsgrad auf

¹ vgl. hierzu ausführlich Gerich 2001, Meyer 2004

² Voraussetzung ist natürlich, dass die Items dafür auch wirklich geeignet sind.

der latenten Variablen) springt die Bejahungswahrscheinlichkeit sofort auf den Wert 1 und verändert sich nicht mehr, da ab dieser Stelle die Ausprägung der latenten Variable Intelligenz immer größer wird. Graphisch lassen sich diese Itemcharakteristiken (tracelines) wie folgt darstellen.

Abb.1: Lösungswahrscheinlichkeit bei Guttman-Items

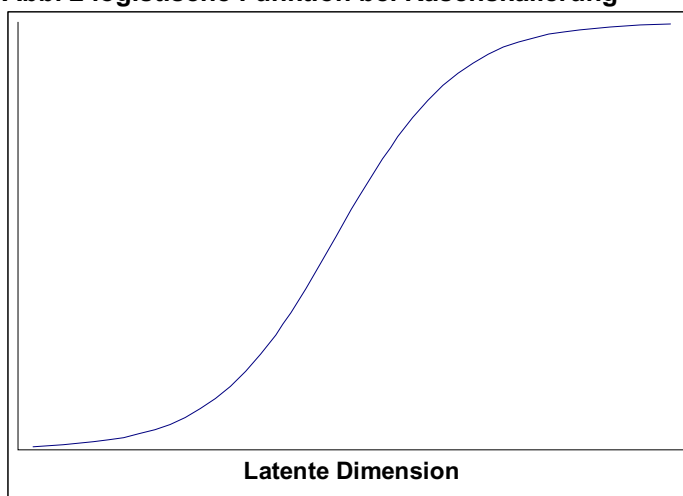


Prinzipiell lässt sich diese Modellannahme mit dem Beispiel des Intelligenztests gut nachvollziehen. Problematisch ist nur, dass die Guttman-Skalierung streng genommen keine Ausnahmen zulässt. Es ist also nicht möglich, eine schwere Aufgabe zu lösen und bei einer leichten zu scheitern. Abgesehen davon, dass die Schwierigkeit der Items im Einzelfall durchaus unterschiedlich empfunden werden kann, lässt sich leicht gegen diese Annahme argumentieren, dass es oftmals die leichten Aufgaben sind, die Personen in Testsituationen irritieren und zu Fehlern führen können.

b) Raschskalierung

Der Determinismus der Skalierungsannahmen ist der entscheidende Kritikpunkt am Verfahren nach Guttman. Bei soziologischen Dispositionsbegriffen ist eine solche deterministische Antwortfunktion unrealistisch. Es ist durchaus plausibel, dass Fehler im Sinne der Guttman-Skalierung auftreten. Eine alternative Herangehensweise ist eine probabilistische Modellkonstruktion wie z.B. die Raschskalierung. Demnach gibt es kein „mechanistisches“ Antwortverhalten der Befragten. Die Bejahungswahrscheinlichkeit ändert sich nicht sprunghaft, sondern nähert sich mit höherer Ausprägung der latenten Variablen dem Wert 1 an. Deshalb verwendet Rasch als Itemcharakteristik eine logistische Funktion (s. Abb. 2).

Abb. 2 logistische Funktion bei Raschskalierung



Die Steigung dieser Funktion ist nicht an jeder Stelle gleich. Für das Intelligenzbeispiel bedeutet dies, dass es Ausprägungsbereiche gibt, in denen sich die Wahrscheinlichkeit der Lösung einer Aufgabe mit zunehmender Intelligenz nur ganz leicht verändert, während in anderen Bereichen eine höhere Intelligenzausprägung zu deutlich größeren Lösungschancen führt.

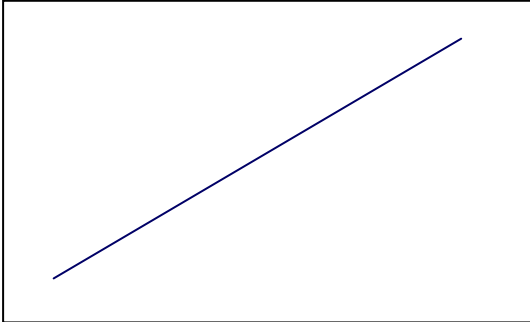
Die Items müssen, wie bereits erwähnt, dichotom sein und **eine Dimension** messen. Im Gegensatz zur Mokkenskalisierung kann die Rasch-Skalierung daher keine mehrdimensionalen Strukturen entdecken.

Weitere Bedingung für die Anwendung der Raschskalierung ist, dass die Itemcharakteristik der Items eine monoton steigende Funktion der latenten Variable repräsentieren (vgl. Gerich 2001: 31).

c) Andere funktionale Zusammenhänge

Messtheoretisch ergibt sich etwa bei der Raschskalierung, dass mit der Spezifizierung der Funktion im Grunde auch eine theoretische Annahme darüber getroffen wird, wie sich Einzelitems zum latenten Kontinuum verhalten. Gerade im sozialwissenschaftlichen Kontext stehen wir aber vor dem Problem, dass die funktionale Beschaffenheit des Zusammenhangs von Item und Dimension in der Regel nicht genau präzisiert werden kann oder schlicht unbekannt ist. Verwendet man etwa die Faktorenanalyse als mehrdimensionale Erweiterung der klassischen Testtheorie, so geschieht dies unter der Hypothese eines linearen funktionalen Zusammenhangs (vgl. Abb.3).

Abb. 3: linearer funktionaler Zusammenhang



Grundsätzlich können monotone Items aber auch andere funktionale Zusammenhänge aufweisen. Zwei mögliche Funktionsverläufe werden in den Abbildungen vier und fünf dargestellt.

Abb. 4: logarithmische Funktion

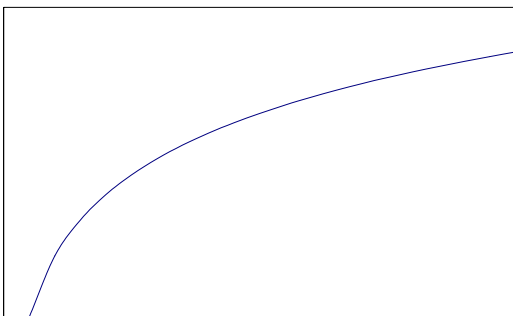
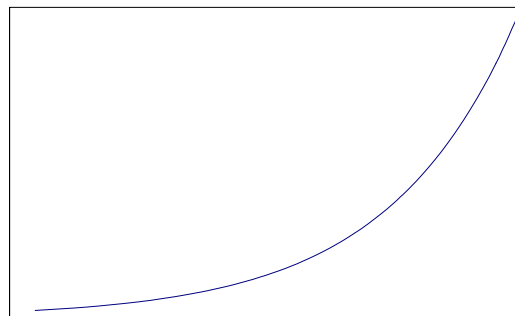


Abb. 5: Exponentialfunktion



Um nicht auf der Basis von unzutreffenden Vorannahmen und weiteren nicht erfüllbaren Modellrestriktionen falsche Schlussfolgerungen zu ziehen, entwickelte Mokken ein Skalierungsverfahren, das mit sehr wenigen Voraussetzungen auskommt und keine bestimmte Form der Itemcharakteristiken festlegt. Auf dieses Verfahren gehen wir im nächsten Abschnitt anhand eines Variablenbeispiels aus der Allgemeinen Bevölkerungsumfrage (ALLBUS) 2004 ein. Die Berechnungen sind mit einem von Leila Akremi und Patrik Budenz entwickelten Zusatzprogramm für SPSS³ durchgeführt worden.

³ Es wird dazu kein weiteres Programm benötigt, da das Mokkentool direkt in SPSS integriert ist. Damit können auch anschließende Analysen mittels SPSS durchgeführt werden. Das Programm zur Mokkenskalierung befindet sich derzeit noch in der Testphase und kann inklusive der Programmdokumentation angefordert werden bei:

Leila Akremi, Technische Universität Berlin, Fakultät VI: Planen – Bauen – Umwelt, Institut für Soziologie, Fachgruppe Methodenlehre, Sekretariat FR 2-5, Franklinstr. 28/29, 10587 Berlin;
Email: leila.akremi@tu-berlin.de

1.3 Einführung des Variablenbeispiels

In der Allgemeinen Bevölkerungsumfrage sollten die befragten Personen u.a. beurteilen, welche Aspekte für den Aufstieg in der deutschen Gesellschaft von Bedeutung sind. Folgende Variablen wurden zur Messung verwendet:

Tab. 1: Variablenliste ALLBUS 2004

Variablen		1 sehr wichtig	2 wichtig	3 weniger wichtig	4 unwichtig
v149	Opportunismus, Rücksichtslosigkeit				
v150	Bildung, Ausbildung				
v151	Politische Betätigung				
v152	Zufall, Glück				
v153	Intelligenz				
v154	Beziehungen, Protektion				
v155	Leistung, Fleiß				
v156	Geld, Vermögen				
v157	Initiative, Durchsetzung				
v158	Herkunft, Richtige Familie				
v159	Bestechung, Korruption				
v160	Kooperation, Offenheit				

Die Befragten konnten bei der Beantwortung zwischen den Kategorien sehr wichtig, wichtig, weniger wichtig und unwichtig abstimmen. Für die Mokkenskalierung benötigen wir jedoch dichotome Variablen, weshalb für dieses Beispiel die Kategorien sehr wichtig und wichtig zusammengefasst wurden (neuer Wert „1“), ebenso weniger wichtig und unwichtig (neuer Wert „0“). Die neu gebildeten Variablen heißen v149a bis v160a. Des Weiteren dürfen keine fehlenden Werte bestehen. Dazu haben wir alle Fälle ausgeschlossen, die bei einer der Variablen einen fehlenden Wert aufwiesen. Insgesamt gehen von den ursprünglich 2946 Fällen 2594 in die Analyse ein. Mittels der Mokkenskalierung soll nun geklärt werden, ob diesen Variablen eine Dimension zugrunde liegt oder ob sich mehrere finden lassen.

2 Erläuterung der wesentlichen Konzepte des Skalierungsverfahrens nach Mokken

Wir haben im vorangegangenen Abschnitt verschiedene Möglichkeiten der Messung latenter Variablen kennen gelernt. Das Skalierungsverfahren nach Mokken bildet nun eine Kombination dieser Ansätze.

Von Guttman übernimmt Mokken dabei die Annahme monotoner Verläufe der Antwortwahrscheinlichkeiten bei ansteigendem Skalenwert. Dies hat konkret zwei Implikationen:

- (1) Zum einen steigt generell und auch für jedes Item die Antwortwahrscheinlichkeit mit steigender Ausprägung der latenten Variable.
- (2) Zum anderen können die Items untereinander in eine eindeutige Rangfolge gebracht werden, die über das gesamte Spektrum der latenten Dimension aufrechterhalten wird.

Mokken geht also wie Guttman davon aus, dass eine Person ein schwierigeres Item nur dann bejahen kann, wenn sie auch das leichtere Item mit „Ja“ beantworten konnte. Er übernimmt aber nicht das deterministische Prinzip Guttmans. Für die Tracelines, d.h. den graphischen Verlauf der Funktion, trifft er keine impliziten

Annahmen. Wir sehen hier bereits die grundlegenden Vorteile der Mokkenskalierung: Im Gegensatz zu anderen Verfahren (vgl. Abschnitt 1) wird das Problem der Anpassung des Modells an die vorgegebenen Daten umgegangen, indem schlichtweg keine Annahmen über den Verlauf der Tracelines getroffen werden. Dadurch ist das Modell nicht nur probabilistisch, also basierend auf einer Wahrscheinlichkeitsfunktion, sondern auch nichtparametrisch⁴. So ergibt sich ein weiterer Vorteil. Da keine Hypothesen über den Zusammenhang zwischen Items und latenter Variable nötig sind, eignet sich das Verfahren besonders dann, wenn geringe oder gar keine Kenntnisse über die zu untersuchende Dimension vorliegen.

Darüber hinaus können durch ein iteratives Vorgehen – im Gegensatz zum Raschmodell – mehrdimensionale Strukturen entdeckt werden. Aus einem Itempool können somit durchaus mehrere Skalen separiert werden. Ähnlich der Faktorenanalyse bietet die Mokkenskalierung die Möglichkeit der explorativen Skalenanalyse.

Die Skalenwerte der einzelnen Befragten werden denkbar einfach berechnet. Der Gesamtpunktwert der Einzelitems, also die Summe aller mit „Ja“ beantworteten Items, bildet einen Schätzer des Skalenwerts⁵. Das Messniveau des so gewonnenen Index der latenten Dimension weist ordinales Niveau auf, da die Informationen über die Ausprägungen der Personen auf der latenten Dimension nur eine Rangfolge widerspiegeln. Dies ist auf die zugrunde liegende Messabsicht zurückzuführen. Die Summe von Nennungen kann sich zwar prinzipiell als Ratioskala definieren lassen. Da der Abstand zwischen den einzelnen Items, die entsprechend ihrer Schwierigkeit geordnet sind, aber nicht definiert ist und auch keine Annahme über diesen Abstand getroffen werden kann, besitzt der Skalenwert einer Mokkenskala lediglich ordinales Messniveau.

Grundsätzlich lässt sich sagen, dass sich die Mokkenskalierung vor allem zu Beginn der Erforschung eines Phänomens anbietet, wenn noch wenige Informationen über den Gegenstand vorliegen. Dies ist hauptsächlich auf den nichtparametrischen Charakter zurückzuführen. Parametrische Modelle, also solche, die eine bestimmte Annahme vom Verlauf der mathematischen Funktion treffen, implizieren dagegen, wie in Abschnitt 1 erläutert, bereits theoretische Annahmen über den Zusammenhang von Einzelitem und latenter Dimension. Diese Vorgaben sind aber – besonders bei einem kaum erschlossenen Forschungsfeld – in den seltensten Fällen auch zu rechtfertigen.

Wenn wir aber keine Informationen über die Art des Zusammenhangs haben und folglich nicht wissen, ob ein linearer, log-linearer oder sonstiger Zusammenhang vorliegt, wie können wir dann konkret bei der Auswertung vorgehen? Die Antwort liegt in der Ordnungsrelation der Einzelitems. Diese werden hierarchisch angeordnet. Ihre Ordnungsrelation kann als Grundlage dienen, um messtheoretisch auf sicherem Boden zu stehen. Die Zusammenhänge werden auf diese Weise zwar auf ordinale Informationen reduziert, es werden aber auch nicht überprüfbare Voraussetzungen ausgeschlossen.

2.1 Anwendungsvoraussetzungen

Die Mokkenskalierung konzentriert sich im Hinblick auf die formale Gültigkeit auf die Prüfung dieser ordinalen Annahmen.

Es können grundsätzlich zwei Elemente der Skalierung unterschieden werden:

Zunächst müssen die Items dahingehend überprüft werden, ob sie überhaupt den formalen Anforderungen des Verfahrens entsprechen und die Skalierungslösung auch inhaltlich valide erscheint. Bilden die ausgewählten Items keine formal gültige Skala, besteht außerdem die Möglichkeit der Fehleranalyse. Dabei wird untersucht, inwieweit durch ein iteratives Vorgehen ungeeignete Items eliminiert werden können oder sich auch mehr als eine Skala und somit eine mehrdimensionale Struktur finden lässt.

Die Kriterien der Mokken-Skalierung können wie folgt zusammengefasst werden:

„Ein zweiparametrisches⁶ probabilistisches Skalierungsmodell, das spezifisch objektive Messungen ermöglicht und das dem im Rahmen der sozialwissenschaftlichen Analyse weit verbreiteten Problem Rechnung trägt, dass häufig nur geringes Wissen über den Gegenstandsbereich vorhanden ist. Letztere Forderung mündet in ein Verfahren, das notwendigerweise wenige explizite Voraussetzungen beinhaltet, die a priori mit dem Datenmaterial verbunden werden müssen“ (Gerich 2001: 50)

Die Bedingungen des Skalierungsverfahrens fasst Mokken in den Begriffen Homogenität und Holomorphie zusammen. Bei der Voraussetzung der Monotonie werden vier allgemeine Anforderungen der Rasch-Skalierung übernommen (vgl. Gerich 2001: 51):

⁴ „Nichtparametrisch“ bedeutet, dass keine Vorgabe einer bestimmten mathematischen Funktion gemacht wird.

⁵ Dies entspricht dem alternativen Vorgehen bei der Faktorenanalyse, anstatt mit Factorscores, die regressionsanalytisch geschätzt werden müssen, mit Summenscores zu arbeiten.

⁶ zweiparametrisch bezieht sich auf die Verwendung eines Personen- und eines Itemparameters. Das Verfahren ist aber dennoch nichtparametrisch, da die Datenstruktur nicht in Bezug auf eine von einem oder mehreren Parametern abhängige Theorie hin untersucht wird.

- (1) Die Wahrscheinlichkeit einer 1-Anzeige (Bejahung) für jedes Item bei jeder Person ist durch einen Personenparameter (θ) und einen Itemparameter (δ) gegeben.
- (2) Diese Wahrscheinlichkeit entspricht der Funktion $\pi(\theta, \delta)$.
- (3) $\pi(\theta, \delta)$ ist eine monoton steigende Funktion bezüglich (θ).
 → Je größer die Ausprägung einer Person auf der latenten Dimension, desto höher ist die Bejahungswahrscheinlichkeit.
- (4) $\pi(\theta, \delta)$ ist eine monoton fallende Funktion bezüglich (δ).
 → Je schwieriger ein Item, desto geringer ist die Bejahungswahrscheinlichkeit.

Diese Anforderungen vereinen mehrere Komponenten. Sie beinhalten, dass anhand der Menge der ausgewählten Items ein zweiparametrischer Raum aufgespannt werden kann. Dieser wird bestimmt vom Personenparameter, dies ist der Grad der Ausprägung der latenten Dimension bei einer Person, und dem Itemparameter, der die Schwierigkeit eines Items widerspiegelt.

Die Wahrscheinlichkeit, dass ein Item bei einer Person die Ausprägung „1“ annimmt, ist dabei nur abhängig von diesen beiden Parametern. Items, die eindimensionale Indikatoren derselben latenten Variable sind, werden als homogene Items bezeichnet.

Diese Homogenitätsbedingung wird teilweise auch durch den Begriff der lokalen stochastischen Unabhängigkeit ersetzt, was nicht bedeutet, dass die Items statistisch voneinander unabhängig sind. Dies kann schon deshalb nicht zutreffen, weil Korrelationen auf die Messung einer gemeinsamen Dimension hinweisen.

Lokale stochastische Unabhängigkeit bedeutet, „dass beobachtete Koinzidenzen bestimmter Ereignisse (z.B. eine Person antwortet auf zwei verschiedene Fragen gleich) stochastisch unabhängige Realisierungen derselben latenten Variablen sind“ (Denz 1982: 23). Die Antwort von Person A auf ein bestimmtes Item hängt nicht von zuvor gegebenen Antworten ab. Entscheidend sind alleine die Ausprägungen der Person und des Items auf der latenten Dimension. Die lokale stochastische Unabhängigkeit ist somit die statistische Realisation der Homogenitätsbedingung.

Bedingung (3) bedeutet, dass mit einem höheren Wert der latenten Variable θ für ein beliebiges Item i die Wahrscheinlichkeit, von einer Person beantwortet zu werden, steigt.

Fasst man die bisherigen Bedingungen zusammen, so wird die Menge der für die Mokkenskalierung geeigneten Items insofern eingeschränkt, als diese nicht nur homogen, sondern auch monoton sein müssen. Wir gehen also von monoton homogenen Items aus: Eine Person, die eine höhere Ausprägung θ_1 auf der latenten Variable aufweist als eine Person mit der Ausprägung θ_2 , weist also auch generell eine höhere Wahrscheinlichkeit auf, ein bestimmtes Item zu bejahen. Dabei handelt es sich nur um eine *Wahrscheinlichkeit*, nicht um einen Determinismus.

Bedingung (4) wiederum sagt aus, dass ein Item, das sich durch eine größere Schwierigkeit δ_1 auszeichnet, eine geringere Wahrscheinlichkeit der Bejahung besitzt als ein leichteres Item mit der Schwierigkeit δ_2 ($\delta_1 > \delta_2$). Bezüglich der Itemparameter herrscht also auch ein monotonen Verhältnis. Rasch bezeichnet das Zusammentreffen von homogener Monotonie und der Monotonie bezüglich des Itemparameters als Holomorphismus. Mokken übernimmt nun diese Anforderung und erweitert sie. Rasch bezieht seine Bedingung nur auf eine Traceline. Zudem weisen diese in seinem Modell immer den gleichen funktionalen Verlauf auf. Bei der Mokkenskalierung wird diese Voraussetzung derart erweitert, dass der Begriff des Holomorphismus durch ein Äquivalent ersetzt werden kann, das sich zum einen auf ein ganzes Itemset und somit mehrere Tracelines bezieht und zum anderen unterschiedliche funktionale Formen der Tracelines zulässt. Dieses Äquivalent bezeichnet Mokken als **doppelte Monotonie** und ein Itemset, das diese Bedingung erfüllt, als doppelt monotone Items. Diese Bedingung kann auch graphisch geprüft werden. Im Fall doppelter Monotonie dürfen sich die Tracelines der Items unterschiedlicher funktionaler Formen nicht überschneiden (vgl. Abb. 7, S.25). Wir sehen also, dass für die Schätzung des funktionalen Verlaufs der Tracelines der Personenparameter und die Itemschwierigkeit von Bedeutung sind. Beide können sehr einfach aus der jeweiligen Stichprobe geschätzt werden. π_i ist dabei der Wert der Populationswahrscheinlichkeit in der Gesamtheit. Da in aller Regel nur eine Stichprobe vorliegt, erfolgt die Schätzung:

$$\hat{\pi}_i = \frac{n_i}{n} \quad \text{mit} \quad \begin{array}{l} n_i = \text{Anzahl derer, die Item } i \text{ mit „ja“ beantworten} \\ n = \text{Anzahl aller dazu betrachteter Personen} \end{array}$$

$\hat{\pi}_i$ entspricht dem relativen Anteil der Ja-Antworten bei einem Item i in einer Stichprobe und ist folglich der Maximum-Likelihood-Schätzer für π_i , welcher bei hinreichend großem n nur mit geringen Stichprobenfehlern behaftet ist und als **Stichprobenschwierigkeit** der Items bezeichnet wird.

Die Stichprobenschwierigkeit der einzelnen Items wird zunächst in unserem Beispiel betrachtet. Die Items sind in Tabelle 2 so geordnet, dass das schwerste Item, welches von den wenigsten Personen beantwortet wurde, an erster Stelle kommt. Es ist ersichtlich, dass nur etwa 30% der Befragten Bestechung und Korruption als möglichen Weg für einen Aufstieg in unserer Gesellschaft ansehen. Daher ist dieses Item das Schwierigste. Im Gegensatz dazu sind Initiative und Durchsetzungsvermögen für nahezu alle Personen wesentliche Erfolgskriterien. Dieses Item ist leicht zu beantworten. Mit steigendem Wert der latenten Variable nimmt die Itemschwierigkeit somit ab.

Tab. 2: Stichprobenschwierigkeit und Itemparameter

Variablen	Stichproben-schwierigkeit $\hat{\pi}_i$	Schwierigkeitsparameter δ_i
v159a Bestechung, Korruption	0,296	0,704
v151a Politische Betätigung	0,415	0,585
v149a Opportunismus, Rücksichtslosigkeit	0,562	0,438
v158a Herkunft, Richtige Familie	0,638	0,362
v152a Zufall, Glück	0,689	0,311
v156a Geld, Vermögen	0,703	0,297
v160a Kooperation, Offenheit	0,800	0,200
v154a Beziehungen, Protektion	0,904	0,096
v153a Intelligenz	0,953	0,047
v155a Leistung, Fleiss	0,954	0,046
v150a Bildung, Ausbildung	0,975	0,025
v157a Initiative, Durchsetzung	0,978	0,022

Die Stichprobenschwierigkeit $\hat{\pi}_i$ eines Itemsets spiegelt die Schwierigkeitsparameter δ_i beim Vorliegen doppelt monotoner Items in inverser Relation wider:

$$\delta_i = 1 - \hat{\pi}_i \text{ und } \delta_1 > \delta_2 \Rightarrow \pi_1 \pi_j \quad \delta_1 > \delta_2 \Rightarrow \pi_1 \leq \pi_j$$

Das inverse Verhältnis zeigt sich daran, dass ein Item umso leichter ist, je größer π_i ist.

Die Schätzung des Personenparameters erfolgt ähnlich über den Gesamtpunktwert. Für monoton homogene Items gilt, dass für zwei Personen α und β folgende Beziehung besteht:

$$\theta_\alpha \geq \theta_\beta \implies \omega_\alpha \geq \omega_\beta$$

Mit steigendem Wert der latenten Variablen θ steigt auch der Personenparameter, der die „Fähigkeit“ einer Person darstellt. Je höher er ausfällt, desto mehr Items beantwortet diese Person. Dieser Parameter ω kann wiederum geschätzt werden:

$$(\hat{\omega}_\alpha) = \frac{s_\alpha}{k} \quad \begin{array}{l} s_\alpha = \text{Anzahl der positiven Antworten von Person } \alpha \text{ (Score)} \\ k = \text{Anzahl der Items} \end{array}$$

Die Ordnung der Personen auf der latenten Variablen wird also aus der Stichprobe durch den Gesamtpunktwert der beantworteten Items geschätzt. Dieser Gesamtpunktwert repräsentiert den Skalenwert einer Person unabhängig von dessen Zustandekommen, d.h. unabhängig davon, welche Items beantwortet wurden.

Während der Schätzfehler des Itemparameters von n abhängt, wird der Schätzfehler des Personenparameters von der Anzahl der zugrunde liegenden Items bestimmt. Da diese in der Regel deutlich geringer ist, wird der Schätzfehler des Personenparameters größer. Für die praktische Umsetzung bedeutet das, dass eine möglichst umfangreiche Stichprobe an Items operationalisiert werden sollte, die das gesamte Spektrum der latenten Variablen abdeckt.

2.2 Skalierungsprozedur mit dem Variablenset

Das Verfahren der Mokkenskalisierung konzentriert sich nun im Wesentlichen darauf, die Gültigkeit der Modellannahmen z.B. für die vorliegenden Variablen v149a bis v160a zu beurteilen. Da die soeben beschriebene Parameterschätzung nur dann Gültigkeit besitzt, wenn die Bedingungen der doppelten Monotonie erfüllt sind, müssen diese überprüft werden.

Mokken schlägt dazu zwei Konzepte vor: Mit Hilfe des Homogenitätskoeffizienten H kann die Bedingung der monotonen Homogenität getestet werden, auf Basis der P-Matrizen die Homogenität der Itemschwierigkeiten.

Überprüfung der Bedingung der monotonen Homogenität

Zur Beurteilung, ob die Bedingungen erfüllt sind, muss bei Itemvergleichen angesetzt werden. Da wir es hier mit zwei dichotomen Variablen zu tun haben, kann dieser Vergleich auch in Form einer zweidimensionalen Tabelle erfolgen.

Tab. 3: Kontingenztabelle für v151a Politische Betätigung und v159a Bestechung, Korruption (relative Häufigkeiten)

leichtes Item i		schweres Item j		Gesamt
		v159a Bestechung, Korruption 0 nicht genannt	1 genannt	
v151a Politische Betätigung	0 nicht genannt	(0,0) 0,4399	(0,1) 0,1449	$1 - \pi_i$ 0,5848
	1 genannt	(1,0) 0,2641	(1,1) 0,1511	π_i 0,4152
Gesamt		$1 - \pi_j$ 0,7039	π_j 0,2961	1,0000

Anmerkung: In den Tabellen sind die Werte aus Darstellungsgründen auf vier Stellen gerundet. Gerechnet wird in SPSS mit ungerundeten Werten.

In unserem Beispiel ist Item i leichter als Item j, d.h. $\pi_i > \pi_j$. Wichtig für alle weiteren Erläuterungen ist die sogenannte (Guttman-)Fehlerzelle (0,1), die hier markiert ist. Dies sind Fälle, die zwar das leichtere Item i nicht beantworten konnten, dagegen Item j schon. In einer perfekten deterministischen Guttmanskala kämen diese Fälle nicht vor.

Bezüglich der monotonen Homogenität leitet Mokken ab, dass $\pi_{ij} \geq \pi_i \pi_j$ gelten muss. Die Wahrscheinlichkeit einer Bejahung beider Items (Zelle (1,1)) muss folglich größer oder gleich der Wahrscheinlichkeit bei Unabhängigkeit der beiden Ereignisse sein. Im Beispiel hat die Zelle (1,1) den empirischen Wert von 0,1511, bei Unabhängigkeit müsste dieser bei 0,1229 (= 0,2961 x 0,4152) liegen. Die beiden Variablen sind folglich monoton homogen zueinander. Im Falle monotoner Homogenität ist die Korrelation zweier Items damit auch stets größer als Null.

Die Interitemkorrelationen werden bei Mokken über den Homogenitätskoeffizienten H_{ij} gebildet. Dieser Koeffizient konzentriert sich auf die Beurteilung der Fehlerzelle. Generell schwankt er – bei erfüllten Modellbedingungen – zwischen 0 und 1. Er nimmt den Wert 0 an, wenn die Antwortwahrscheinlichkeit der Fehlerzelle der Wahrscheinlichkeit bei statistischer Unabhängigkeit entspricht. Der Wert 1 liegt dann vor, wenn die Items der Beziehung einer perfekten Guttmanskala entsprechen, die Fehlerzelle also nicht besetzt ist. Ein negativer Wert kann dann auftreten, wenn der tatsächliche Fehler höher ist als der Fehler, der bei Unabhängigkeit zu erwarten ist. Dies bedeutet aber eine Verletzung der Skalierungsvoraussetzungen, genauer, der monotonen Homogenität.

Was bedeutet das konkret für unser Beispiel? Für die Berechnung des Homogenitätskoeffizienten benötigen wir zwei Maßzahlen: Da hier Item j schwieriger ist als Item i, entspricht – wie bereits beschrieben – die Zelle (0,1) der Fehlerzelle. Die Wahrscheinlichkeit entspricht der relativen Antwortwahrscheinlichkeit dieser Zelle:

$$\pi_{eij} = \pi_{ij}(0,1) = 0,1449.$$

Die erwartete Wahrscheinlichkeit bei statistischer Unabhängigkeit wiederum ist wie folgt definiert:

$$\pi_{0ij} = \pi_j(1 - \pi_i) = 0,1731.$$

Der Homogenitätskoeffizient H_{ij} lässt sich daraus folgendermaßen berechnen:

$$H_{ij} = \frac{\pi_{0ij} - \pi_{eij}}{\pi_{0ij}}$$

Für die zwei Variablen v159a und v151a ergibt sich $H_{ij} = \frac{0,1731 - 0,1449}{0,1731} = \mathbf{0,1628}$.

Diese zwei Items weisen demnach nur eine schwache Homogenität auf, jedoch keine Verletzung der monotonen Homogenität. Werden für alle weiteren Variablen die Homogenitätskoeffizienten berechnet, kann visuell anhand einer Korrelationstabelle geprüft werden, ob die Items überhaupt für die Skalierung geeignet sind.

Tab. 4: Übersicht über die paarweisen Homogenitätskoeffizienten der Variablen v149a bis v160a

	Paarweise Homogenitäts-Koeffizienten H_{ij}												
	v149a	v150a	v151a	v152a	v153a	v154a	v155a	v156a	v157a	v158a	v159a	v160a	
v149a	1,0000												
v150a	-,3739	1,0000											
v151a	,1066	-,0583	1,0000										
v152a	,0571	-,0997	,0121	1,0000									
v153a	-,4136	,3481	-,1845	-,0590	1,0000								
v154a	,5448	,0443	,4894	,2278	-,0069	1,0000							
v155a	-,6150	,3804	-,3448	-,0645	,3004	-,0790	1,0000						
v156a	,2367	-,1427	,2877	,0492	-,0847	,3853	-,2569	1,0000					
v157a	,0019	,0999	,0704	,0068	,1163	,1069	,1170	,1263	1,0000				
v158a	,2245	-,2594	,2185	,0769	-,1055	,4981	-,3461	,3430	-,1006	1,0000			
v159a	,5745	-1,2006	,1628	,0960	-,7442	,7973	-1,3362	,5181	-,1259	,4970	1,0000		
v160a	-,2924	,4125	-,0236	,0788	,3849	-,0606	,4268	-,1192	,4734	-,1913	-,6303	1,0000	

In Tabelle 4 finden sich einige deutlich positiv von Null verschiedene Korrelationen. Als das homogenste Itempaar können die Variablen v154a Beziehungen / Protektion und v159a Bestechung /Korruption bezeichnet werden. Allerdings bestehen viele negative Korrelationen. Dies stellt eine Verletzung der Monotoniebedingung der Mokkenskalierung dar. In eine Skala dürfen nur Items aufgenommen werden, die untereinander einen positiven Zusammenhang aufweisen. Folglich kann beispielsweise die Variable v149a Opportunismus / Rücksichtslosigkeit nicht gemeinsam mit den Variablen v150a Bildung / Ausbildung, v153a Intelligenz, v155a Leistung / Fleiß und v160a Kooperation / Offenheit skaliert werden. Es liegt die Vermutung nahe, dass der Itempool mehr als nur eine Dimension misst.

Die Homogenitätskoeffizienten H_{ij} allein sind jedoch nicht ausreichend für die Skalenkonstruktion. Sobald ein homogenes Itempaar gefunden wurde und weitere Items dazu genommen werden sollen, braucht man einen allgemeinen Koeffizienten für die Skalenhomogenität (H). Dieser ist zu vergleichen mit Cronbachs Alpha, das aus der Reliabilitätsanalyse der klassischen Testtheorie bekannt ist. H_{ij} und H entsprechen den Koeffizienten H_i und H_j von Loevinger. Daher wird der Homogenitätskoeffizient im Rahmen der Mokkenskalierung als Loevingers H bezeichnet. Auch H basiert auf dem Verhältnis von empirischer Besetzung der Fehlerzelle und theoretischer Besetzung bei Unabhängigkeit. Für mehrere Items lässt sich daher die obige Formel verallgemeinern und zwei Größen bestimmen:

$$E = \sum_{j=1}^{k-1} \sum_{i=j+1}^k \pi_{eij}$$

Um E zu erhalten, werden alle möglichen Itempaare kreuztabelliert (s. Tab. 3) und die Fehlerzellen all dieser Tabellen aufaddiert⁷. Insgesamt sind dies bei k Items $\frac{k^2 - k}{2}$ Kreuztabellen.

Genauso verfährt man für alle theoretischen Fehlerzellen bei Unabhängigkeit:

$$E_0 = \sum_{j=1}^{k-1} \sum_{i=j+1}^k \pi_{0ij}$$

⁷ Dabei muss immer darauf geachtet werden, welches Item das leichtere und welches das schwerere ist, um die Fehlerzelle richtig zu bestimmen.

Für die Definition des Homogenitätskoeffizienten ergibt sich somit:

$$H = 1 - \frac{E}{E_0} = \frac{E_0 - E}{E_0} = \frac{\sum_{j=1}^{k-1} \sum_{i=j+1}^k \pi_{0ij} - \sum_{j=1}^{k-1} \sum_{i=j+1}^k \pi_{eij}}{\sum_{j=1}^{k-1} \sum_{i=j+1}^k \pi_{0ij}}$$

Wir sehen hier, dass die Homogenität einer Skala umso höher ist, je kleiner E ist und damit die Guttmanfehler. Dies entspricht unserer Intention, die (monotone) Homogenität abzubilden.

Grundlage für H sind die Informationen der einzelnen Paarvergleiche. H gibt Aufschluss über eine Menge an Items und setzt sich aus den „Homogenitätsbeiträgen“ der Paarvergleiche zusammen. Es ist daher von zentraler Bedeutung, diejenigen Paare zu identifizieren, die die Homogenität eher verringern oder erhöhen.

Mokken entwickelte noch einen weiteren Koeffizienten H_i , der einen Item-Homogenitätskoeffizienten darstellt (vergleichbar mit dem Trennschärfekoeffizienten) und wie folgt definiert ist:

$$H_i = 1 - \frac{E_i}{E_{0i}}$$

H_i stellt ein Maß der Homogenität eines Items im Vergleich zu den übrigen Items dar. Es beinhaltet im Grunde die „summierte“ Homogenität eines Items i aus allen Paarvergleichen, in denen i enthalten ist. E_i stellt dabei den Fehler dar, der sich aus der Summe aller Fehlerwahrscheinlichkeiten eines Items i mit den übrigen Items ergibt. E_{0i} ist der entsprechende Fehler bei statistischer Unabhängigkeit.

Die Entwicklung des Itemkoeffizienten H_i ist für die Mokkenskalierung deshalb so wichtig, weil dadurch beurteilt werden kann, ob und in welcher Weise einzelne Items hinsichtlich des gesamten Itemsets Homogenität aufweisen. Er hilft also besonders bei der Fehleranalyse und ermöglicht die Eliminierung ungeeigneter Items.

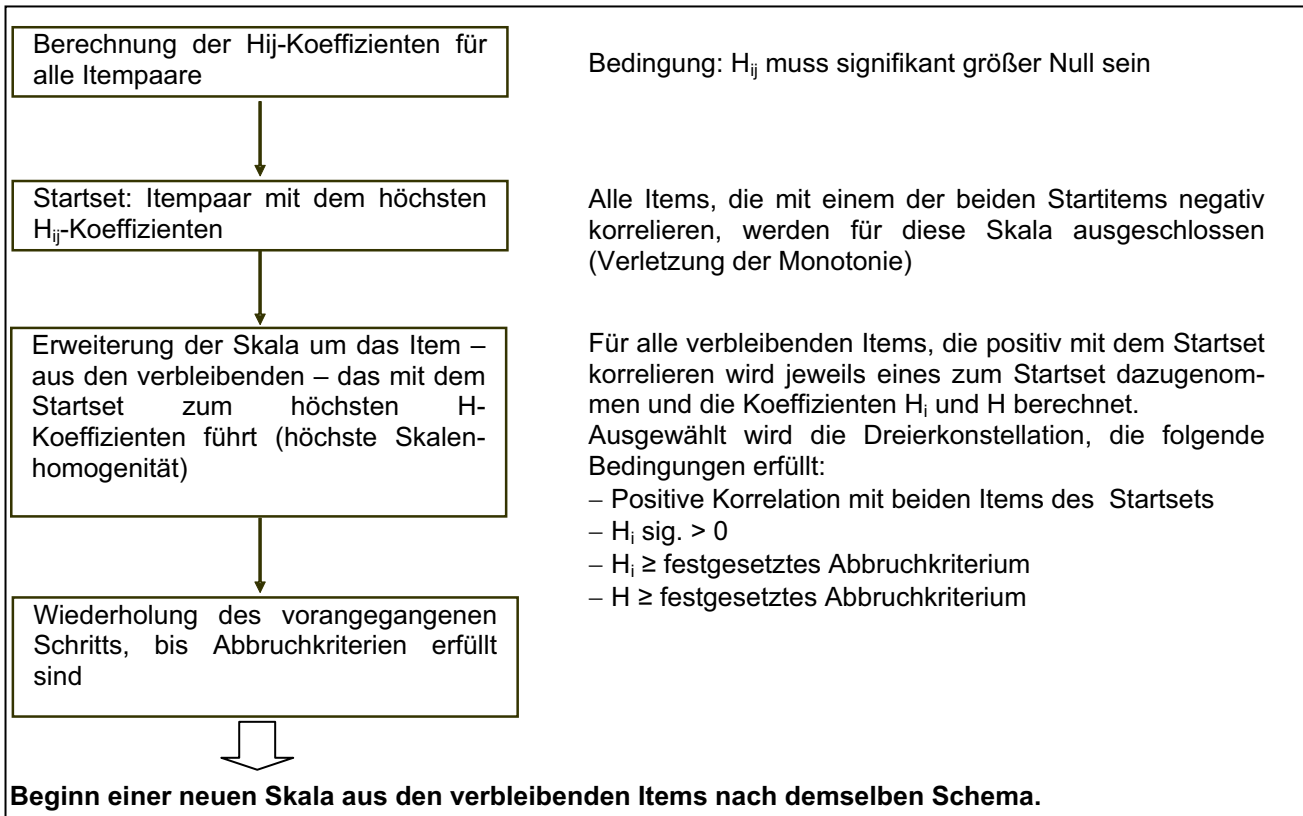
Wir haben bisher gesehen, dass H die Homogenität eines Itemsets und damit in gewisser Weise auch die Qualität dieses Sets misst. Doch wie beurteilen wir nun die Qualität einer Skala?

Es gibt keine festen Regeln, ab wann man von einer homogenen und wirklich brauchbaren Lösung sprechen kann. Mokken schlägt dazu einige anwendungsorientierte Faustregeln vor:

- starke Skala: Homogenitätskoeffizient H größer/ gleich 0,5
- mittlere Skala: H zwischen 0,4 und 0,5
- schwache Skala: H zwischen 0,3 und 0,4

Da die Itemvergleiche Grundlage des Modells sind, ist zudem von der Skala zu erwarten, dass die einzelnen Items auch eine gewisse Homogenitätsgrenze nicht unterschreiten. Mokken schlägt in diesem Zusammenhang vor, Itemsets nur dann als Skala zu definieren, wenn alle Itemkoeffizienten H_i Werte größer oder gleich einer festgelegten Konstanten c aufweisen. In der Praxis hat sich dabei $c=0,3$ als praktikabel erwiesen.

Für eine Skalierungsprozedur sind alle wichtigen Maßzahlen beschrieben worden. Nachdem alle für eine Skalierungsprozedur wichtigen Maßzahlen beschrieben wurden, soll nun die praktische Umsetzung anhand unseres Beispiels erfolgen.

Abb. 6: Ablaufschema der Skalenkonstruktion nach Mokken

Quelle: Eigene Zusammenstellung nach Gerich 2001: 209f.

Startset finden: In Tabelle 4 haben wir bereits das homogenste Itempaar gefunden: v154a Beziehungen / Protektion und v159a Bestechung / Korruption weisen eine Korrelation von 0,7973 auf. Die Korrelation weicht signifikant positiv von Null ab. Befinden sich nur zwei Items in der Skala, so sind H_{ij} und H identisch (H_i ist also erst ab dem dritten Item sinnvoll). Nachdem der Skalenkern bestimmt ist, können alle Items ausgeschlossen werden, die mit einer der beiden Variablen negativ korrelieren. Dies sind die Variablen:

Tab. 5: Für Skala 1 ausgeschlossene Variablen

v150a	Bildung, Ausbildung
v153a	Intelligenz
v155a	Leistung, Fleiß
v157a	Initiative, Durchsetzung
v160a	Kooperation, Offenheit

Offensichtlich repräsentieren diese Items nicht dieselbe Hintergrundvariable wie die Items des Startsets. Doch welches der verbleibenden Items passt zu diesem Paar?

Erweiterung der Skala: Zur Erweiterung der Skala stehen nur noch die folgenden Variablen zur Verfügung:

Tab. 6: Mögliche Variablen zur Skalenbildung für Skala 1

v149a	Opportunismus, Rücksichtslosigkeit
v151a	Politische Betätigung
v152a	Zufall, Glück
v156a	Geld, Vermögen
v158a	Herkunft, Richtige Familie

Jedes dieser Items wird einzeln zum Startset dazu genommen und für jede Kombination werden die Koeffizienten H_i und H nach obigen Formeln berechnet. Zur besseren Nachvollziehbarkeit der Berechnungen sind

in Tab. 7 alle Fehlerzellen aus den gebildeten Kreuztabellen eingetragen und in Tab. 8 die korrespondierenden Werte bei Unabhängigkeit. Zuerst wird z.B. die Skalenhomogenität H überprüft, wenn zu v154a und v159a die Variable v149a hinzukommt. Die vereinfachte Formel für H ist: $H = \frac{E_0 - E}{E_0}$. Um E zu bestimmen,

müssen wir die Werte für die drei Variablen aus Tab. 7 herausuchen und aufaddieren (s. Markierung).

$$E(v154a, v159a, v149a) = 0,0247 + 0,0551 + 0,0058 = \mathbf{0,0856}$$

Tab. 7: Empirische Fehlerzellenbesetzung

	v149a	v151a	v152a	v154a	v156a	v158a	v159a
v149a	,0000						
v151a	,1623	,0000					
v152a	,1650	,1276	,0000				
v154a	,0247	,0204	,0513	,0000			
v156a	,1276	,0879	,1947	,0416	,0000		
v158a	,1581	,1176	,1831	,0308	,1245	,0000	
v159a	,0551	,1449	,0833	,0058	,0424	,0540	,0000

Für E_0 wird mit Tab. 8 analog verfahren:

$$E_0(v154a, v159a, v149a) = 0,0542 + 0,1295 + 0,0285 = \mathbf{0,2122}$$

$$H(v154a, v159a, v149a) = \frac{0,2122 - 0,0856}{0,2122} = \mathbf{0,5966}$$

Tab. 8 Fehlerzellenbesetzung bei Unabhängigkeit

	v149a	v151a	v152a	v154a	v156a	v158a	v159a
v149a	,0000						
v151a	,1817	,0000					
v152a	,1750	,1292	,0000				
v154a	,0542	,0400	,0664	,0000			
v156a	,1672	,1234	,2048	,0677	,0000		
v158a	,2038	,1505	,1984	,0615	,1895	,0000	
v159a	,1295	,1731	,0921	,0285	,0880	,1073	,0000

Ebenso einfach lassen sich die H_i - Koeffizienten für diese Konstellation berechnen. Für v149 müssen Kreuztabellen mit den beiden anderen Variablen gebildet und wiederum die Fehlerzellen addiert werden.

$$H_i = 1 - \frac{E_i}{E_{0i}} \quad H_i(v149) = 1 - \frac{0,0247 + 0,0551}{0,0542 + 0,1295} = \mathbf{0,5657}$$

Alle möglichen Kombinationen für H und H_i werden im Mokkenskalisierungsprogramm durchgerechnet. Auf dieser Basis wird diejenige Lösung ausgewählt, die den H Koeffizienten von einem Anfangswert 0,7973 am wenigsten vermindert. Gleichzeitig sollte der H_i -Koeffizient für das betreffende Item nicht unter einem festgesetzten Wert liegen. In unserem Beispiel haben wir bereits das beste Item gefunden. Alle anderen verbleibenden Items hätten schlechtere Werte erzielt.

Diese Prozedur stellt sicher, dass auch wirklich die homogensten Items in eine Skala aufgenommen werden. Tabelle 9 zeigt die vollständige Skalierung für ein Abbruchkriterium von 0,3 sowohl für H als auch für H_i .

Tab. 9: Skala 1

	H _i und H für c=0,3 und c _i =0,3		
	H _i während Skalierung	(Loevingers)H	H _i nach Skalierung
v159a Bestechung, Korruption			,5549
v154a Beziehungen, Protektion	,7973	,7973	,5143
v149a Opportunismus, Rücksichtslosigkeit	,5657	,5969	,3412
v156a Geld, Vermögen	,3446	,4446	,3440
v158a Herkunft, Richtige Familie	,3464	,3943	,3464

Es können keine weiteren Variablen aufgenommen werden.

H_i wird dabei zweimal berechnet. Während der Skalierung sind noch nicht alle Items aufgenommen, weshalb sich die H_i-Werte nur auf die Items beziehen können, die schon vorhanden sind. Ist die Skala fertiggestellt, werden alle H_i-Werte noch einmal für die komplette Skala berechnet. Es zeigt sich beispielsweise für v149a in der Dreierkonstellation mit v159a und v154a ein H_i von 0,5657. Inhaltlich passen diese Items sehr gut zusammen. Sind dagegen alle Variablen aufgenommen, wird dieser Wert deutlich schlechter und auch inhaltlich verändert sich der Fokus etwas. An dieser Stelle bleibt es dem Forscher überlassen, welche Lösung er wählt. Er muss zwischen inhaltlichen und statistischen Kriterien abwägen.

Nach dem gleichen Schema lassen sich aus unserer Itemliste zwei weitere Skalen bilden.

Tab. 10: Skala 2

	H _i und H für c=0,3 und c _i =0,3		
	H _i während Skalierung	(Loevingers)H	H _i nach Skalierung
v160a Kooperation, Offenheit			,4418
v157a Initiative, Durchsetzung	,4734	,4734	,2795
v155a Leistung, Fleiß	,3147	,3516	,3147

Es können keine weiteren Variablen aufgenommen werden.

Obwohl v157a zusammen mit v160a das Startset bildet, fällt der H_i-Wert nach Erweiterung um die Variable v155a unter die Grenze von 0,3. Auch hier muss entschieden werden, ob die Skala so bestehen bleiben oder verworfen bzw. modifiziert werden soll. Die gemeinsame Hintergrunddimension könnte beispielsweise für spezielle Charaktereigenschaften bzw. zielgerichtetes Verhalten auf dem Weg zum Erfolg sprechen.

Tab. 11: Skala 3

	H _i und H für c=0,3 und c _i =0,3		
	H _i während Skalierung	(Loevingers)H	H _i nach Skalierung
v153a Intelligenz			,3481
v150a Bildung, Ausbildung	,3481	,3481	,3481

Es können keine weiteren Variablen aufgenommen werden.

Für die letzte Skala, die im weitesten Sinne Bildung umfasst, können nur noch zwei Variablen verwendet werden. Die Variablen v151a „Politische Betätigung“ und v152a „Zufall, Glück“ passen in keine der drei Skalen.

Überprüfung der Bedingung der Monotonie der Itemschwierigkeit: P-Matrizen, Empirische Tracelines

Mit der Bildung der Skalen auf der Basis des Homogenitätskoeffizienten ist die Prüfung des Modells noch nicht abgeschlossen. Zur endgültigen Bestimmung der Skalierbarkeit muss die Bedingung der Monotonie der Itemschwierigkeit überprüft werden. Dies geschieht über die Betrachtung der P-Matrizen. Im Anschluss kann dann die doppelte Monotonie mit Hilfe der empirischen Tracelines bewertet werden.

Wenn diese Bedingungen erfüllt sind, liegen holomorphe Itemsets vor. Mokken leitet für solche holomorphe Itemsets, die eine Skala bilden, ein weiteres zentrales Theorem ab, das sich auf den Vergleich von Itemtriaten bezieht und in der Form von vier Ungleichungen formuliert ist.

Nimmt man eine Menge von k holomorphen Items an, so muss für alle Items i, j_1, j_2 aus dieser Menge von k Items, wobei j_1 schwieriger ist als j_2 , und beliebigem i aus k Items gelten:

- (1) $\pi_{ij_1}(1,1) \leq \pi_{ij_2}(1,1)$
- (2) $\pi_{ij_1}(1,0) \geq \pi_{ij_2}(1,0)$
- (3) $\pi_{ij_1}(0,1) \leq \pi_{ij_2}(0,1)$
- (4) $\pi_{ij_1}(0,0) \geq \pi_{ij_2}(0,0)$

Was bedeuten nun diese Ungleichungen?

Ungleichung (1) zeigt, dass die gleichzeitige Bejahung eines beliebigen Items i aus einem holomorphen Itemset mit einem schwereren Item (j_1) weniger häufig zutrifft als die gemeinsame Bejahung mit einem leichteren Item (j_2).

Ungleichung (4) betrifft die gemeinsame 0-Anzeige der Items. Die gemeinsame Verneinung eines beliebigen Items i mit einem schwereren Item j_1 ist größer oder gleich der gemeinsamen Verneinung des Items i mit einem leichteren Item j_2 .

(1) und (4) gelten auch im allgemeineren Fall der doppelten Monotonie, d.h., wenn sich die Tracelines nicht überschneiden.

Ungleichung (2) wiederum zeigt Folgendes: Die Wahrscheinlichkeit einer 1-Anzeige bei Item i unter der Bedingung, dass Item j_1 die Ausprägung 0 aufweist, ist größer oder gleich der Wahrscheinlichkeit, dass Item i die Ausprägung 1 besitzt, wenn ein leichteres Item die Ausprägung 0 aufweist.

Ungleichung (3) stellt die Umkehrung zu Ungleichung (2) dar.

Auf Basis dieser vier Ungleichungen, die notwendige Bedingungen für die Monotonie der Itemschwierigkeiten sind, schlägt Mokken zur Prüfung dieser Bedingung die Bildung und Inspektion der P-Matrizen vor. Dazu werden zwei Matrizen, P und P_0 , gebildet. Die symmetrische P-Matrix enthält dabei als Einträge die Anteilswerte der 1-Anzeigen jeweils zweier Items. Die P_0 -Matrix enthält die entsprechenden Anteile der gleichzeitigen 0-Anzeigen zweier Items. Sowohl in der P- als auch in der P_0 -Matrix werden die Zeilen und Spalten nach aufsteigender Schwierigkeit geordnet.

Tab. 12: P-Matrix für Skala 1 (Anteile der 1-1 Kombination)

	v154a	v156a	v158a	v149a	v159a
v154a Beziehungen, Protektion	—	0,6611	0,6068	0,5378	0,2903
v156a Geld, Vermögen	0,6611	—	0,5131	0,4348	0,2537
v158a Herkunft, Richtige Familie	0,6068	0,5131	—	0,4044	0,2421
v149a Opportunismus, Rücksichtslosigkeit	0,5378	0,4348	0,4044	—	0,2409
v159a Bestechung, Korruption	0,2903	0,2537	0,2421	0,2409	—

Tab. 13: P_0 -Matrix für Skala 1 (Anteile der 0-0 Kombination)

	v154a	v156a	v158a	v149a	v159a
v154a Beziehungen, Protektion	—	0,0547	0,0655	0,0717	0,0906
v156a Geld, Vermögen	0,0547	—	0,1727	0,1696	0,2548
v158a Herkunft, Richtige Familie	0,0655	0,1727	—	0,2043	0,3084
v149a Opportunismus, Rücksichtslosigkeit	0,0717	0,1696	0,2043	—	0,3824
v159a Bestechung, Korruption	0,0906	0,2548	0,3084	0,3824	—

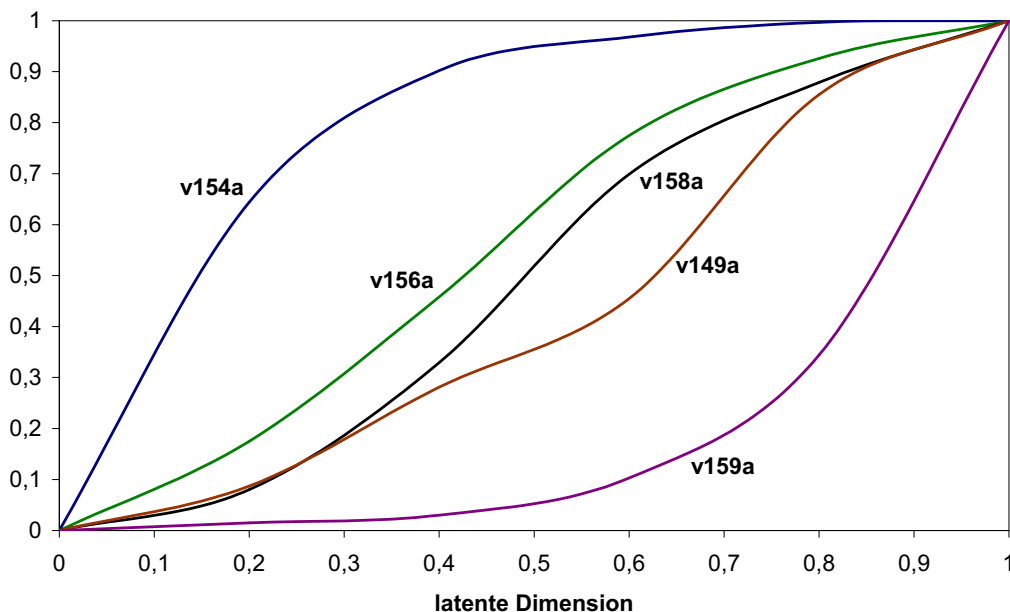
Wenn nun die zugrunde liegenden Theoreme gelten, muss die P-Matrix die Eigenschaft aufweisen, dass die Anteilswerte von oben nach unten und von links nach rechts sinkende oder zumindest gleiche Werte aufweisen. Für die P_0 -Matrix gilt die gegenteilige Eigenschaft: Die Anteilswerte müssen von oben nach unten und von links nach rechts ansteigen.

Bezüglich der Abweichungen von diesem Muster werden verschiedene Verfahren vorgeschlagen. Mokken regt an, die P-Matrizen visuell zu prüfen und Abweichungen, die sich innerhalb einer tolerierbaren Differenz von 0,03 bewegen, zu ignorieren. Andere Autoren tendieren dagegen dazu, auftretende Abweichungen mittels statistischer Tests auf Signifikanz hin zu überprüfen (Bacher 1990: 55).

Die Tabellen 12 und 13 enthalten beispielhaft die Matrizen für Skala 1. Es zeigt sich bei der P_0 -Matrix eine Abweichung von den theoretischen Anforderungen bei den Variablen v149a und v156a. Da diese aber unter dem von Mokken vorgeschlagenen Toleranzwert von 0,03 liegt, erhalten wir die Skala auch weiterhin aufrecht.

Die Prüfung der doppelten Monotonie anhand der Sichtprüfung der empirischen Tracelines schließt die eigentliche Skalierung ab. Bei der Mokkenskalierung wird keine Annahme über den funktionalen Verlauf der Tracelines getroffen. Da wir folglich auch nicht wissen, welche Form die Tracelines annehmen, erfolgt die Prüfung der Tracelines anhand der empirischen Daten. Über das „wahre“ Aussehen der Tracelines kann i.d.R. erst nach der wiederholten Prüfung (mehrere Studien) von Skala und Items entschieden werden. Abbildung 7 zeigt den Verlauf der empirischen Tracelines für Skala 1.

Abb.7: Empirische Tracelines für Skala1



Das von uns im Rahmen der Prüfung der P-Matrizen bereits analysierte Homogenitätsproblem lässt sich auch hier erkennen. Es hat zur Folge, dass sich die Tracelines der Variablen v149a und v158a überschneiden. Diese Inkonsistenzen sind aber vernachlässigbar. Die Tracelines zeigen grundsätzlich eine doppelmonotone Struktur, die eine Mokkenskalierbarkeit der Items nahe legt.

2.3 Abschluss des Skalierungsprozesses: Bildung der Skalenwerte am Beispiel

Alle formalen Kriterien sprechen für die gefundene dimensionale Struktur des Variablensets. Unabhängig von inhaltlichen und statistischen Optimierungsmöglichkeiten, die nicht Thema dieser Einführung sein sollen, endet die Skalierungsprozedur mit der Bildung der Skalenwerte. Dadurch lässt sich einerseits die Verteilung der Dimension in der Stichprobe darstellen und andererseits stehen die Skalen für weitergehende Analysen wie z.B. kausalanalytische Fragestellungen zur Verfügung.

Berechnung der Skalenwerte in SPSS

Die Gesamtpunktwerte für die einzelnen Skalen werden aus der Stichprobe mit der bereits vorgestellten Formel geschätzt:

$$(\hat{\omega}_\alpha) = \frac{s_\alpha}{k} \quad \begin{array}{l} s_\alpha = \text{Anzahl der positiven Antworten von Person } \alpha \text{ (Score)} \\ k = \text{Anzahl der Items} \end{array}$$

Für jede Person werden alle bejahten Items der jeweiligen Skala aufaddiert und durch die Gesamtzahl aller Items der Skala dividiert. Damit sind alle Skalen zwischen 0 und 1 normiert. In SPSS wird dies mit einem COMPUTE-Befehl umgesetzt (s.Abb.8).

Abb.8: Syntaxfenster mit Berechnung der Skalenwerte

```

\\*Mokkenskalisierung mit den Variablen v149a bis v160a - ALLBUS 2004*\\
\\*Berechnung der Skalenwerte*\\

COMPUTE skala1 = (v159a + v154a + v149a + v156a + v158a)/5.
EXECUTE.
COMPUTE skala2 = (v160a + v157a + v155a)/3.
EXECUTE.
COMPUTE skala3 = (v153a + v150a)/2.
EXECUTE.

FORMATS skala1 skala2 skala3 (F5.4).

FREQUENCIES skala1 skala2 skala3.

```

Häufigkeitsverteilungen der Skalen

Aus den Tabellen 14 bis 16 können die einzelnen Häufigkeitsverteilungen der drei Skalen verglichen werden.

Tab.14: Häufigkeitsverteilung für Skala 1

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ,0000	80	3,1	3,1	3,1
,2000	275	10,6	10,6	13,7
,4000	502	19,4	19,4	33,0
,6000	648	25,0	25,0	58,0
,8000	620	23,9	23,9	81,9
1,0000	469	18,1	18,1	100,0
Gesamt	2594	100,0	100,0	

Dimension 1 stellte eine Mischung aus illegitimen Mitteln auf dem Weg zum Erfolg und persönlichem Hintergrund mit Geld und Vermögen und der richtigen Herkunft dar. Insgesamt zeigt sich eine relativ große Streuung über die einzelnen Ausprägungsgrade der Dimension. Immerhin 18% der Befragten halten diese Dimension für unerlässlich um erfolgreich zu sein, da sie alle Items der Skala bejaht haben.

Wesentlich ungleicher sind die Verteilungen der Dimensionen 2 und 3: 77% bzw. rund 94% der Befragten haben hier maximale Ausprägungen (vgl. Tab.15 und Tab.16) und sehen somit zielgerichtetes persönliches Verhalten im Leben und Bildung als zentrale Elemente für einen gesellschaftlichen Aufstieg.

Tab.15: Häufigkeitsverteilung für Skala 2

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ,0000	5	,2	,2	,2
,3333	92	3,5	3,5	3,7
,6667	498	19,2	19,2	22,9
1,0000	1999	77,1	77,1	100,0
Gesamt	2594	100,0	100,0	

Tab.16: Häufigkeitsverteilung für Skala 3

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ,0000	25	1,0	1,0	1,0
,5000	138	5,3	5,3	6,3
1,0000	2431	93,7	93,7	100,0
Gesamt	2594	100,0	100,0	

Aufgrund der großen prozentualen Unterschiede der höchsten Ausprägung zwischen den drei Skalen lässt sich vermuten, dass die Befragten den drei Dimensionen unterschiedliche Wichtigkeit für ein erfolgreiches Leben beimessen. An erster Stelle steht die Bildungsdimension (Skala 3), die sich nicht nur auf erworbene Qualifikation beschränkt, sondern auch mitgebrachte Fähigkeiten (Item „Intelligenz“) berücksichtigt. Ein weiterer wichtiger Bereich ist das zielstrebige Verhalten (Skala 2) und weitaus weniger bedeutend sind unstatthafte Wege (Skala 1).

Literatur:

Bacher, Johann (1990): Einführung in die Logik der Skalierungsverfahren. In: Historical Social Review 15, S.1-170

Denz, Hermann (1982): Analyse latenter Strukturen. München; Franke

Gerich, Joachim (2001): Nichtparametrische Skalierung nach Mokken. Linz; Trauner Verlag

Meyer, Harald (2004) Theorie und Qualitätsbeurteilung psychometrischer Tests. Stuttgart: Verlag W. Kohlhammer.