

### Zum Einfluß verbal-numerischer Antwortmodelle sowie der Frage-/ Aussageformulierung auf das Befragungsergebnis bei Meinungsindikatoren: (institutsinterner Forschungsbericht zur Methoden- Studie FB-72)

Ettrich, Klaus Udo; Schreiber, Dieter

Forschungsbericht / research report

#### Empfohlene Zitierung / Suggested Citation:

Ettrich, K. U., & Schreiber, D. (1973). *Zum Einfluß verbal-numerischer Antwortmodelle sowie der Frage-/ Aussageformulierung auf das Befragungsergebnis bei Meinungsindikatoren: (institutsinterner Forschungsbericht zur Methoden-Studie FB-72)*. Leipzig: Zentralinstitut für Jugendforschung (ZIJ). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-370476>

#### 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.

## ZENTRALINSTITUT FÜR JUGENDFORSCHUNG

## Sektor Methodik

Vertrauliche Dienststudie				
Nachweis-Bereich	Umfang	Jahr	Ausführung	Blatt
217	26	73	19	1-84

00186

gelöscht  
25.1.1984 Jm

Zum Einfluß verbal-numerischer Antwortmodelle sowie der Frage-/Ausageformulierung auf das Befragungsergebnis bei Meinungsindikatoren.

(Institutsinterner Forschungsbericht zur Methoden-Studie PB-72)

K.U. Eitrich / D. Schreiber

Leipzig, Juli 1973

<u>Gliederung</u>	<u>Seite</u>
1. Ziel- und Problemstellung	1
2. Untersuchungsmethodik	12
2.1. Fragebogen zur Analyse der Antwortmodelle	12
2.2. Fragebogen zur Analyse der Frage-/Aussage- wirkung	16
3. Zur Interviewsituation und Analysestichprobe	18
4. Untersuchungsergebnisse	21
4.1. Zum Einfluß verbal-numerischer Antwortmodelle auf das Befragungsergebnis bei Meinungsindika- toren	21
4.1.1. Kontrolle der Abhängigkeit des Befragungser- gebnisses von der Reihenfolge der Antwort- modelle	21
4.1.2. Stabilität und Trennschärfe der Antwortmodelle	27
4.1.3. Zur Transformation der Antwortmodelle	34
4.1.4. Trennscharfenanalyse einzelner Stufen der AI und deren Zusammenfassungen	51
4.1.5. Abhängigkeit der Indikatorfunktion vom Ant- wortmodell	57
4.1.6. Quantitative Äquivalenz verbaler Urteile	59
4.2. Zum Einfluß der Frage-/Aussageformulierung auf das Befragungsergebnis bei Meinungs- indikatoren	65
4.2.1. Zur Abhängigkeit des Befragungsergebnisses von der Frage-/Aussageformulierung	65
4.2.2. Stabilität und Trennschärfe von Frage- und Aussageitems	69
4.2.3. Zur Transformation der Befragungsergebnisse bei Frage- und Aussageformulierung	72
5. Zusammenfassung und Schlussfolgerungen	76
5.1. Schlussfolgerungen zur Antwortmodell- problematik	76
5.2. Schlussfolgerungen zur Frage-/Aussageformu- lierung	80
Literatur	82

## 1. Ziel- und Problemstellung

Die kollektive Diskussion von Umfrageergebnissen in unserem Institut läßt deutlich werden, daß in die Interpretation der Befragungsergebnisse ein gewisser Prozentsatz "unaufgeklärter Varianz" eingeht, der daraus resultiert, daß bestimmte Probleme der Befragungsmethodik bisher ungenügend analysiert werden konnten. Insbesondere kristallisieren sich in der Diskussion zwei Probleme heraus:

1. Inwiefern können die Ergebnisse zu gleichen Befragungszeitpunkten verglichen werden, wenn ihnen unterschiedliche Antwortmodelle zugeordnet sind?

Beispielsweise bevorzugt der Sektor EOP vierstufige, der Sektor Studentenforschung sechsstufige Antwortmodelle. Die Ursachen für diese Entwicklung sind bekannt und stehen, wohl sekundär für unser Anliegen, hier nicht zur Diskussion. Aus der unterschiedlichen Stufenzahl der Antwortmodelle ergeben sich für umfassendere Verallgemeinerungen (also Freigeb, die über die Auswertung einer aktuellen Studie hinausgehen) sofort eine Reihe Probleme, von denen einige genannt sein sollen:

a) Wie müssen die Stufen der Antwortmodelle zusammengefaßt werden, damit vergleichbare Ergebnisse entstehen?

b) Bedeutet z.B. die Stufe "mit gewissen Einschränkungen meine Meinung" in einem vierstufigen Antwortmodell das gleiche wie in einem sechsstufigen Antwortmodell? Wenn nein, wie ist dann der quantitative Charakter dieser Stufe in unterschiedlichen Antwortmodellen zu verstehen und zu werten?

c) Wie ändert sich das Befragungsergebnis, wenn man von einem niederstufigen Antwortmodell in ein höherstufiges übergeht? Ebenso muß man natürlich fragen, was bei der Umkehrung des Prozesses geschieht.

d) Ändert sich durch die Verwendung eines anderenstufigen Antwortmodells etwa die Indikatorfunktion des Items? Eine sehr wichtige Frage, auf die wir gegebenenfalls keine Antwort vorfinden.

Wir brechen hier die Reihe relevanter Fragen vorerst ab und wenden uns dem zweiten Diskussionsaspekt zu.

## 2. Welchen Einfluß hat die Frageformulierung bzw. die Aussageformulierung eines Items auf das Befragungsergebnis?

Über die Häufigkeitsverteilung der Befragungsergebnisse bei Aussage- oder Frageformulierung gibt es einige Voraussetzungen, die sich auf die psychologische Wirkung von Frage und Aussage als Stimulus verbales Antwortverhalten beziehen.

So wird allgemein angenommen, daß Frage und Aussage unterschiedlich die "Werte- bzw. Wertinstanzen" der Persönlichkeit tangieren. Von der Aussage meinen wir, daß sie unperzönllicher, indirekter als die Frage wirkt (vgl. FREUDENBERG 1970, S. 56 ff.). Das Befragungsergebnis wird damit eher eine unreflektierte, der Selbstzensur entlegene Widerspiegelung darstellen und somit der Realität besser entsprechen. Von der Frage erwartet man eine stärkere suggestive Wirkung, so daß dadurch stärker werbezogene Antworten induziert werden. Bisher wurde nur ungenügend erforscht, ob Frage oder Aussage tatsächlich unterschiedliche Befragungsergebnisse ergeben und bei welchen Indikatorsbereichen der erwartete Effekt eintritt. Durch einige Forschungsberichte (U 69, U 70) wird die Vermutung nahegelegt, daß die beschriebene Wirkung von Frage- und Aussageformulierungen zumindest bei ideologisch relevanten Fragestellungen zu beobachten ist.

In der Untersuchung U 70 wurde beispielsweise die Einstellung junger Bürger zur DDR einmal mit einer Frage- und einmal mit einer Aussageformulierung gleichen Inhaltes abgefragt. Die Frage und ihr Antwortmodell lauteten:

Sind Sie stolz, ein junger Bürger unseres sozialistischen Staates zu sein?

1. ja, sehr	21
2. ja	43
3. schon etwas	21
4. weder noch	5
5. nein	2
6. zur Zeit keine feste Meinung	7
- k.A.	1

Als Aussage und Antwortmodell wurden vorgegeben:

Ich bin stolz, ein junger Bürger unseres sozialistischen Staates zu sein.

1. das ist vollkommen meine Meinung	41
2. das ist mit gewissen Einschränkungen meine Meinung	48
3. das ist kaum meine Meinung	7
4. das ist überhaupt nicht meine Meinung	2
- k.A.	1

Hinter den Antwortstufen haben wir die relativen Antworthäufigkeiten vermerkt.

Die Interpretation der aufgefundenen Ergebnisse ist, wie nachstehende Überlegung zeigt, nicht unproblematisch: Fassen wir beim Frageitem die Stufen 1 und 2 des Antwortmodells zusammen, ergibt das einen Prozentsatz von 64 % positiv eingestellter Analysepersonen. Nehmen wir gleichzeitig beim Aussageitem die Stufen 1 und 2 des Antwortmodells zusammen, erhalten wir 89 % positiv eingestellter Analysepersonen. Für den Interpreten ergibt sich hier eine Differenz von 25 % im Befragungsergebnis, die nur schwer erklärbar ist. Gegen obige Interpretation lassen sich berechtigt Bedenken anmelden.

Es ist hier ungeklärt, ob die Stufen 1 und 2 des Antwortmodells zum Frageitem der Stufen 1 und 2 des Antwortmodells zum Aussageitem äquivalent sind. Wir können ebenso vermuten, daß die Stufe 1 des Aussageitems der Stufe 1 und 2 des AM zum Frageitem entspricht.

Stellen wir die Ergebnisse dieses Vergleiches gegenüber, so erhalten wir 64 % positiv eingestellte Analysepersonen bei der Frage und 41 % bei der Aussage. Das Ergebnis dieser Gegenüberstellung würde insofern befriedigen, weil es mit unserer Erwartung über die psychologische Wirkung von Frage und Aussage auf die Analyseperson konform geht.

Würden wir diese Auslegung des Befragungsergebnisses akzeptieren, müßten wir weiter fragen, wohin denn die Stufe 2 des AM zum Aussageitem gehört. Haben etwa die Probanden, die diese Stufe wählten schon eine negative Einstellung? Diese Frage ist ohne spezielle empirische Analyse nicht zu beantworten.

Wir sehen am Beispiel, daß eine Entscheidung über die Wirkung von Frage und Aussage nicht möglich ist, weil der Versuchsplan, mit dem der Sachverhalt angezielt wurde, ungeeignet ist. <sup>1)</sup> Im Beispiel treten nicht nur Frage und Aussage als unabhängige Variable, sondern auch die Antwortmodelle als solche auf. Die Unterschiede in den Häufigkeitsverteilungen sind damit nicht mehr aus der Frage- oder Aussageformulierung des Items zu erklären, denn sie können ebenfalls Wirkung der unterschiedlichen Antwortmodelle sein.

Will man trotzdem die Stufen in den unterschiedlichen Antwortmodellen vergleichen, müssen wir uns das quantitative Niveau, das die verbalen Urteile verkörpern, vergegenwärtigen. Auch bei diesem Vergleich sind wir zum Teil nur auf Vermutungen angewiesen, teilweise können wir aber methodenkritische Ergebnisse zu Rate ziehen. So ist BISSER (1959) der numerischen Äquivalenz verbaler Quantitätsurteile nachgegangen. Hier zeigte sich, daß die Stufe "mit gewissen Einschränkungen" einem quantitativen Niveau (bezogen auf eine Hunderterteile) von 72 % Zustimmung bei einer Streuung von 7,2 % entspricht. In den Indikatursbereich von "mit gewissen Einschränkungen" des Aussagen-Antwortmodells fällt die Stufe "sehen etwas" des Frage-Antwortmodells hinein, so daß die Zusammenfassung der ersten drei Stufen des Frage-Antwort-

1) Da gerecht zu sein, weisen wir darauf hin, daß der methodenkritische Vergleich von Frage und Aussage nicht Gegenstand der Studie, der wir das Beispiel entnehmen, war.









modelle in seinem Bedeutungsgehalt den ersten beiden Stufen des Aussage-Antwortmodells entspricht. Im Ergebnis stehen sich dann die Werte 85 % (Frage) und 89 % (Aussage) gegenüber. Aus dem Ergebnis geht hervor, daß die Differenz in den Befragungsergebnissen bei Berücksichtigung der quantitativen Niveaustufen der verbalen Urteile schmälert.

Wir sehen also, daß unser gegenwärtiges Wissen über den Einfluß von Frage und Aussage auf das Befragungsergebnis über Vermutungen und Plausibilitätsüberlegungen nicht hinausgeht.

Die Analyse des Einflusses von Frage- und Aussageformulierung sowie des Einflusses der Stufenzahl von Antwortmodellen auf das Befragungsergebnis ist, wie unser kurzer Problemaufriß zeigt, von großer Bedeutung für die Verbesserung der Fragebogenkonstruktion und damit ein wesentlicher Beitrag für die Erzielung gegenstandsadäquater Ergebnisse.

Die Anlage spezieller methodenkritischer Untersuchungen zu beiden Problemen ist nicht nur wissenschaftswert, sondern garantiert auf der empirischen Ebene des Forschungsprozesses den Erkenntnisfortschritt unserer Wissenschaftsdisziplin.

In der Studie FD- 72 wollen wir einen ersten Schritt in dieser Richtung gehen.

Es sei verangewendet, daß es uns in einer einzigen Studie nicht möglich ist auf alle Einzelfragen, die die aufgeworfenen Probleme implizieren, auch umfassend zu antworten. <sup>2)</sup>

FD- 72 steht am Anfang einer Reihe systematischer Analysen zur Befragungsmethodik.

Bei den Analysen zur Stufenzahl von Antwortmodellen werden wir uns in FD- 72 (a) auf die Analyse des Antwortmodells zu Meinungsitems und (b) auf vier-, fünf- und sechstufige verbal-numerische AM hierzu beschränken.

Eine Begründung, weshalb wir in FD- 72 Antwortmodelle zu Meinungsindikatoren untersuchen, erscheint, gemessen an der

2) Ein solches Vorhaben würde die personelle und finanzielle Kapazität unseres Instituts auf Jahre völlig binden.

Bedeutung, die diesen Items in der Forschungspraxis unseres Instituts zukommt, überflüssig. Anders ist es jedoch bei der Stufenzahl der Antwortmodelle. Möglicherweise erscheint den einen oder anderen die Analyse zwei-, drei- oder siebenstufiger Antwortmodelle ebenfalls dringlich.

Eine erste Begründung für die Wahl vier- bis sechstufiger Antwortmodelle besteht darin, daß sie in der Forschung zur Zeit dominieren und ihre methodenkritische Absicherung - auch im Nachhinein - wünschenswert ist.

Jedoch ist u.ä. diese Begründung nicht ausreichend. Wir wollen deshalb einen kurzen Exkurs zur Funktion der Antwortmodelle im Interview geben.

Die Zuordnung von Antwortmodellen zu Fragen und Aussagen geschieht im Interview mit der Absicht, das Antwortverhalten der Analysepersonen unabhängig von der subjektiven Intention des Auswerters zu kategorisieren und wenn möglich, quantitativ zu kennzeichnen.

Antwortmodelle im obigen Sinne konstituieren sich aus einer Menge geordneter Entscheidungsalternativen, die ein Verhaltenskontinuum abdecken. Wie genau und wie differenziert die Entscheidungsalternativen das Kontinuum repräsentieren hängt zumindest von zwei Voraussetzungen ab, die voneinander nicht unabhängig sind:

1. Zunächst können wir davon ausgehen, daß die Abbildung des Verhaltens umso präziser ist, je mehr Entscheidungsalternativen eines Kontinuums wir den Befragten vorgeben. Aus dem Sicht des Forschers sind höherstufige Antwortmodelle niederstufigen vorzuziehen, weil sie a priori eine differenzierte Abbildung des Untersuchungsgegenstandes versprechen.

Der Differenzierungsfähigkeit des Antwortmodells wird aber durch das Differenzierungsvermögen der Analysepersonen Grenzen gesetzt. Die Analysepersonen denken und urteilen über sozialwissenschaftliche Phänomene, über psy-

chischen Erleben in Alltagsbegriffen, die meist unter dem Differenzierungsniveau des Sozialwissenschaftlers liegen. Die Analysepersonen werden deshalb mit einem niederstufigen Antwortmodell im allgemeinen besser erreicht kommen. Aus der Divergenz zwischen Zielstellung des Forschers und Differenzierungsvermögen der Analysepersonen haben sich in der Praxis gewissermaßen optimale Stufenzahlen von Antwortmodellen herauskristallisiert, die bei verbalen Antwortmodellen zwischen zwei und sechs liegen.

Die obere Begrenzung der Stufenzahl von verbalen Antwortmodellen läßt sich auch durch Ergebnisse der Gedächtnisforschung begründen. Hier konnte immer wieder gezeigt werden, das das Simultangedächtnis (sog. Gedächtnisspanne) des Erwachsenen in der Regel nicht mehr als vier bis sechs Elemente speichern kann.

Bei Befragungen setzen wir ja voraus, das die Analysepersonen beim Abarbeiten einer Umfrage-Batterie das Antwortmodell vollständig präsent haben. Durch die Gedächtnisspanne ergibt sich also eine natürliche Schwelle, die die Stufenzahl verbaler Antwortmodelle begrenzt.

2. Von den Entscheidungsalternativen fordern wir, das sie das Kontinuum vollständig abdecken. Jede Analyseperson soll die Möglichkeit haben, sich unbedingt einer der vorgegebenen Entscheidungsalternativen zuordnen zu können.

Bei verbalen Antwortmodellen ergibt sich die Notwendigkeit, verbale quantitative Urteile zu finden, die diese Indikatorenfunktion erfüllen. In einem Modell, das die Intensität der Meinungsäußerung erfasst, wird das Kontinuum durch verbale Urteile begrenzt, die an dem einen Extrem völlige Ablehnung und an anderen vollkommenen Zustimmung zum Ausdruck bringen. Die Wahl der Bezugspunkte, die das Kontinuum als Ankerpunkte determinieren, bereitet hier keine Schwierigkeiten. Diese Wertmarken werden mit Sicherheit von den Analysepersonen eindeutig verstanden. Größere Probleme be-



reitet dagegen die Aufteilung des Kontinuums. Wir müssen Wertmarken (verbale quantitative Urteile) finden, die das Kontinuum mit Stetigkeit, also monoton wachsend, repräsentieren.

Abbildung 1 soll den Sachverhalt verdeutlichen.

Abb. 1: Quantitative Entsprechung verbaler Urteile in verschiedenenstufigen Antwortmodellen (hypothetisch)

zweistufig	vierstufig	sechsstufig	dreistufig	fünfstufig	siebenstufig
AM	AM	AM	AM	AM	AM
100 %	100 %	100 %	100 %	100 %	100 %
					84 %
(75 %)	67 %	80 %		75 %	67 %
		60 %	50 %	50 %	50 %
		40 %			34 %
(25 %)	33 %	20 %		25 %	17 %
0 %	0 %	0 %	0 %	0 %	0 %

Jede Stufe eines beliebigen Antwortmodells in Abbildung 1 muß mit einer Wortmarke belegt werden, die von den Analysepersonen eindeutig als ein anderer Grad von Zustimmung erlebt wird.

Wenn auch verbale quantitative Urteile in beträchtlicher Zahl vorliegen, ist doch die Anzahl von eindeutig differenzierbaren Urteilen, die auch noch in eine Antwortdimension passen, gering. Wir werden mit Wahrscheinlichkeit deshalb mit einem niederstufigen Antwortmodell (etwa mit einem vierstufigen) genauere Ergebnisse erzielen, als mit einem sehr differenzierten Antwortmodell (mehr als sechs Stufen).

Übertragen wir gebräuchliche Antwortmodelle aus unserer Forschungsarbeit auf Abbildung 1, so zeigt sich, daß die Entscheidungsalternativen (die Extrempunkte ausgenommen) immer einen unterschiedlichen Indikatorbereich verkörpern. So umfaßt die Stufe "kaum meine Meinung" in einem vierstufigen Antwortmodell etwa 50 % des Kontinuums, in einem fünfstufig-

schärfenbestimmung). Die Trennschärfenanalyse darf dabei nicht nur das gesamte AM betrachten, sondern muß ebenfalls einzelne Stufen und Zusammenfassungen der Stufen einbeziehen.

2. Lassen sich die Antwortmodelle ineinander transformieren? Die Beantwortung dieser Frage ist insbesondere für Verallgemeinerungen von Befragungsergebnissen aus Studien mit unterschiedlichen AM wesentlich. Für die Analyse werden Korrelationen und Analysen der Veränderungen in der Interkorrelationsstafel herangezogen.
3. Welches quantitative Äquivalent entspricht den verbalen Urteilen in den einzelnen Antwortmodellen? Aufschluß erhalten wir über Paarvergleich und Zuordnung von Prozentzahlen durch die Analysepersonen.
4. Ist es möglich, ein Antwortmodell aufzustellen, das sowohl die Kriterien unter 1. erfüllt und gleichzeitig eine höhere Äquidistanz der verbalen Urteile sichert?
5. Ändert sich durch Verwendung eines anderstufigen Antwortmodells die Indikatorfunktion eines Items? Diese Frage können wir mittels Faktorenanalyse beantworten.

Bei der Frage-/Aussageproblematik läßt sich die Zielstellung von FB-72 wie folgt präzisieren:

1. Unterscheiden sich die Häufigkeitsverteilungen zu Frageitems signifikant von Häufigkeitsverteilungen zu Aussageitems? Als Methode stehen uns Verteilungsanalysen mittels Kolmogorov-Smirnov-Test zur Verfügung.
2. Lassen sich die Ergebnisse zu beiden Itemformulierungen ineinander transformieren?  
- Korrelationsanalyse



3. Unterscheiden sich die Ergebnisse hinsichtlich Stabilität und Differenzierungsfähigkeit?
  - Varianzanalyse
  - Trennschärfenanalyse
  
4. Inwieweit sind die Antwortstufen bei Frage- und Aussageformulierung einander quantitativ äquivalent?
  - Zuordnung von Prozentzahlen durch Analysepersonen

## 2. Untersuchungsmethodik

### 2.1. Fragebogen zur Analyse der Antwortmodelle

Gegenstand der Analyse von FI-72 sind (erstens) Antwortmodelle, die die Intensität einer Meinungsäußerung abbilden. Im einzelnen werden nachstehende Antwortmodelle untersucht:

#### A: vierstufiges Antwortmodell

- 1 vollkommen meine Meinung
- 2 mit gewissen Einschränkungen meine Meinung
- 3 kaum meine Meinung
- 4 überhaupt nicht meine Meinung

#### B: fünfstufiges Antwortmodell

- 1 vollkommen meine Meinung
- 2 mit gewissen Einschränkungen meine Meinung
- 3 teilweise meine Meinung
- 4 kaum meine Meinung
- 5 überhaupt nicht meine Meinung

#### C: sechsstufiges Antwortmodell

- 1 vollkommen meine Meinung
- 2 in großen und ganzen meine Meinung
- 3 mit gewissen Einschränkungen meine Meinung
- 4 ungefähr meine Meinung
- 5 kaum meine Meinung
- 6 überhaupt nicht meine Meinung

Die Antwortmodelle A, B, C sind die unabhängigen Variablen unserer Untersuchung. Ihren Einfluß auf das Antwortverhalten können wir nachweisen, wenn

- a) die Items konstant gehalten werden: (in unserem Fall sind es Ausprägungen, die auf ideologische Einstellungen abzielen)

b) jede Versuchsperson zu jedem Item und jedem Antwortmodell Stellung nimmt.

Für die Untersuchung wählten wir aus den wesentlichen Studien unseres Institutes sechs Items aus, die mit gewisser Regelmäßigkeit zur Abbildung ideologischer Einstellungen verwendet werden.

a) Ich bin davon überzeugt, daß sich die sozialistische Gesellschaftsordnung in der ganzen Welt durchsetzen wird.

b) Ich bin stolz, ein Bürger unseres sozialistischen Staates zu sein.

c) Ich bin der festen Meinung, daß die Sowjetunion unser bester Freund ist.

d) Die SED besitzt mein Vertrauen.

e) In der Zukunft werden sich Sozialismus und Kapitalismus so weit annähern, daß alle wesentlichen Unterschiede zwischen ihnen verschwinden.

f) Der Aufbau des Sozialismus in der DDR ist nur unter der Führung der SED zu verwirklichen.

Die methodische Gestaltung des Fragebogens zur Analyse von Antwortmodellen ist schwierig, weil die Analysepersonen zu jedem Befragungsinhalt (Item a bis f) dreimal antworten müssen. Die massierte Darbietung der Items und Antwortmodelle ließ unerwünschte Nebenwirkung auf die Untersuchungsergebnisse befürchten. Bei massierter Darbietung sind Sättigung, Ermüdung und Frustrationen der Analysepersonen nicht unwahrscheinlich. Beispielsweise können die Befragten bei diesem Vorgehen annehmen, daß wir die "Logik" ihres Antwortverhaltens kontrollieren. Eine solche Vermutung der Analysepersonen würde dazu führen, daß zumindest ein Teil der Befragten die Konsistenz der Antworten durch Selbstzensur sichert. Aus diesen und

Ähnlichen Gründen erschien es uns nicht ratsam, die Items und Antwortmodelle unmittelbar hintereinander darzubieten, sondern die sechs Standarditems in größere Itemgruppen gleichen Inhalts einzubetten. Ferner wurden zur Kaschierung der Befragungsabsicht neben ideologisch relevanten Items auch solche, die Einstellungen zu Familie, zu Intimbeziehungen, zum Wehrdienst induzieren, in den Fragebogen aufgenommen. Ebenfalls mit der gleichen Absicht gingen Kurzskalen zur Messung "Sozialer Aufgeschlossenheit", "Emotionaler Labilität", Angst und "ehrliches versus unehrliches Aussageverhalten" in die Enquete ein.

Zur Sicherung präziser Untersuchungsergebnisse war noch ein weiterer Störfaktor, der sich aus der Expositionsfolge der Antwortmodelle ergeben kann, auszuschließen.

Wir müssen bei der Planung der Untersuchung davon ausgehen, daß ein Antwortmodell, das anderen AM vorangeht, die Reaktionsverteilung der Analysepersonen bei nachfolgenden Antwortmodellen beeinflusst. Bekanntlich sind in den Antwortmodellen A, B, C (vgl. S. 12) einige Stufen verhal. identisch. Es ist also damit zu rechnen, daß diese Stufen in einem nachfolgenden AM aus Gründen der Bekanntheit oder einer vorangehenden Wahlentscheidung bevorzugt werden.

Den Einfluß dieser Fehlerquelle auf das Befragungsergebnis versuchten wir durch Balancierung der Bedingungen auszuschließen (s. Abb. 2):

Abb. 2: Versuchsplan zu FB-72 (1 bis 3)

Fragebogenvariante	Reihenfolge der AM		
	1	2	3
FB 72 - 1	A	B	C
FB 72 - 2	B	C	A
FB 72 - 3	C	A	B

Wir erarbeiteten deshalb drei Fragebogenvarianten (FB 72-1, FB 72-2, FB 72-3). In FB 72-1 werden die Antwortmodelle in

der Reihenfolge A, B, C, in der Variante FB 72-2 in der Anordnung B, C, A, und in FB 72-3 in der Abfolge C, A, B dargeboten. Durch diesen Versuchsplan wird erreicht, daß jedes Antwortmodell einmal an erster, an zweiter und an dritter Stelle zur Wirkung kommt. Auf diese Weise können wir (a) den Einfluß aufeinanderfolgender Antwortmodelle analysieren und (b) durch Zusammenfassung der Versuchsergebnisse aller drei Varianten die Wirkung dieser Störgröße durch Balancierung eliminieren.

Bei der Gestaltung des Fragebogens sicherten wir ferner, daß die Probanden vor der Bearbeitung der Standarditems Zeit und Gelegenheit hatten, sich mit dem jeweiligen Antwortmodell vertraut zu machen. Zu diesem Zweck mußten die Analysepersonen vor der Bearbeitung der Standarditems zu anderen Umfragebatterien (UB), denen das gleiche Antwortmodell beigegeben war, Stellung nehmen.

Abb. 3 informiert über den Aufbau der Fragebogen 1 bis 3 zu FB-72. <sup>3)</sup> In der Übersicht wurden identische Elemente ausgelassen. Dies betrifft die Instruktion zur Einführung in die Befragungsabsicht, eine Startfrage sowie die Fragen 79 bis 107, die den Paarvergleich zur Quantifizierung verbaler Urteile umfassen.

3)

Die Abkürzungen in Abb. 3 bedeuten:

UB - Umfragebatterie

N - Anzahl der Items in UB

AM - Antwortmodell

Ideol. Ein. - Ideologische Einstellung, UB schließt jeweils die Standarditems ein.

Abb. 3: Makroaufbau der Fragebogen zu FB 72 (1 bis 3)

FB 72-1	FB 72-2			FB 72-3			N	AM
	UB	N	AM	UB	N	AM		
Familie	6	A	Intim	6	B	Persönlichkeitskalen	36	C
Ideol. Ein.	10	A	Wehrdienst	4	B	Ideol. Ein.	6	C
Intim	6	B	Ideol. Ein.	9	B	Familie	6	A
Wehrdienst	4	B	Persönlichkeitskalen	36	C	Ideol. Ein.	10	A
Ideol. Ein.	9	B	Ideol. Ein.	6	C	Intim	6	B
Persönlichkeitskalen	36	C	Familie	6	A	Wehrdienst	4	B
Ideol. Ein.	6	C	Ideol. Ein.	10	A	Ideol. Ein.	9	B
Zuordnung von Prozentzahlen zu verbalen Urteilen	6	C	Zuordnung von Prozentzahlen zu verbalen Urteilen	4	A	Zuordnung von Prozentzahlen zu verbalen Urteilen	5	B

Aus Abb. 3 geht hervor, daß die Umfragebatterie, die in unserer Untersuchung vor allen die Einarbeitung in das Antwortmodell gewährleisten, vor jedem kritischen Befragungsbereich (Standarditem) die gleichen sind. Vor den Standarditem zum Antwortmodell A steht in allen Fragebogenvarianten immer die UB "Familie", vor B die Batterien "Intimbeziehungen" und "Wehrdienst" und vor C der Komplex "Persönlichkeitskalen". Dieser Aufbau der Fragebogen ist notwendig, damit mögliche Veränderungen in den Befragungsergebnissen eindeutig auf die Antwortmodelle bezogen werden können und diese Erklärung nicht mit einer Alternativhypothese, die auf dem Kontextbezug fußt, konkurriert.

## 2.2. Fragebogen zur Analyse der Frage-/Aussagewirkung

Für die Untersuchung unserer zweiten Problematik ist es notwendig, den gleichen Sachverhalt den Analysepersonen einmal als Frage und einmal als Aussage vorzulegen, wobei die Antwortmodelle konstant gehalten werden.

Den Fragen und Aussagen haben wir jeweils ein vierstufiges Antwortmodell zugeordnet. Das Modell für die Aussagesysteme ist mit dem Antwortmodell A unter 2.1. identisch.

Für die Frageformen mußte das Antwortmodell neu formuliert werden:

- 1 ja
- 2 mit gewissen Einschränkungen
- 3 kaum
- 4 nein

Die Extrempositionen wurden in ja bzw. nein umgebannt. Die Stufen 2 und 3 sind dann als einschränkende Grade des ermittelnden Urteils zu verstehen.

Die Aussagesysteme sind ebenfalls mit denen unter 2.1. angeführten Standardisema identisch. Die Transformation dieser Aussagen in Frageformen ergab nachstehende Formulierungen:

- a) Sind Sie überzeugt, daß sich die sozialistische Gesellschaft in der ganzen Welt durchsetzen wird?
- b) Sind Sie stolz, Bürger unseres sozialistischen Staates zu sein?
- c) Sind Sie der festen Meinung, daß die Sowjetunion unser bester Freund ist?
- d) Besitzt die SED Ihr Vertrauen?
- e) Werden sich in der Zukunft Sozialismus und Kapitalismus so weit annähern, daß alle wesentlichen Unterschiede zwischen ihnen verschwinden?
- f) Ist der Aufbau des Sozialismus in der DDR nur unter Führung der SED zu verwirklichen?

Beim Aufbau des Fragebogens ergaben sich ähnliche Probleme wie beim Fragebogen zur Analyse von Antwortmodellen.

Hier müssen ebenfalls damit rechnen, daß sich die Versuchsbedingungen (Frage-F) und Aussage-R) gegenseitig beeinflussen. Um die hier untersuchte Wechselwirkung zu erkennen, mußten zwei Fragebogenentwürfe (FB 72-4 und FB 72-5) entwickelt

werden. In FB 72-4 werden die Standarditems - kaschiert durch andere Umfragebatterien und Persönlichkeitskalen - zunächst in Frageform und später in Aussageform dargeboten. In Fragebogen FB 72-5 wird die Reihenfolge dieser Bedingungen vertauscht. Abbildung 4 gibt einen Überblick über den Mischbau von FB 72-4 und 5.

Abb. 4: Mischbau der Fragebogen zu FB 72-4 und 5

FB 72-4		FB 72-5	
UR	N D/E	UR	N D/E
Familie	6 D	Persönlichkeitskalen	36 E
Ideol. Einstellung	10 D	Ideol. Einstellung	6 E
Intimbeziehungen	6 D	Familie	6 D
Wehrdienst	4 D	Intimbeziehungen	6 D
Ideol. Einstellung	9 D	Ideol. Einstell.	10 D
Persönlichkeitskalen	36 E	Wehrdienst	4 D
Ideol. Einstellung	6 E	Ideol. Einstell.	9 E
Zuordnung von Prozentzahlen zu verbalen Urteilen	4 E	Zuordnung von Prozentzahlen zu verbalen Urteilen	4 D

Der formale Aufbau von FB 72-4 und FB 72-5 ist mit den Fragebogen von FB 72 (1 bis 3) identisch, so daß es möglich ist, bestimmte Sachverhalte auch zwischen diesen methodischen Varianten zu vergleichen.

### 1. Zur Interviewsituation und Analysestrategie

Entscheidende Voraussetzungen für eine präzise Analyse des Einflusses von Antwortmodell bzw. Itemformulierung haben wir durch besonders methodische Abweichungen im Fragebogaufbau (vgl. 2.) getroffen. In Ergebnis dieser Überlegungen erhalten wir für die Methoden-Studie FB 72 fünf Fragebogenvarianten, wodurch sich eine besondere Gestaltung der Interviewsituation notwendig macht.



Die Fragebogen werden in Gruppeninterviews den Analysepersonen vorgelegt. Innerhalb eines jeden Gruppeninterviews werden die Analysepersonen nach Zufall den einzelnen Fragebogenvarianten zugeordnet.

Den Analyseleistern erleichterten wir diese Aufgabe durch verschiedenfarbige Fragebogen. Die Randomisierung der Analysepersonen in der Interviewsituation sichert die innere Gültigkeit unserer Versuchsanordnung. Wir können bei diesem Verfahrensmodus erwarten, daß keine Differenzen in den Befragungsergebnissen entstehen, die aus Stichprobensystematik erklärt werden könnten. Unterschiede in den Ergebnissen werden durch Randomisierung der Ana. eindeutig auf die Versuchsbedingungen (I bis III) rückführbar.

Für die Untersuchung FB 72 standen uns ca. 600 Analysepersonen zur Verfügung, die sich auf die Fragebogenvarianten wie folgt verteilten:

Tabelle 1: Stichprobensammensetzung zu FB 72

Fragebogenvariante	männlich	weiblich	gesamt
FB 72-1	50	65	115
FB 72-2	49	74	123
FB 72-3	49	73	122
FB 72-4	54	66	120
FB 72-5	54	67	121
<b>Summe:</b>	<b>254</b>	<b>345</b>	<b>609</b>

Die Gesamtstichprobe setzt sich jeweils aus ca. 200 Ober-  
 schülern, Lehrlingen und Studenten zusammen. Die Teilstich-  
 proben entsprechen damit den Hauptbefragungsgruppen unseres  
 Instituts. Die Lehrlinge, Oberschüler und Studenten verteil-  
 ten sich zu etwa gleichen Anteilen auf die Stichproben zu  
 den Fragebogenvarianten. Wir können hier auf eine Spezifi-  
 zierung der Teilstichproben verzichten, weil wir in diesen

Bericht auf Stichprobenspezifika nicht eingehen werden. Es sei nur soviel vermerkt, daß die Differenzen zwischen den Teilgruppen statistisch unerheblich sind und somit die Zusammenfassung der Daten gerechtfertigt ist.

Zu jedem Gruppeninterview wurde von dem Analytiker ein kurzes Verlaufsprotokoll gefordert, anhand dessen wir die Einhaltung der Versuchsbedingungen kontrollierten.<sup>3)</sup>

Vor der Übergabe der Fragebogen an die Abt. DV wurde durch die Forschungsleiter geprüft, ob die Bereitschaft zur Mitarbeit in allen Interviewgruppen als annähernd gleich eingeschätzt werden konnte, ob Ereignisse (Störungen) eintreten, die besser auf eine Auswertung des Materials verzichten ließen, und ob die Zufallsaufteilung der Ayn. strikt durchgeführt wurde.

4) Es sei nur am Rande erwähnt, daß 120 Fragebogen unausgewertet vernichtet werden mußten, weil sich ein Interviewer nachweislich nicht an die Versuchsleiterinstruktion gehalten hatte.

#### 4. Untersuchungsergebnisse

##### 4.1. Zum Einfluß verbal-numerischer Antwortmodelle auf das Befragungsergebnis bei Meinungsindikatoren

###### 4.1.1. Kontrolle der Abhängigkeit des Befragungsergebnisses von der Reihenfolge der Antwortmodelle

Die Kontrolle der Abhängigkeit des Befragungsergebnisses von der Reihenfolge der Antwortmodelle im Fragebogen ist in unserer Untersuchung zunächst von methodenkritischem Interesse. Die Reihenfolge der Bedingungen A, B, C hatten wir unter 2.1. als Störvariable gekennzeichnet, die das Untersuchungsziel in Frage stellt. Wir hatten uns deshalb zur Eliminierung dieses (hypothetischen) Einflusses durch Balancierung der Bedingungen entschlossen.

Neben dem methodenkritischen Aspekt verdient vorliegende Auswertung darüber hinaus Beachtung für die Fragebogenkonstruktion im allgemeinen. Wir gewinnen hier Daten, die Aufschluß über den Einfluß eines Wechsels von Antwortmodellen in einer Befragung geben. Zu dieser Problematik stehen sich auch in unserem Institut zwei Meinungen gegenüber. Die Anhänger der einen plädieren für die Verwendung möglichst nur eines Antwortmodells in einer Befragung. Ein Wechsel der Antwortmodelle wird immer als Ursache einer möglichen Fehlerquelle betrachtet. Die Anhänger der anderen Meinung bevorzugen unterschiedliche Antwortmodelle in einer Befragung. Ihre Absicht ist es, den Fragebogen stärker an die natürliche Gesprächssituation anzupassen.

In Tabelle 2 haben wir die relativen Häufigkeitsverteilungen zu den Items a bis f beim vier-, fünf- und sechstufigen Antwortmodell zusammengefaßt. Bei jedem Item und Antwortmodell ist die Verteilung, bei der das betreffende Antwortmodell als erstes zur Wirkung kommt, unterstrichen.

Tabelle 2: Relative Häufigkeitsverteilungen zu den Items  
a bis f in FB 72-1 bis 3

Antwortmodell A:

Item	FB 72	Stellung des AM in Fragebogen	Stufen			
			1	2	3	4 5)
a	1	1	1,6	4,9	21,9	71,5
	2	3	0,8	7,3	20,4	71,6
	3	2	2,5	6,6	29,5	61,5
	1-3	-	1,6	6,2	23,9	68,4
b	1	1	0,8	4,1	31,1	64,0
	2	3	0,0	1,6	21,2	77,4
	3	2	0,8	2,5	32,2	64,5
	1-3	-	0,6	2,7	28,2	68,6
c	1	1	0,8	4,9	31,4	63,6
	2	3	0,8	7,3	27,6	64,1
	3	2	1,7	4,1	37,2	57,0
	1-3	-	0,8	5,5	32,0	61,8
d	1	1	3,3	4,9	39,9	52,1
	2	3	0,0	6,5	38,5	55,0
	3	2	2,5	4,1	39,7	53,6
	1-3	-	1,9	5,2	39,3	53,5
e	1	1	4,1	17,1	22,1	56,1
	2	3	5,8	14,2	16,7	63,4
	3	2	4,4	24,8	22,1	48,6
	1-3	-	4,8	18,6	20,5	56,1
f	1	1	0,8	2,4	19,5	77,2
	2	3	0,8	2,4	24,4	72,4
	3	2	0,0	3,3	23,2	73,6
	1-3	-	0,6	2,7	22,3	74,4

5) Die Antwortstufen wurden so codiert, daß aus einer qualitativ hohen Stufe des AM auch ein hoher Einstimmungsgrad erschlossen werden kann.

Unsere Hypothese lautet:

Die Expositionserfolge der Antwortmodelle hängen von folgenden

Artenmerkmalen in den Befragungsergebnissen

ab: Die Hauptkriterien sind die den Brandgefährdungen unter-

schieden nach in den Fragebogenantworten (B 12-1 bis 3) abge-

lesen.

Tabelle 2 (Fortsetzung)

Antwortmodell 2:

Item 20 72	Stellung des AB	in Fragebogen		
		1	2	3
1-3	1	0,8	4,9	17,1
	2	1,6	6,5	19,5
	3	1,6	4,1	29,5
1-3	1	1,4	3,3	22,0
	2	0,8	4,9	24,6
	3	0,8	1,7	27,3
1-3	1	0,8	1,6	21,0
	2	0,8	4,9	24,6
	3	0,8	1,7	27,3
1-3	1	0,8	2,5	29,9
	2	0,8	4,9	29,9
	3	2,5	2,5	34,4
1-3	1	0,8	2,5	28,7
	2	2,4	4,9	30,9
	3	0,8	3,3	33,4
1-3	1	0,8	3,3	33,4
	2	1,7	4,2	33,4
	3	1,7	3,9	32,9
1-3	1	4,1	15,5	19,6
	2	3,3	8,9	20,4
	3	4,2	16,8	16,8
1-3	1	3,8	13,7	8,8
	2	0,8	6,8	4,9
	3	0,8	0,8	24,0
1-3	1	0,8	0,8	24,0
	2	0,8	0,8	24,0
	3	0,8	0,8	24,0
1-3	1	0,8	0,8	24,0
	2	0,8	0,8	24,0
	3	0,8	0,8	24,0

Die Entscheidung darüber, inwieweit wir vorhergehende Hypothese akzeptieren können, soll von zwei beobachtbaren Ereignissen abhängig gemacht werden:

Tabelle 2 (Fortsetzung)

Antwortmodell C:

Item	FB 72	Stellung des AM in Fragebogen	Stufen					
			1	2	3	4	5	6
a	1	3	1,6	4,1	0,0	13,1	12,3	68,9
	2	2	0,0	5,7	1,6	6,5	19,5	66,6
	3	1	1,6	2,5	2,5	12,3	22,2	59,0
	1-3	-	1,1	4,1	1,4	10,6	17,9	65,0
b	1	3	0,8	3,3	0,8	11,6	22,3	61,2
	2	2	0,0	0,8	0,0	4,0	15,3	79,9
	3	1	0,8	1,7	2,5	9,9	23,2	62,0
	1-3	-	0,5	1,9	1,1	8,5	20,2	67,8
c	1	3	0,0	2,5	0,0	17,2	16,4	64,0
	2	2	0,0	3,3	2,4	10,6	20,4	63,5
	3	1	1,7	3,3	4,1	9,1	27,3	54,6
	1-3	-	0,5	3,0	2,2	12,3	21,6	60,6
d	1	3	0,8	2,5	2,5	15,6	27,1	51,6
	2	2	0,0	2,4	1,6	13,0	29,2	53,6
	3	1	1,7	2,5	3,3	18,2	24,8	49,6
	1-3	-	0,8	2,5	2,5	15,6	27,0	51,6
e	1	3	1,6	10,6	12,3	0,8	20,5	54,1
	2	2	5,7	5,7	9,8	4,9	21,9	52,1
	3	1	3,5	11,5	13,3	1,8	20,4	49,6
	1-3	-	3,5	9,2	11,7	2,5	21,0	52,0
f	1	3	0,0	1,6	1,6	11,5	21,3	64,0
	2	2	10,8	0,8	0,0	9,7	14,5	74,1
	3	1	0,0	1,7	4,1	4,1	24,0	66,1
	1-3	-	0,3	1,4	1,9	8,4	19,9	68,2

1. Die erste beobachtbare Größe bildet das Signifikanzniveau. In methodenkritischen Arbeiten betrachtet man allgemein Häufigkeitsverteilungen als divergent, wenn ihr Unterschied auf dem 1 %-Niveau gesichert ist (vgl. LIENERZ 1961). In unserer Analyse setzen wir jedoch das Prüfniveau mit 5 Prozent fest, da wir auch geringfügigere Differenzen in Richtung unserer Annahme werten müssen. Erinnerung sei hier, daß wir die Expositionsfolge als Störgröße im Versuchsplan betrachten, so daß ein höheres Risiko vertretbar ist.
2. Die zweite beobachtbare Größe bezieht sich auf die Regelmäßigkeit, mit der die Unterschiede zwischen den Verteilungen auftreten. Durch Senkung des Signifikanzniveaus besteht die Möglichkeit, daß Unterschiede auch aus anderen hier nicht kontrollierten Faktoren entstehen. Unter Regelmäßigkeit soll hier ein deutliches Überwiegen divergenter Häufigkeitsverteilungen bei den einzelnen Antwortmodellen verstanden werden (exakt mehr als vier signifikante Differenzen von sechs möglichen Vergleichen). Treffen beide Ereignisse gleichzeitig zu, ist dies Indiz, die aufgestellte Hypothese zu akzeptieren. Im anderen Fall werden wir sie zurückweisen.

#### Ergebnisse zum Antwortmodell A:

Antwortmodell A tritt im Fragebogen FB 72-1 an erster Stelle, im Fragebogen FB 72-3 an zweiter und in FB 72-2 an dritter Stelle auf. Die Vergleiche der Häufigkeitsverteilungen zu diesem Antwortmodell ergaben bei keinem Standarditem signifikante Unterschiede. Die Häufigkeitsverteilungen zum vierstufigen Antwortmodell werden also von der Expositionsfolge der Antwortmodelle in Fragebogen nicht nachweisbar beeinflusst.

#### Ergebnisse zum Antwortmodell B:

Als Bezugspunkt für die Vergleiche zum Antwortmodell B dient die Häufigkeitsverteilung in FB 72-2. Hier wird das Antwort-

modell (vgl. Tabelle 2) an erster Stelle geboten. Die Häufigkeitsverteilungen zu den Standarditems bei FB 72-1 und FB 72-3 sind in jedem Falle statistisch homogen. Zusätzlich haben wir bei diesem Antwortmodell geprüft, ob die Mittelposition des AM von der Expositionsfolge beeinflusst wird. Auch hier ergaben sich keine bedeutsamen Differenzen.

Wir halten fest: Die Häufigkeitsverteilungen zum fünfstufigen Antwortmodell werden von der Reihenfolge der Antwortmodelle im Fragebogen nicht beeinflusst.

#### Ergebnisse zum Antwortmodell C:

Das sechsstufige Antwortmodell tritt an erster Stelle in der Fragebogenvariante FB 72-3 auf. Bei diesem Antwortmodell finden wir eine signifikante Differenz beim Standarditem b. Die Analysepersonen zur Fragebogenvariante FB 72-2 (AM an zweiter Stelle) wählen in höherem Maße ( $p = 0,05$ ) die Stufe "sechs". Da nun (vgl. Tab. 2) die gleiche Differenz auch zu FB 72-1 (AM an dritter Stelle) besteht, müssen wir andere Faktoren als die Expositionsfolge für die Unterschiede in der Häufigkeitsverteilung verantwortlich machen. Der Unterschied ist offensichtlich nicht hypothesenrelevant.

Ferner haben wir auch beim sechsstufigen Antwortmodell keine weiteren Hinweise für die Annahme unserer Hypothese gefunden. Zwischen allen anderen Häufigkeitsverteilungen wird die Nullhypothese bestätigt.

#### Zusammenfassung

Die Analyse ergab, daß die Expositionsfolge der Antwortmodelle die Häufigkeitsverteilungen zu den Standarditems nicht nachweisbar beeinflusst. In der weiteren Auswertung können die Ergebnisse der einzelnen Fragebogenvarianten zusammengefaßt werden. Auch sind Vergleiche, bei denen z.B. die Veränderungen beim Übergang von einem vier- zu einem sechsstufigen AM bzw. von einem sechs- zu einem vierstufigen AM betrachtet werden, statthaft.



Für den Aufbau von Fragebogen können wir aus unseren Ergebnissen folgern, daß ein Wechsel dimensionsgleicher AM im Fragebogen die Ergebnisse einer Befragung nicht nachteilig beeinflusst.

Methodisch wäre es legitim, bei sehr ähnlichen Umfrageskalen die Monotonie des Fragebogens durch einen Wechsel dimensionsgleicher Antwortmodelle zu durchbrechen.

#### 4.1.2. Stabilität und Trennschärfe der Antwortmodelle

Im folgenden wenden wir uns der Frage zu, welches Antwortmodell eine differenziertere und stabilere Abbildung von Meinungsäußerungen gestattet.

Wir betrachten zunächst nur die Trennschärfe von Antwortmodellen insgesamt. Die Trennschärfe einzelner Stufen der Antwortmodelle und deren Zusammenfassungen gehen wir in einem gesonderten Abschnitt nach.

Für die Trennschärfenanalyse haben wir drei Kriterien gebildet:

1. Die Trennschärfe der Antwortmodelle (bezogen auf ein bestimmtes Item) wird an Hand der Summenwerte aus den Zuordnungsgraden der Analysepersonen zu den sechs Standarditems berechnet.
2. Die Trennschärfe der Antwortmodelle A, B und C wird aus einer Summe von 10 inhaltsgleichen Items bei einem vierstufigen AM bestimmt.
3. Die Trennschärfe der Antwortmodelle A, B und C wird an einem Kriterium ermittelt, das aus den Summenwerten von 9 Items bei einem fünfstufigen AM besteht.

Tabelle 3 gibt die Trennschärfenkoeffizienten zu den Standarditems beim Kriterium wieder.

Tabelle 3: Trennschärfenkoeffizienten zu den Standarditems beim Kriterium 1

Item		FB 72-1	FB 72-2	FB 72-3	
I.	AM A	a	0,68	0,59	0,69
		b	0,75	0,55	0,77
		c	0,52	0,64	0,78
		d	0,71	0,70	0,76
		e	0,51	0,39	0,49
		f	0,58	0,66	0,72
		$\bar{x}$	0,63	0,59	0,70
II.	AM B	a	0,76	0,57	0,63
		b	0,73	0,52	0,79
		c	0,60	0,59	0,78
		d	0,71	0,62	0,81
		e	0,57	0,46	0,37
		f	0,68	0,59	0,77
		$\bar{x}$	0,68	0,58	0,66
III.	AM C	a	0,69	0,62	0,72
		b	0,80	0,58	0,76
		c	0,65	0,61	0,73
		d	0,77	0,69	0,84
		e	0,52	0,43	0,41
		f	0,75	0,71	0,79
		$\bar{x}$	0,70	0,61	0,71
IV. Balanzierung der Expositionsfolge					
		AM A	AM B	AM C	
Item	a	0,65	0,65	0,68	
	b	0,69	0,69	0,71	
	c	0,65	0,66	0,66	
	d	0,72	0,71	0,77	
	e	0,47	0,48	0,45	
	f	0,65	0,69	0,75	
	$\bar{x}$	0,64	0,63	0,67	

Im Abschnitt I bis III der Tabelle 3 sind die Ergebnisse zum vier-, fünf- und sechststufigen Antwortmodell getrennt nach den Fragebogenvarianten dargestellt.

Die Inspektion der Koeffizienten zwischen den Fragebogenvarianten eines jeden AM macht deutlich, daß die Unterschiede auf Zufallsschwankungen zurückführbar sind. Ein exakter Vergleich über den "t-Test zur Prüfung des Unterschiedes zwischen zwei Korrelationskoeffizienten" (CLAUSE/EBNER 1967) bestätigt diesen Befund.

Wir können auch bei der Trennschärfenanalyse festhalten, daß die Expositionsfolge keinen Einfluß auf die Differenzierungsfähigkeit der Antwortmodelle nimmt. Es wäre immerhin denkbar, daß die Entscheidungssicherheit der Probanden bei mehrfacher Darbietung gleicher Items zunimmt und als Folge des veränderten Entscheidungsverhaltens der Analysepersonen die Trennschärfenkoeffizienten sich erhöhen.

Vernachlässigen wir die (unbedeutenden) Unterschiede zwischen den einzelnen Fragebogenvarianten, indem wir die mittleren Trennschärfenkoeffizienten bei jedem Item und Antwortmodell bestimmen (vgl. Tab. 3/IV), so zeigen sich beim Kriterium 1 keine nennenswerten Unterschiede in der Differenzierungsfähigkeit der Antwortmodelle. In der Tendenz fallen die Ergebnisse zum sechststufigen Antwortmodell etwas günstiger als bei den anderen Modellen aus ( $p = 0,072$ ; FRIEDMAN-Test). In fünf Fällen sind die Koeffizienten des sechststufigen AM numerisch höher als beim vierstufigen und beim fünfstufigen AM. Die gleiche Tendenz zwischen vier- und fünfstufigen Antwortmodellen besteht nicht.

Allerdings ist die Überlegenheit des sechststufigen Antwortmodells keineswegs so deutlich, daß sich daraus praktische Folgerungen für die Gestaltung von Fragebogen ableiten ließen.

In den folgenden Tabellen verzichten wir auf eine detaillierte Darstellung zu den einzelnen Fragebogenvarianten. Auch hier erwiesen sich die Differenzen zwischen den Trennschär-

lenkoeffizienten zu den einzelnen Items und Antwortmodellen als nicht signifikant.

Tabelle 4: Trennschärfenkoeffizienten zu den Standarditems beim Kriterium 2 (Balanzierung der Expositionsfolge)

Item	AM A	AM B	AM C
a	0,63	0,57	0,61
b	0,68	0,59	0,63
c	0,65	0,61	0,61
d	0,73	0,69	0,71
e	0,40	0,36	0,31
f	0,64	0,66	0,66
$\bar{x}$	0,62	0,58	0,59

Tabelle 5: Trennschärfenkoeffizienten zu den Standarditems beim Kriterium 3 (Balanzierung der Expositionsfolge)

Item	AM A	AM B	AM C
a	0,57	0,62	0,57
b	0,66	0,69	0,72
c	0,59	0,64	0,63
d	0,73	0,76	0,75
e	0,24	0,27	0,30
f	0,61	0,70	0,66
$\bar{x}$	0,57	0,61	0,60

Das Kriterium 2 für die Trennschärfenbestimmung wird aus den Summenwerten der Analyseparameter über 10 Items bei einem vierstufigen Antwortmodell gebildet. In der Analyse werden somit vierstufige Antwortmodelle durch die Art der Kriteriums-bildung bevorzugt. Trotz dieser Tatsache ist die Analyse in zweifacher Hinsicht interessant:

1. Sie gibt über die Konstanz der Differenzierungsfähigkeit der Antwortmodelle bei einem Wechsel des Kriteriums Aufschluß.
2. Für die Fragebogenkonstruktion erhalten wir Auskunft, welchen Einfluß die Kriteriumsdefinition auf die methodenkritische Qualifikation des Indikators (Item plus AM) nimmt.

Aus Tabelle 4 ist zu entnehmen, daß die Trennschärferkoeffizienten nur unbedeutend bei jedem Item differieren. Alle Unterschiede liegen im Zufallsbereich. Gleichzeitig geht aus dem Datenmaterial hervor, daß die Differenzierungsfähigkeit bei allen AM konstant ist.

Erwartungsgemäß liegen die Ergebnisse aus vierstufigen AM numerisch etwas höher als beim fünf- und sechsstufigen AM. Diese Tendenz wurde auf Grund der Kriteriumsdefinition erwartet. Die Mittelwerte über alle Items (AM A: 0,62; AM B: 0,56; AM C: 0,59) lassen erkennen, daß die Unterschiede keine praktischen Konsequenzen für die Fragebogenkonstruktion ergeben. Nach diesem Befund ist es unproblematisch, wie wir das Kriterium für die Trennschärfeanalyse in einer konkreten Untersuchung bilden. Natürlich wird bei dieser Aussage vorausgesetzt, daß alle Items hypothetisch ein Merkmal abbilden und die AM die gleiche Dimension erfassen.

Auch in Tabelle 5 gehen wir obigen Sachverhalten nach. Nur ist hier das Kriterium anders definiert. Es besteht aus der Summe über neun Items bei einem fünfstufigen AM. Auch hier erwarten wir auf Grund der Kriteriumsdefinition, daß sich beim fünfstufigen Antwortmodell etwas höhere Trennschärferkoeffizienten ergeben. Anschaulich tritt dieser Effekt beim Vergleich vier- und fünfstufiger Antwortmodelle hervor. In allen Fällen sind die Koeffizienten beim fünfstufigen Antwortmodell numerisch höher. Beim Vergleich mit dem sechsstufigen (vgl. Tabelle 5) tritt diese Tendenz etwas schwächer auf.

Auch beim Kriterium 3 ist die Differenzierungsfähigkeit bei allen Antwortmodellen hochgradig konstant.

Wir wenden uns nun der Stabilitätsanalyse zu den Antwortmodellen A, B, C zu. Zur Ermittlung der Messgenauigkeit der Antwortmodelle sendeten wir einen varianzanalytischen Versuchplan (vgl. HENNERT, 1961) an. Im Ergebnis erhalten wir Kennwerte, die eine Beurteilung des Messfehlers, der auf den Antwortmodellen als Meßinstrument beruht, ermöglicht. In Tabelle 6 sind die  $r_{tt}$ -Koeffizienten in der Expositionsfolge der AM im Fragebogen angegeben. Die Indices informieren über die Stufenzahl des jeweiligen AM.

Tabelle 6: Stabilitätskoeffizienten der Antwortmodelle

Fragebogenvariante	Expositionsfolge		
	1	2	3
FB 72-1	0,57 <sub>4</sub>	0,73 <sub>5</sub>	0,76 <sub>6</sub>
FB 72-2	0,51 <sub>5</sub>	0,57 <sub>6</sub>	0,55 <sub>4</sub>
FB 72-3	0,75 <sub>6</sub>	0,76 <sub>4</sub>	0,65 <sub>5</sub>
$\bar{x}$	0,54	0,68	0,65

Wir wollen prüfen, ob die Stabilität der Antwortmodelle von der Expositionsfolge beeinflusst wird. Hierzu haben wir die Mittelwerte über die  $r_{tt}$ -Werte einer jeden Spalte gebildet. Wir abstrahieren hier also von der Stufenzahl des Antwortmodells. Ähnlich wie bei der Trennschärfe kann man vermuten, daß die Stabilität der Antworten bei mehrfacher Darstellung des gleichen Items zunimmt und damit die speziellen Reliabilitätskoeffizienten künstlich beeinflusst.

In Tabelle 6 kann man sich überzeugen, daß die erwartete Wirkung in unserer Untersuchung nicht eintritt. Die Unterschiede zwischen den mittleren  $r_{tt}$ -Werten sind statistisch unbedeutend. Ebenfalls hat die Zahl der Stufen eines AM (vgl. Tab. 7) keinen nennenswerten Einfluß auf deren Stabilität.

Tabelle 7: Stabilitätskoeffizienten der Antwortmodelle

Fragebogenvariante	AM A	AM B	AM C
FB 72-1	0,67	0,73	0,76
FB 72-2	0,55	0,51	0,57
FB 72-3	0,76	0,65	0,75
$\bar{x}$	0,67	0,64	0,70

Die mittleren  $R_{\text{st}}$ -Werte ergeben folgende Rangfolge über die Stabilität der Antwortmodelle:

- sechststufiges AM (0,70)
- vierstufiges AM (0,67)
- fünfstufiges AM (0,64).

Die Unterschiede zwischen den Koeffizienten sind so gering, daß sich keine praktischen Konsequenzen für die Bevorzugung eines AM ableiten lassen.

Zusammenfassung

1. Ergebnisse zu den Trennschärfebestimmungen der AM:

- Hinsichtlich der Differenzierungsfähigkeit erweisen sich alle analysierten Antwortmodelle als gleichwertig.
- Die Differenzierungsfähigkeit der Antwortmodelle bleibt bei wechselnder Kriteriumsdefinition konstant.

2. Ergebnisse zur Stabilitätsanalyse der AM:

- Die Reihenfolge der Expositionen beeinflusst die Stabilität der Antwortmodelle nicht.
- Ein deutlicher Einfluß der Stufenzahl auf die Maßgenauigkeit der Antwortmodelle konnte nicht nachgewiesen werden. Vom Aspekt der Maßgenauigkeit her ist jedes Antwortmodell zur Abbildung der relevanten Sachverhalte qualifiziert.

#### 4.1.3. Zur Transformation der Antwortmodelle

In der bisherigen Untersuchung erwiesen sich die analysierten vier-, fünf- und sechststufigen Antwortmodelle hinsichtlich Stabilität und allgemeiner Trennschärfe (AM insgesamt) als gleichwertig. Damit erlangt ihre wechselseitige Transformation besondere Bedeutung. Wir werden untersuchen, wie sich die Antwortverteilungen innerhalb der einzelnen Stufen der AM verändern, wenn wir von einem niederstufigen AM in ein höherstufiges übergehen und umgekehrt.

Wir hoffen, durch die Analyse Aufschluß darüber zu gewinnen, wie die Zwischenstufen der Antwortmodelle zusammengefaßt werden müssen, damit vergleichbare Ergebnisse entstehen.

Zunächst wollen wir unsere Erwartungen über die Veränderungen in der Antwortverteilung beim Übergang von einem niederstufigen in ein höherstufiges Antwortmodell schematisch darstellen (Abb. 5). Der Doppelpfeil in Abbildung 5 deutet unsere Erwartung darüber an, in welche Richtung sich die Mehrzahl der Analysepersonen bei einem Wechsel des Antwortmodells bewegen werden. Der einfache Pfeil zeigt somit schwächere Stichprobenanteile an.

Beim Übergang in ein niederstufiges Antwortmodell ist die Transformationsrichtung in Abb. 5 umzukehren.

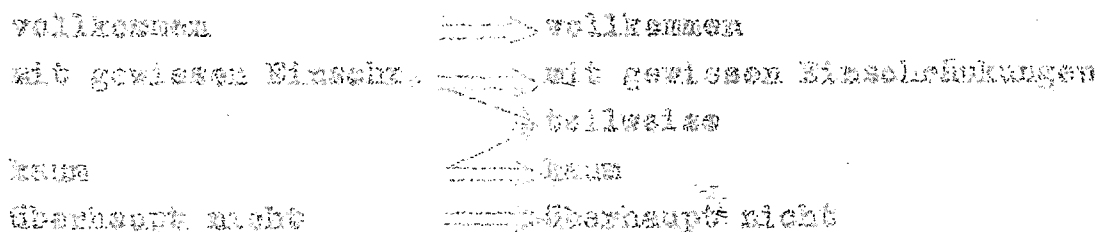
Aus Abb. 5 lassen sich nun einige Hypothesen formulieren, deren Haltbarkeit am Datenumaterial zu FB 72 geprüft werden soll:

1. Die Stufen "willkommen" und "überhaupt nicht" sind in den Antwortmodellen A bis C die Bezugspunkte des Reaktionsfeldes. Sie determinieren als Ankerreize das Kontinuum. Als solche sind sie gegenüber Veränderungen der Zwischenstufen in ihrer Indikatorfunktion invariant. Analysepersonen, die sich beim Antwortmodell A für diese Stufen entschieden haben, wählen sie auch in einem anderenstufigen Antwortmodell wieder. Der gleiche Sachverhalt ist zu beobachten, wenn die Antwortmodelle B oder C an erster Stelle gehalten werden.

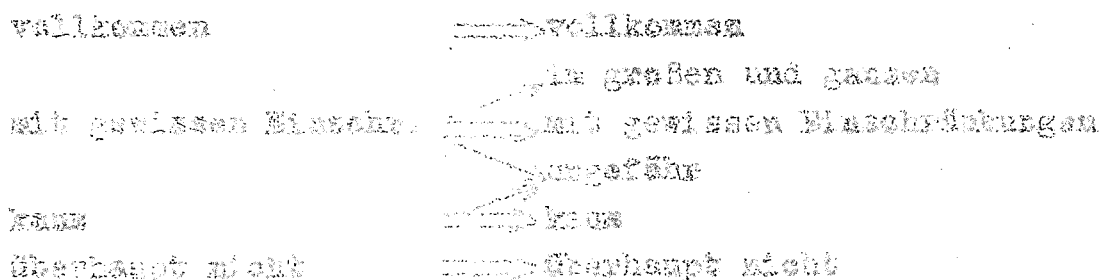


Abb. 5: Veränderungen in der Antwortverteilung beim Übergang von einem niedrigeren in ein höherstufiges Antwortmodell (hypothetische Annahme)

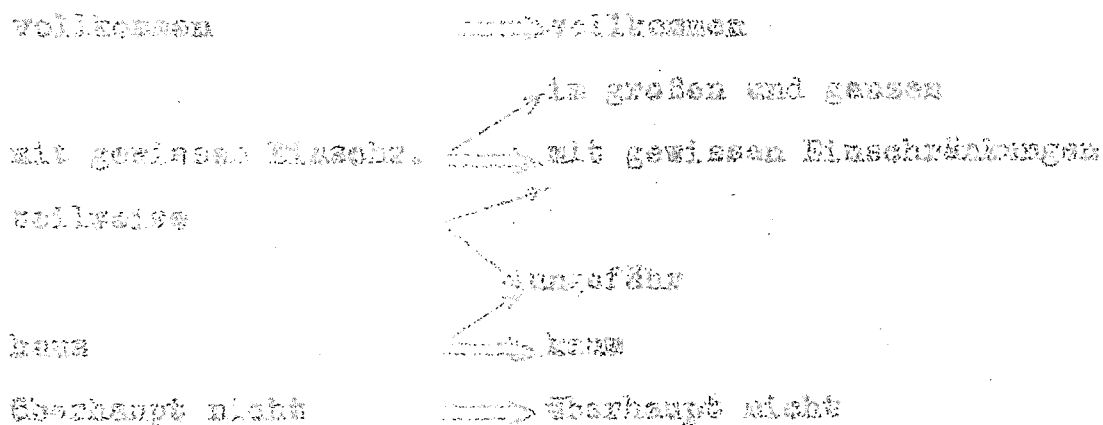
1. Übergang von Antwortmodell A zu AM B



2. Übergang von Antwortmodell A zu Antwortmodell C



3. Übergang von Antwortmodell B zu Antwortmodell C



2. Je differenzierter das Inventarmodell, desto geringer ist die Bedeutung der Stufen "mit gewissen Einschränkungen" und "vollständig".

Die Übertragung von III in II ist vornehmlich durch die Wirkung der Stufe "mit gewissen Einschränkungen" auf die Stufe "vollständig" bedingt. Dabei erwarten wir, daß im allgemeinen die Stufe "vollständig" häufiger gewählt wird als die Stufe "mit gewissen Einschränkungen", obwohl sie hier einem regelmäßigen Zusammenhang korrelieren, wobei wie zu erwarten, bei der Verteilung auf die beiden Stufen wahrscheinlich stark von Flüssigkeit betroffen wird.

Beim Übergang von AM A zu AM C erwarten wir eine Durchlässigkeit der Beobachtungsgruppe, die zwischen "mit gewissen Einschränkungen" liegt. Die Protokolle werden sich wahrscheinlich auf die Stufen "vollständig" und "mit gewissen Einschränkungen" und "unvollständig" verteilen.

Die wesentliche Übertragung bei der Stufe "vollständig" können Abweichungen festzustellen werden.

Bevor wir uns auf eine Hypothese beziehen, wollen wir zunächst von der oben beschriebenen Wechsel der Rollen die gleiche Struktur wieder abbilden. Die Parallelität der Aussagen können wir durch Korrelationsanalysen überprüfen. In Tabelle 8 haben wir die entsprechenden Parallelitätskoeffizienten zusammengestellt. In den Spalten I bis III der Tabelle 8 geben wir Bezug auf die Hypothesen der Antwortmöglichkeiten in Fragebogen. Unsere Überlegung gibt dabei, daß die Parallelitätskoeffizienten unvollständigen und vollständigen der AM höher als zwischen entfernt platzierten AM ausfallen. Beachtenswert ist auch in Tabelle 8 unterstrichen.

Tabelle 8: Paritätskoeffizienten des Antwortmodells

I. Item/Korrelationen	AM A/AM B	AM A/AM C	AM B/AM C
a	0,80	0,86	0,83
b	0,81	0,81	0,90
c	0,85	0,81	0,90
d	0,89	0,85	0,90
e	0,88	0,84	0,90
f	0,77	0,72	0,76
F	0,83	0,80	0,86

II. Item/Korrelationen	AM B/AM A	AM C/AM A	AM B/AM C
a	0,78	0,84	0,82
b	0,78	0,78	0,75
c	0,94	0,85	0,90
d	0,78	0,75	0,85
e	0,93	0,82	0,81
f	0,74	0,77	0,78
F	0,78	0,81	0,81

III. Item/Korrelationen	AM A/AM B	AM C/AM B	AM C/AM A
a	0,83	0,89	0,80
b	0,86	0,86	0,85
c	0,93	0,85	0,91
d	0,86	0,83	0,88
e	0,91	0,90	0,88
f	0,77	0,78	0,75
F	0,86	0,85	0,84

IV. Balanceierung der Hypothesenfolge	Item/Korrelationen	AM A/AM B	AM A/AM C	AM B/AM C
a		0,80	0,86	0,83
b		0,81	0,81	0,90
c		0,85	0,81	0,90
d		0,89	0,85	0,90
e		0,88	0,84	0,90
f		0,77	0,72	0,76
F		0,83	0,80	0,86

In der Tat liegen die Koeffizienten beobachtet als in Fragebogen immer um ein Geringes höher als die Korrelationen zwischen räumlich entfernter angeordneten Antwortmodellen. Jedoch ist der Unterschied so unbedeutend, daß er auf die Höhe der Parallelitätskoeffizienten kaum Einfluß hat. Die Höhe der Parallelitätskoeffizienten kann mit großer Wahrscheinlichkeit aus der Ähnlichkeit der Antwortmodelle erklärt werden.

Aus Tabelle 3/IV geht hervor, daß die durchschnittliche Korrelation mit der Ähnlichkeit der Antwortmodelle zunimmt. So ist die Korrelation zwischen vier- und fünfstufigem bzw. fünf- und sechsstufigem Antwortmodell numerisch etwas höher als zwischen vier- und sechsstufigen Antwortmodellen. Auch hier sind die Unterschiede statistisch unbedeutend. Sie erleichtern uns aber das Verständnis für die wechselseitigen Transformationen der Antwortmodelle. Die Parallelitätskoeffizienten der Antwortmodelle liegen zwischen 0,75 und 0,91 (vgl. Tab. 3/IV). Damit können wir mit etwa 35 bis 95 Prozent konstanter Antwortreaktionen beim Übergang von einem Antwortmodell in ein anderes (bei gleichen Items) rechnen.

Für uns ist aber der Anteil wechsellagerter Analysepersonen von besonderem Interesse. Es soll in folgendem interessiert sein, welche Antwortstufen die Analysepersonen wählen, wenn ihnen die gleichen Items zum zweiten oder dritten Mal mit einem anderen Antwortmodell vorgelegt werden.

Zunächst gehen wir auf die Extrempositionen "vollkommen" bzw. "überhaupt nicht" (s. obige Hypothese) ein.

Tabelle 9: Veränderungen in der Extremposition "vollkommen  
meine Meinung" beim Übergang in ein andersstufi-  
ges Antwortmodell

Transformation	(in anderen AM gewählte Stufe in %)					
	1	2	3	4	5	6
AM A --- AM B	0,0	0,0	0,8	4,5	<u>94,7</u>	-
AM B --- AM A	0,3	1,0	5,2	<u>93,5</u>	-	-
AM A --- AM C	0,0	0,0	0,5	1,5	8,6	<u>89,4</u>
AM C --- AM A	0,0	0,6	3,9	<u>95,5</u>	-	-
AM B --- AM C	0,0	0,0	0,0	0,9	7,7	<u>91,4</u>
AM C --- AM B	0,0	0,5	0,3	4,8	<u>94,6</u>	-

Anmerkung: Anteil konstanter Ipn. in Tabelle 9 unterstrichen.

In Tabelle 9 haben wir jeweils den Übergang von niederstufi-  
gen in ein höherstufiges AM und den Übergang von höherstufi-  
gen in ein niederstufiges AM gegenübergestellt. Die Prozent-  
werte wurden aus der Reaktionsverteilung bei allen sechs Items  
berechnet. Unsere Betrachtung abstrahiert also vom konkreten  
Inhalt der Items. Bei allen Transformationen erweist sich  
die extreme Position als sehr konstant. Auch im ungünstig-  
sten Fall (AM A --- AM C) bleiben 89,4 Prozent der Wahlent-  
scheidungen positionstreu. Bei Inspektion der Tabelle 9  
könnte man vermuten, daß die Stufenzahl des AM die Entschei-  
dung der Probanden für die extreme Position etwas verändert.  
Die geringsten konstanten Werte finden wir jeweils bei Trans-  
formationen in ein sechstufiges AM. Wir sind diesem Sach-  
verhalt mittels verteilungsfreier Varianzanalyse nach FRIED-  
MAN (S. LINHART 1962) nachgegangen. In der Tat ist die Feld-  
besetzung der Stufe "vollkommen" beim sechstufigen AM zu  
geringsten ( $\chi^2 = 9,3; p = 0,0055$ ).

Dieser Befund zeigt, daß auch der Indikatorbereich der Stufe "vollkommen" nicht absolut konstant ist. Bei einem niederstufigen AM ist er offensichtlich größer, wodurch ein konstantes Entscheidungsverhalten der Analysepersonen begünstigt wird.

Die Tatsache, daß die Feldbesetzung der Stufe "vollkommen" von der Stufenzahl des Antwortmodells nicht unabhängig ist, veranlaßt uns Hypothese 1 dahingehend zu präzisieren, daß wir von relativer Invarianz der Extremposition sprechen. Die modifizierte Hypothese wird akzeptiert, wenn die Reaktionsverteilung in der Extremposition konstanter als in den Zwischenstufen des AM ist.

Über die Konstanz der Position "überhaupt nicht meine Meinung" können wir wegen zu kleinen  $n$  bei allen AM keine Aussage treffen.

Tabelle 9 können wir ferner entnehmen, daß ein gewisser Prozentsatz (3,9 % bis 8,6 %) der Wählentscheidungen von der Stufe vollkommen auf die nächst niedrigere Stufe des AM flutet. Diese Stufe steht also in gewisser Konkurrenz mit dem Indikatorbereich der Stufe "vollkommen". Auf Grund des empirischen Materials müssen wir entgegen unserer hypothetischen Annahmen (s. Abb. 5) feststellen, daß auch ein Teil der Analysepersonen, die zunächst die Stufe "mit gewissen Einschränkungen" signifierten, in einem anderen AM sich nicht - wie erwartet - nur für die Stufen "teilweise", "in großen und ganzen" oder "ungefähr", sondern auch für die Stufe "vollkommen" entscheiden.

Tabelle 10: Veränderungen in der Stufe "mit gewissen Einschränkungen meine Meinung" beim Übergang in ein anderestufiges Antwortmodell

Transformations-		(in anderem AM gewählte Stufe in Prozent)					
		1	2	3	4	5	6
AM A	→ AM B	0,0	3,0	12,9	71,8	12,4	-
AM B	→ AM A	0,6	5,3	77,5	16,6	-	-
AM A	→ AM C	0,0	3,3	3,9	24,8	47,5	10,5
AM C	→ AM A	0,0	12,1	84,9	3,0	-	-
AM B	→ AM C	0,0	0,2	2,3	34,6	54,9	7,5
AM C	→ AM B	0,0	4,6	16,7	71,1	7,6	-

In Tabelle 10 sind die Veränderungen in der Stufe "mit gewissen Einschränkungen" beim Übergang in ein anderes Antwortmodell dargestellt. Aus den Daten in Tabelle 10 ist ersichtlich, daß der Anteil der Probanden, die beim Wechsel des Antwortmodells die Stufe "vollkommen meine Meinung" wählen immerhin zwischen drei und 16,6 % liegt. Dieser Prozentsatz ist zu groß, um als Versuchsfehler (Nachlässigkeiten beim Signieren, Irrtümer der App. beim Eintragen der Zahlen) vernachlässigt zu werden. Das Ergebnis zeigt eher, daß die Indikatorbereiche der Stufen in den Antwortmodellen selbst zur Extremposition nicht eindeutig abgegrenzt sind. Beim Übergang von einem vierstufigen in ein fünfstufiges Antwortmodell sind 71,8 % der Wahlen konstant, 12,4 % gehen zur Extremposition, 12,9 % zur Stufe "teilweise" und sogar 3,0 % zur Stufe "kaum" über.

An dieser Stelle ist es zweckmäßig, eine Konvention darüber einzuführen, nach der wir Fluktuationsquoten noch als symptomatisch betrachten bzw. sie als Versuchsfehler klassifizieren. Wir werden einen Prozentwert von 5,0 und größer als eine symptomatische Veränderung und Werte darunter als Versuchsfehler auffassen. In obigen Fall können wir dann auf eine Bewertung des Übergangs zur Stufe "kaum" verzichten. Gehen wir dagegen von einem fünfstufigen Antwortmodell in ein vierstufiges über, entfallen auf die Stufe "vollkommen" 16,6 % und auf die Stufe "kaum" 5,3 %, bei 77,9 % konstanter Wahlen.

Die Zahl konstanter Wahlen sinkt beim Übergang von einem vierstufigen in ein sechsstufiges Antwortmodell rapide ab (vgl. Tabelle 10). Ja sogar die Majorität dieser Stufe geht verloren, denn 47,5 % der Wahlen gehen auf die Stufe "in großen und ganzen" über. Daraus ist zu erkennen, daß der Indikatorbereich der Stufe "mit gewissen Einschränkungen" im vierstufigen Antwortmodell dem Indikatorbereich der Stufen "mit gewissen Einschränkungen" und "in großen und ganzen" des sechsstufigen AM entspricht. Unsere Überlegungen über den Wechsel des Indikatorbereiches gleicher verhalten

Stufen in unterschiedlichen Antwortmodellen (s.S. 9/10) wird hier bestätigt. Ebenfalls geht ein beträchtlicher Prozentsatz (10,5 %) der Wähler beim Übergang von AM A nach AM C auf die Stufe "vollkommen" über, während die erwartete Differenzierung zur Stufe "ungefähr" (3,9 %) ausbleibt. Beim Übergang von Antwortmodell C nach Antwortmodell A ist die Zahl konstanter Fälle beträchtlich (54,9 %). Wir können diesen Sachverhalt so verstehen, daß Probanden, die bei einem höherstufigen Antwortmodell sich für die Stufe "mit gewissen Einschränkungen" entschieden haben, diese Stufe auch mit hoher Wahrscheinlichkeit in einem niedrigerstufigen Antwortmodell signalisieren. Diese Aussage wird erhärtet durch die Verteilungen beim Übergang von AM C nach AM B und beim Übergang von AM B nach AM A (vgl. Tab. 10).

Die Fluktuation beim Wechsel des Antwortmodells von AM C nach AM A beschränkt sich hier auf die Stufe "kaum". 12,1 % der Wähler entfallen auf diese Stufe des niedrigerstufigen AM. Der Übergang von AM B nach AM C weist ähnliche Verteilungsgesetzmäßigkeiten wie der Übergang von AM A nach AM C auf, auch hier sind nur etwa ein Drittel der Wähler (34,6 %) konstant. Der Hauptteil (54,9 %) geht zur Stufe "in großen und ganzen" und ein geringerer Anteil, nämlich 7,5 % zur Stufe "vollkommen" über. Die anderen Stufen des Antwortmodells werden vom Wechsel des Modells kaum berührt.

Es zeigt sich wiederum, daß sich der Indikatorbereich der Stufe "mit gewissen Einschränkungen" in einem niedrigerstufigen Antwortmodell in die Stufen "mit gewissen Einschränkungen" und "in großen und ganzen" in einem höherstufigen Antwortmodell differenziert.

Gehen wir von AM C nach AM B über, so stellen wir dagegen - wie schon weiter oben angedeutet - einen beträchtlichen Prozentsatz konstanter Wahlentscheidungen (71,1 %) fest. Als kennzeichnende Indikatorbereiche erweisen sich die beiden benachbarten Stufen. Auf "vollkommen" entfallen 7,6 % und auf "teilweise" 16,7 % der ursprünglichen Entscheidungen.



An Hand unseres Datennaterials können wir nun zu den obigen Hypothesen Stellung nehmen:

1. Die Invarianz der Extremposition "vollkommen" ist gegenüber einem Wechsel der Stufenzahl von Antwortmodellen größer als die der Zwischenstufe "mit gewissen Einschränkungen". Obwohl auch die Wahl der Extremposition von der Stufenzahl des AM (s. oben) nicht unabhängig ist, kann die Extremposition als ein Ankerreis der Antwortmodelle, die den Grad der Zustimmung zu Meinungsäußerungen erfassen, betrachtet werden.

2. Auf Grund unseres Datennaterials gilt es als erwiesen, daß die Entscheidungskonstante der Apm. bei der Stufe "mit gewissen Einschränkungen" mit dem Differenzierungsgrad der Antwortmodelle abnimmt.

Daneben ließ sich über den FRIEMAN-Test zeigen, daß auch die Wahlhäufigkeit mit der Differenzierung der AM signifikant zusammenhängt ( $K^2 = 12,0; p = 0,0001$ ).

Tabelle 11: Veränderungen in der Stufe "kann keine Meinung" beim Übergang in ein anderenstufigen Antwortmodell

Transferaktion	(in anderen AM gewählte Stufe in Prozent)					
	1	2	3	4	5	6
AM A ---- AM B	2,1	50,0	35,4	8,3	4,2	-
AM B ---- AM A	4,2	25,8	0,0	0,0	-	-
AM A ---- AM C	2,4	48,8	26,8	17,1	4,9	0,0
AM C ---- AM A	7,4	68,7	25,9	0,0	-	-
AM B ---- AM C	3,3	29,0	40,0	3,3	3,3	0,0
AM C ---- AM B	8,0	48,0	40,0	4,0	0,0	-

Auch bei der Stufe "kann" gilt, daß die Konstante des Entscheidungsverhaltens der Analysepersonen beim Übergang von einem niederenstufigen in ein höherstufigen AM (s. Tabelle 11)

geringer ist als bei der Stufe "vollkommen". Dieses Ergebnis bestätigt damit Hypothese 1.

Ebenso wie bei der Stufe "mit gewissen Einschränkungen" haben wir bei der Position "kaum" geprüft, ob das Antwortverhalten der Analysepersonen von der Stufenzahl beeinflusst wird. Auch hier können wir eine Abnahme der Fehlbesetzung mit zunehmendem Differenzierungsgrad des Antwortmodells beobachten ( $\chi^2 = 9,3; p = 0,0055$ ).

Die Mobilitätsanalyse beim Übergang von Stufe zum Antwortmodell ergab, daß neben der konstanten Antwort (50,4 %) bevorzugt die Stufe "teilweise" signalisiert wird.

Auf die verbleibenden 35,4 % der Wähler (Wir müssen bei der Interpretation zu Tabelle 11 den Versuchsfehler durch Korrekturen von Definitionen. Wir werden Werte unter 10 % wegen der geringen n hier nicht korrigieren).

Bei der Umkehrung der Transformation, also von Antwortmodell B nach AM A, orientiert sich die Wähler ein hochgradig konstant (95,6 %).

Beim Übergang von Antwortmodell A zum Antwortmodell B spalten sich die Wähler zur Stufe "kaum" in wesentliches in zwei Gruppen auf. 25,6 % wechseln zur Stufe "ungefähr", 17,1 % zur Stufe "mit gewissen Einschränkungen" und 48,6 % bleiben konstant. Die Mobilität ist beim Übergang von Antwortmodell A nach AM B größer als wir erwarteten (vgl. Tab. 5).

In der Transformation von AM C nach AM A bleiben zwei Drittel der Wähler konstant und ein weiteres Viertel geht zur Stufe "mit gewissen Einschränkungen" über. Dieses Ergebnis ist ein Hinweis darauf, daß die Indikatorgrenzen der einzelnen Stufen in den verschiedenen Antwortmodellen großen Schaulücken unterliegen.

Die Ergebnisse beim Übergang von AM B nach AM C sind weitgehend numerisch fast gleichlautend (s. Tabelle 11).

Wie haben aber eine andere inhaltliche Bedeutung. Beim Übergang zum höherrangigen AM gehen 40 % der Stufe "kaum" zur Stufe "ungefähr" über. Der Wechsel zum niederrangigen AM

Führt dazu, daß 40 % der Wähler in die Stufe "teilweise" eingehen.

Bisher haben wir die allen Antwortmetallen gemeinsamen Stufen "überhaupt nicht", "kaum", "mit gewissen Einschränkungen" und "vollkommen" betrachtet. Nun wenden wir uns den speziellen Stufen der AM B und C zu.

Zunächst soll uns interessieren, wie sich die Wähler in der Stufe "teilweise" ändern, wenn die Stufenzahl des Antwortmodells wechselt.

Tabelle 12: Veränderungen in der Stufe "teilweise" beim Übergang in ein andersstufiges Antwortmodell

Transformation	(in anderem AM gewählte Stufe)					
	1	2	3	4	5	6
AM B → AM A	2,2	32,3	56,0	4,6	-	-
AM B → AM C	0,0	16,2	15,2	40,7	21,5	5,3

Beim Übergang des fünfstufigen Antwortmodells in ein vierstufiges verteilen sich die Wähler ersatzweise auf die Stufen "kaum" (32,3 %) und "mit gewissen Einschränkungen" (56,0 %). Sehr breit ist dagegen die Streuung der Stufe "teilweise", wenn wir in ein sechsstufiges Antwortmodell überspringen. Hier entfallen bedeutende Häufigkeiten auf die Stufen 2 bis 5 des AM C, wobei der Schwerpunkt auf der Stufe 4 "mit gewissen Einschränkungen" liegt (40,7 %).

Tabelle 13: Veränderungen in der Stufe "in großen und ganzen" beim Übergang in ein andersstufiges Antwortmodell

Transformation	(in anderem Antwortmodell gewählte Stufe)					
	1	2	3	4	5	6
AM C → AM A	0,0	1,2	76,1	20,1	-	-
AM C → AM B	0,0	1,2	4,1	71,1	21,6	-

Die Transformation der Stufe "in großen und ganzen" führt sowohl beim Übergang in ein vier- als auch in ein fünfstufiges Antwortmodell zu recht eindeutigen Befunden. In wesentlichen geht diese Stufe in die Position "mit gewissen Einschränkungen" ein (78,1 % bzw. 71,1 %). Ein weiteres Fünftel der Wahlentscheidungen geht im niederstufigen Antwortmodell in die Stufe "vollkommen" über. Auch dieses Ergebnis weist darauf hin, daß der Indikatorbereich der Stufe "vollkommen" im vier- und fünfstufigen AM größer als im sechststufigen ist.

Tabelle 14: Veränderungen in der Stufe "ungefähr meine Meinung" beim Übergang in ein vierstufiges Antwortmodell

Transformation	(in anderen Antwortmodell gewählte Stufe)				
	1	2	3	4	5
AM C → AM A	2,9	45,7	48,5	2,9	-
AM C → AM B	3,1	37,6	25,0	34,4	0,0

Tabelle 14 veranschaulicht, daß sich die Stufe "ungefähr meine Meinung" im vierstufigen Antwortmodell etwa zu gleichen Teilen (45,7 % zu 48,5 %) auf die Stufen zwei und drei verteilen. Im fünfstufigen Antwortmodell findet eine Dreiteilung der Wahlentscheidungen statt. Hier gehen die Antworten auf die Stufen "kaum", "teilweise" und "mit gewissen Einschränkungen" über.

Die Betrachtungen der Transformationsprozesse beim Wechsel der Stufenzahl von Antwortmodellen ließ vielfache Beziehungen zwischen den Stufen hervortreten, die teilweise mit unseren Erwartungen über die Transformation nicht übereinstimmen. Wir wollen deshalb das empirische Ergebnis systematisieren. Der Anschaulichkeit halber wählen wir wiederum die Form einer Abbildung. Abbildung 6 stellt das empirische Paradigma zu Abbildung 5 dar.

Abt. 6: Feststellungen in den Substanzproben, falls Übergang von einem nichtstrafbaren Substanztyp zu ein höherstrafbares und umgekehrt (empirisches Ergebnis)

1. Übergang von AM A nach AM B

vollkommen	→	vollkommen
mit gewissen Hinweism.	→	mit gewissen Hinweism.
keine	→	keine
überhaupt nicht	→	überhaupt nicht

2. Übergang von AM A nach AM D

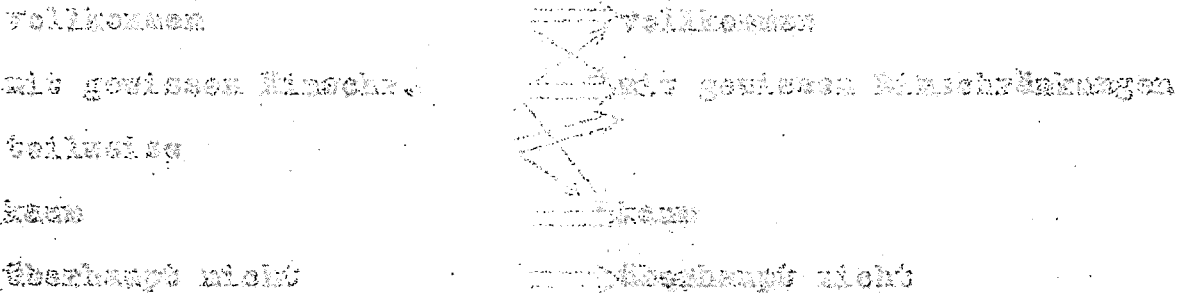
vollkommen	→	vollkommen
mit gewissen Hinweism.	→	in großen und gewissen mit gewissen Hinweism.
keine	→	ungefähr
überhaupt nicht	→	keine
	→	überhaupt nicht

3. Übergang von AM B nach AM D

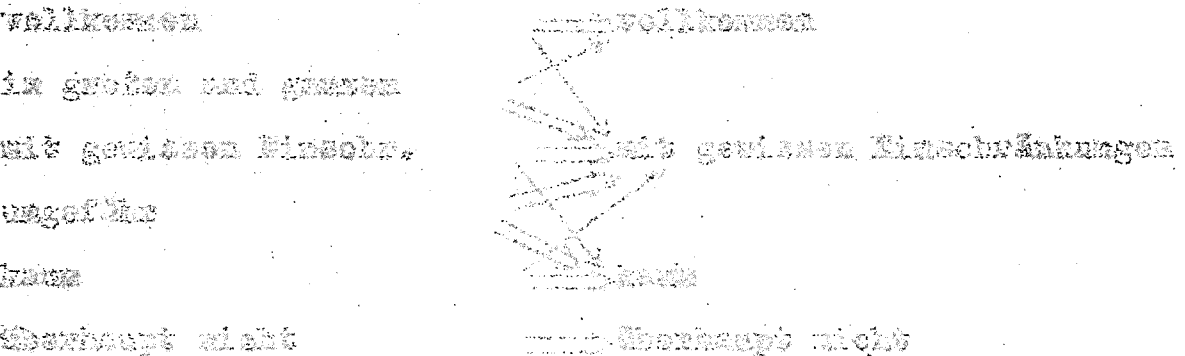
vollkommen	→	vollkommen
mit gewissen Hinweism.	→	in großen und gewissen mit gewissen Hinweism.
teilweise	→	ungefähr
keine	→	keine
überhaupt nicht	→	überhaupt nicht

Abb. 5 (Fortsetzung)

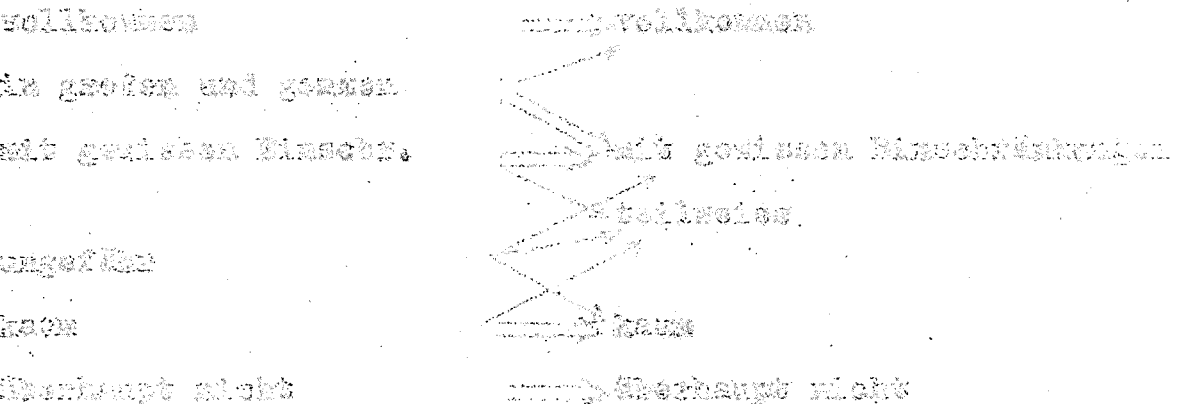
2. Übergang von AM 3 nach AM A



3. Übergang von AM 3 nach AM A



4. Übergang von AM 3 nach AM A



Sie verdeutlicht, daß die Übergänge zwischen den Indikatorbereichsbereichen aller Stufen bei allen Antwortmodellen fließend sind. (Über die Stufe "Überhaupt nicht" konnten wir, siehe oben, keine Aussage treffen. Dieser Sachverhalt ist durch den gestrichelten Pfeil angedeutet.)

Eindeutige Zuordnungen von Stufen verschiedenstufiger Antwortmodelle sind nicht möglich. Selbst der Indikatorbereich der Extremposition "vollkommen" ist in den einzelnen AM nicht gleichbedeutend.

Im allgemeinen entspricht eine Stufe eines beliebigen Antwortmodells mehreren Stufen in einem anderen AM, wobei in der Regel eine Stufe am deutlichsten korrespondiert.

Wollen wir Ergebnisse aus verschiedenstufigen Antwortmodellen, die zu einem Item vorliegen, vergleichen, lassen sich aus der Transformationsanalyse folgende Empfehlungen ableiten, wobei es grundsätzlich nur möglich ist, ein höherstufiges AM in ein niederstufiges zu transformieren:

1. Vergleich der Ergebnisse aus fünf- und vierstufigen AM

Hier ergeben sich zwei Reduktionsmöglichkeiten:

a) Rückführung der Ergebnisse von fünfstufigen in ein vierstufiges AM.

Bei der Transformation ist es erforderlich, die Stufe der Stufe "teilweise" auf die Stufen "mit gewissen Einschränkungen" und "kaum" aufzuteilen. Nach unseren Ergebnissen müßten etwa zwei Drittel der Apr. der Stufe "mit gewissen Einschränkungen" zugeordnet werden.

Will man die Aufteilung der Stufe "teilweise" umgehen, lassen sich

b) beide Modelle auf ein dreistufiges AM reduzieren. Es werden dann nur die Extrempositionen in alter Form beibehalten, alle Zwischenstufen werden zu einer neuen Position "eingeschränkte Zustimmung" vereinigt.

2. Vergleich der Ergebnisse aus sechs- und vierstufigen AM

Auch hier können wir Reduktion (b) vornehmen. Folien wie dagegen das sechststufige AM auf ein vierstufiges AM durchfüh-

ren, ist die Zusammenfassung der Stufen "in großen und ganzen" sowie "mit gewissen Einschränkungen" des sechsstufigen AM empirisch am besten gesichert. Sie bereitet keine Schwierigkeiten. Problematisch ist dagegen die Stufe "ungefähr". Sie müßte zu gleichen Teilen den Stufen "kaum" und "mit gewissen Einschränkungen" zugeordnet werden.

### 3. Vergleich der Ergebnisse aus sechs- und fünfstufigen AM

Die Reduktion des sechsstufigen AM auf das fünfstufige bereitet enorme Schwierigkeiten, weil zu viele Aufteilungen von Zwischenstufen (vgl. Abb. 6) vorgenommen werden müßten. Neben der schon oben beschriebenen Reduktion auf ein dreistufiges AM wäre die Rückführung beider Antwortmodelle auf vier Stufen zu empfehlen. Beim sechsstufigen AM werden (a) die Stufen "in großen und ganzen" und "mit gewissen Einschränkungen" und (b) die Stufen "ungefähr" sowie "kaum" zusammengefaßt. Beim fünfstufigen AM ziehen wir die Stufen "teilweise" und "kaum" zusammen.

### 4.1.4. Trennschärfenanalyse einzelner Stufen der AM und deren Zusammenfassungen

In Abschnitt 4.1.2. hatten wir die Trennschärfe der Antwortmodelle unter dem Aspekt von response sets betrachtet. Wir konnten dort nachweisen, daß vier-, fünf- und sechsstufige Antwortmodelle hinsichtlich ihrer Differenzierungsfähigkeit gleichwertig sind. Bei dieser Analyse blieb offen, ob die einzelnen Stufen der Antwortmodelle hinreichend diskriminieren, und welche Zusammenfassungen von Antwortstufen bei der Interpretation von Befragungsergebnissen berechtigt sind. Dieser Problematik soll nun für die Antwortmodelle A, B und C nachgegangen werden.



- a) Tabelle 15 gibt über die Diskriminationsfähigkeit der einzelnen Stufen und deren Zusammenfassungen beim vierstufigen Antwortmodell Auskunft. Die Trennschärfenkoeffizienten in Tabelle 15 verstehen sich als punktblissziale Korrelationskoeffizienten.

Tabelle 15: Trennschärfenkoeffizienten der Stufen und deren Zusammenfassungen im vierstufigen Antwortmodell

Item/Stufe	4/3,2,1	3/2,1	2/1	4,3/2,1	4,3,2/1
a	0,62	0,29	0,21	0,45	0,13
b	0,70	0,26	0,15	0,43	0,09
c	0,52	0,13	0,06	0,26	0,00
d	0,59	0,41	0,27	0,52	0,48
e	0,55	0,29	0,23	0,41	0,18
f	0,77	0,19	0,14	0,29	0,09
$\bar{r}_{1it}$	0,62	0,26	0,18	0,39	0,16

Zunächst interessieren die  $r_{1it}$ -Werte für die einzelnen Stufen dieses AM. Die Stufe "vollkommen" (4/3,2,1) erweist sich bei allen Items gegenüber den restlichen Entscheidungsalternativen als konsistent und trennscharf ( $\bar{r}_{1it} = 0,62$ ). Für die inhaltliche Interpretation können wir festhalten, daß Befragte, die die Stufe "vollkommen" des AM wählen, gegenüber Befragten, die die Stufen "mit gewissen Einschränkungen", "kaum" und "überhaupt nicht" anwählen, eine positivere Meinung (Einstellung) zum Befragungsgegenstand haben.

Diese Aussage, die vielleicht selbstverständlich erscheint, können wir für die Stufe "mit gewissen Einschränkungen" (3/2,1) nicht mehr aufrechterhalten. Beim Item e und f ( $r_{1it} = 0,13$  bzw. 0,19) ist die Wahl der Stufe "mit gewissen Einschränkungen" kein Ausdruck einer positiveren

Meinung als die Entscheidung für die Alternativen "kaum" und "überhaupt nicht". Insgesamt geht die Trennschärfe von Stufe (3) gegen den verbalen Urteilen 2 und 1 beträchtlich zurück. Der mittlere  $\kappa_{12}$ -Wert liegt bei 0,26. Dieser Wert liegt an der unteren Grenze des Signifikanzniveaus.

Die Diskriminationsfähigkeit zwischen den Stufen "kaum" (2/1) und "überhaupt nicht" fällt im Vergleich zu (4/3,2,1) und (3/2,1) noch weiter ab. Nur für die Items e, d und a lassen sich noch signifikante  $\kappa_{12}$ -Werte ermitteln.

Das Ergebnis der Diskriminationsanalyse einzelner Stufen des AM (A) verdeutlicht, daß die a priori unterstellte differenzierte Abbildung der Meinungsäußerung nicht generell, d.h. für alle Fragegegenstände gegeben ist.

Mit hoher Wahrscheinlichkeit läßt sich auf Grund unserer Analyse für die Interpretation von Meinungsumfragen nur die Aussage treffen, daß Befragte, die die Stufe "vollkommen" wählen, eine positivere Meinung (Einstellung) zum Fragegegenstand haben als alle anderen Probanden. Eine differenziertere Abbildung der Meinung bedarf immer erst der empirischen Trennschärfebestimmung.

Die Zusammenfassung der Stufen "vollkommen" und "mit gewissen Einschränkungen" ist bei der Interpretation von Befragungsergebnissen eine vielfach geübte Praxis. In Tabelle 15 sind die entsprechenden Trennschärfeeffizienten in der Spalte (4,3/2,1) eingetragen. Alle Koeffizienten sind signifikant, so daß die Interpretation berechtigt erscheint. Allerdings können wir aus den vorhergehenden Ergebnissen folgern, daß bei einzelnen Items eine Scheintrennschärfe vorliegt. Die Zusammenfassungen bei den Items e und f ergeben nur durch den Einfluß der Stufe "vollkommen" signifikante Trennschärfe.

Stellen wir die Trennschärfeeffizienten der Spalten (4/3,2,1) und (4,3/2,1) gegenüber, so wird ersichtlich, daß durch die Zusammenfassung der beiden Stufen die Abbildgenauigkeit zurückgeht. Der mittlere  $\kappa_{12}$ -Wert sinkt

von 0,62 auf 0,39 ab.

Die verallgemeinernde Aussage, daß es sich bei den Befragten, die die Stufen 4/3) signieren, um positiv eingestellte Probanden handelt, läßt sich also empirisch nicht durchgängig bestätigen. Auch ist die Zusammenfassung der Befragten, die die Stufen 4,3 und 2 wählten, zu einer Analysegruppe kaum möglich. Nur in einem Fall (Item d) ergab die Analyse einen signifikanten Koeffizientenz.

b) Tabelle 16 ist zu entnehmen, daß die Ergebnisse der Trennschärfenanalyse zum fünfstufigen Antwortmodell den vorhergehenden sehr ähnlich sind.

Auch hier erweist sich die Stufe "vollkommen" (5) gegenüber allen anderen Stufen des AM (4,3,2,1) als durchgängig trennscharf ( $r_{1t} = 0,52$ ). Die Diskriminationsfähigkeit der Stufe "mit gewissen Einschränkungen" (4/3,2,1) ist mit  $r_{1t} = 0,24$  bedeutend geringer und nicht für alle Items gegeben. Dieser für die Interpretation ungenutzte Trend setzt sich bei den Stufen "teilweise" (3/2,1) und "kaum" (2/1) fort ( $r_{1t} = 0,16$  bzw.  $0,06$ ).

Tabelle 16: Trennschärfenkoeffizienten der Stufen und deren Zusammenfassungen im fünfstufigen AM

Item/ Stufe	5/4,3,2,1	4/3,2,1	3/2,1	2/1	5,4/3,2,1	5,4,3/2,1	5,3,2,1
a	0,55	0,30	0,22	0,00	0,45	0,29	0,18
b	0,51	0,07	0,00	0,00	0,22	0,00	0,00
c	0,54	0,24	0,13	0,06	0,46	0,27	0,10
d	0,50	0,31	0,22	0,11	0,49	0,44	0,39
e	0,42	0,29	0,20	0,18	0,39	0,34	0,16
f	0,60	0,21	0,20	0,00	0,31	0,26	0,10
$\bar{r}_{1t}$	0,52	0,24	0,16	0,06	0,39	0,27	0,08

Demit bleiben auch die Zusammenfassungen der Stufen 5 und 4 bzw. 5,4 und 3 bei den Interpretationen für dieses AM problematisch. In einigen Fällen entsteht wiederum eine Scheitersensibilität (vgl. Tabelle 16).

Die Zusammenfassung der Stufen 5,4,3 und 2 führte in vorliegender Analyse zu keinem einzigen signifikanten Koeffizienten. Diese Form der Datenauswertung sollte deshalb vermieden werden.

Für die Interpretation der Ergebnisse empfiehlt sich auch beim fünfstufigen Antwortmodell die Stufe "vollkommen" als positiver Ausdruck der Meinungsäußerung den anderen Stufen des AM entgegenzustellen.

Die Zusammenfassung der Stufen "vollkommen" und "mit gewissen Einschränkungen" führt zu einer erheblichen Abnahme der Indikaterschärfe.

e) Die Ergebnisse zum sechsstufigen Antwortmodell (vgl. Tabelle 17) weichen in einigen Passagen von den vorhergehenden ab. So finden wir durchgängig eine hinreichende Differenzierungsfähigkeit der Stufe "im großen und ganzen".

Tabelle 17: Trennschärfenkoeffizienten der Stufen und deren Zusammenfassungen im sechsstufigen Antwortmodell

Item/ Stufe	6/5,4,3,2,1	5/4,3,2,1	4/3,2,1	3/2,1	2/1
a	0,49	0,52	0,29	0,00	0,19
b	0,49	0,49	0,26	0,00	0,19
c	0,50	0,53	0,13	0,00	0,19
d	0,50	0,45	0,16	0,00	0,00
e	0,50	0,36	0,09	0,00	0,26
f	0,48	0,45	0,19	0,00	0,13
$\bar{F}_{it}$	0,49	0,47	0,19	0,00	0,16

Tabelle 17 (Fortsetzung)

Item/ Stufe	6,5/4,3,2,1	6,5,3/3,2,1	6,5,4,3/2,1	6,5,4,3,2/1
a	0,56	0,42	0,36	0,33
b	0,63	0,48	0,38	0,37
c	0,63	0,45	0,34	0,33
d	0,61	0,38	0,28	0,34
e	0,35	0,25	0,25	0,29
f	0,66	0,48	0,48	0,33
$\bar{T}_{1,2}$	0,57	0,38	0,31	0,33

(5/4,3,2,1). Das Mittel der Transschärfenkoeffizienten liegt für diese Stufe bei 0,47 und ist damit nur geringfügig niedriger als für die Stufe "vollkommen" ( $\bar{T}_{1,2} = 0,49$ ).

In der nächstfolgenden Stufe des AM finden wir dagegen nur noch zwei signifikante Koeffizienten. Auch zeigt das Mittel der Transschärfenkoeffizienten für die Stufe "mit gewissen Einschränkungen" eine deutliche Abnahme (über Transschärfe) ( $\bar{T}_{1,2} = 0,39$ ). Der beobachtete monotone Rückgang in den Transschärfenkoeffizienten (s. Tabellen 15, 16, 17) wird in der Stufe "ungefähr" (3/2,1) und "kaum" durchbrochen. In der Stufe "ungefähr" sinkt die Transschärfe für alle Items ein wenig ab, während sie bei "kaum" wieder zur durchschnittlichen 0,36 ansteigt.

Es wird deutlich, daß das sechststufige Antwortmodell im unteren Bereich entscheidende Konstruktionsanalogie aufweist.

Die Zusammenfassung der Stufen (6,5) des AM C ist nach unserem Ergebnis berechtigt. Jede dieser Stufen verhält sich nämlich als Transschärfe, so daß auch für die Kombination dieser situationsfähigkeit zu erwarten ist. Die empirische Analyse ergibt sogar (vgl. Tabelle 17), daß die Transschärfenkoeffizienten für die Kombination (6,5) etwas höher als für jede

Stufe ausfallen ( $\bar{r}_{it} = 0,57$  gegenüber 0,40 bzw. 0,47). Die inhaltliche Interpretation kann somit die Befragten, die die Stufen 6 und 5 signierten, zu einer Gruppe zusammenfassen.

Obwohl wir auch bei der Kombination (6,5,4) durchweg signifikante Trennschärfekoeffizienten vorfinden, ist die Zusammenfassung aber nur für die Items a und b berechtigt. Alle anderen Koeffizienten müssen wir als Scheintrennschärfe charakterisieren, weil - wie in Tabelle 17 nachweisbar - Stufe (4) nicht generell zwischen positiv und negativ eingestellten Befragten diskriminiert.

Ebenso müssen die Koeffizienten in der Spalte (6,5,4,5/3,3) als Scheintrennschärfe aufgefaßt werden, da die Stufe (3) nicht in einem einzigen Fall methodisch qualifiziert war.

Die Analyse der Diskriminationsfähigkeit der Antwortmodellstufen und ihrer Zusammenfassungen führte zu folgenden Ergebnissen:

1. Eine sehr differenzierte Interpretation der Befragungsergebnisse kann methodenkritisch nur wenig abgesichert werden.

Beim vier- und fünfstufigen Antwortmodell kann im allgemeinen die Stufe "vollkommen" als Ausdruck positiver Meinung oder Einstellung gegenüber dem Befragungsgegenstand betrachtet werden. Für die Stufen "mit gewissen Einschränkungen" und "teilweise" ist dagegen eine generelle Aussage anhand unseres Materials nicht möglich. Für die differenzierte Interpretation muß von Fall zu Fall ein empirischer Nachweis über Diskriminationsfähigkeit der Antwortstufen erbracht werden.

Beim sechststufigen Antwortmodell ließ sich für die Stufen "vollkommen" und "mit gewissen Einschränkungen" die Trennschärfe durchgängig sichern. Alle anderen Stufen bedürften ebenfalls einer speziellen empirischen Trennschärfenbe-

stimmung. Die Analyse dieses Antwortmodells läßt erkennen, daß die unteren verbalen Antwortstufen noch nicht einen praktischen Einsatz des Modells gestatten.

2. Die durchgängig geübte Praxis der Zusammenfassung von Stufen einzelner Antwortmodelle in der Interpretationsphase ist aus methodenkritischer Sicht kaum gerechtfertigt, da die Trennschärfe der einzelnen Stufen nicht generell erwiesen ist. Mit Ausnahme der Kombination (4,5) ist bei den Zusammenfassungen der AM-Stufen ein starker Abfall der Indikatorschärfe zu beobachten. Dergleichen konnte wiederholt eine Sensitivitätsanalyse dieser Kombinationen aufgezeigt werden.

#### 4.1.5. Abhängigkeit der Indikatorfunktion vom Antwortmodell

Wir werden jetzt der Frage nachgehen, ob sich durch Veränderung eines antwortstufigen Antwortmodells die Indikatorfunktion der Items ändert. Hierzu prüfen wir, ob die Items bei antwortstufigen Antwortmodellen in unterschiedlichen Faktoren laden.

Die Faktorenanalyse der Interkorrelationen ergab für jedes Antwortmodell drei Faktoren. In Tabelle 18 sind die Ladungen in den unrotierten Faktoren zum Antwortmodell A, B und C zusammengestellt.

Die Korrelationen der Items mit dem ersten Faktor zeigen für alle Antwortmodelle ein gut übereinstimmendes Ladungsmuster. Die Rangkorrelationen zwischen AM A und B betragen 0,77, zwischen AM A und C 0,77 und AM B und C 0,88. Auf Grund dieser Werte und der Richtungskonstanz der Faktorladungen können wir die Ähnlichkeit zwischen dem ersten Faktor behaupten.

Anzumerken ist, daß alle Faktorladungen in jeweiligen Faktor 1 signifikant sind, wobei Item 4 mit diesem Faktor am niedrigsten korreliert.

Tabella 18: Gegenüberstellung der unrotierten Faktoren der vier-, fünf- und sechstufigen Antwortmodelle

Faktor 1

Item	$4^{\text{F}1}$	$5^{\text{F}1}$	$6^{\text{F}1}$
a	-0,55	-0,66	-0,53
b	-0,76	-0,72	-0,54
c	-0,52	-0,63	-0,60
d	-0,70	-0,73	-0,62
e	-0,26	-0,29	-0,24
f	-0,50	-0,68	-0,74

Faktor 2

Item	$4^{\text{F}2}$	$5^{\text{F}2}$	$6^{\text{F}2}$
a	0,24	0,24	0,36
b	-0,06	-0,06	-0,06
c	-0,42	-0,30	-0,22
d	-0,10	-0,16	-0,21
e	0,51	0,54	0,52
f	0,17	0,05	0,06

Faktor 3

Item	$4^{\text{F}3}$	$5^{\text{F}3}$	$6^{\text{F}3}$
a	0,22	0,16	0,23
b	-0,32	0,23	-0,23
c	0,14	0,18	0,23
d	-0,11	-0,15	-0,20
e	0,07	0,10	-0,16
f	-0,25	-0,30	0,11



Der zweite unrotierte Faktor ist jeweils bipolar. Item a besitzt hier die höchsten positiven signifikanten Ladungen. Der negative Pol des Faktors wird in allen drei Fällen durch Item c markiert. Die Übereinstimmung der Rangkorrelationen der Faktorladungen ist für diesen Faktor vollkommen. Alle möglichen Korrelationen sind 1,00.

Der Vergleich der Faktorladungen im dritten unrotierten Faktor fällt dagegen weniger günstig aus. Auch diese Faktoren sind für die AM A, B und C bipolar. Unterschiede treten aber in den Markieritems der Faktoren auf. So wird beim Antwortmodell A der Faktor durch Item a und f charakterisiert, beim AM B durch Item b und f und beim AM C durch Item e und f. Die Korrelationen zwischen den Faktoren fallen entsprechend niedrig aus. Zwischen AM A und B, AM A und C sowie AM B und AM C erhalten wir Werte von nur 0,60, 0,45 und -0,03.

Das Ergebnis der Faktorenanalyse läßt vermuten, daß die ersten beiden Faktoren jeweils durch den Iteminhalt determiniert werden, der letzte Faktor aber zu Lasten der Unterschiede in den Antwortmodellen geht.

Damit dürfte die spezifische Struktur der Antwortmodelle die Indikatorfunktion der Items mitbestimmen. Würden wir eine inhaltliche Interpretation der Faktoren ausführen, so würden wir auf Grund der AM zu leicht divergierende Aussagen kommen. Die Unterschiede im dritten Faktor lassen sich nicht aus Differenzen zwischen den Analysestichproben erklären, da das Material an korrelierenden Stichproben bei balancierter Darbietungsfolge der AM gewonnen wurde. Der Anteil der modellspezifischen Varianz an der aufgeklärten gemeinsamen Varianz liegt zwischen 5,5 und 7,7 Prozent.

#### 4.1.6. Quantitative Äquivalenz verbaler Urteile

Oft wird die Frage diskutiert, mit wieviel Prozent die Analysepersonen dem Iteminhalt zustimmen, wenn sie die Stufen "vollkommen", "mit gewissen Einschränkungen" usw. wählen.

Bisher ist nur wenig bekannt wie das subjektive quantitative Äquivalent zu den spezifischen verbalen Urteilen eines AP, also zu seinen Stufen, aussieht. In FB-72 haben wir auf zwei unterschiedlichen Wegen eine Klärung gesucht:

1. Zuordnung von Prozentzahlen zu den verbalen Urteilen.

In der Fragebogenvariante FB 72-1 nahmen die Aps. zum sechsstufigen AM, in FB 72-2 zum vierstufigen AM und in FB 72-3 zum fünfstufigen AM Stellung. Die Probanden bearbeiteten immer das Antwortmodell, das ihnen zuletzt im Kontext dargeboten wurde. Auf diese Weise wollten wir erreichen, daß der Bezug zum Fragegegenstand erhalten blieb.

2. Paarvergleich verbaler Urteile.

In einem unvollständigen Paarvergleich wurden den Befragten nachstehende verbale quantitative Urteile vorgelegt: "überhaupt nicht", "recht schwach", "kaum", "beiläufig", "in etwa", "nicht ganz", "ungefähr", "mit gewissen Einschränkungen", "im großen und ganzen", "vollkommen". Mit Ausnahme der Extremurteile wurden alle möglichen paarweisen Kombinationen aus den verbalen quantitativen Urteilen gebildet. Die Befragten hatten jeweils zu entscheiden, welche der beiden Alternativen einen höheren Grad an Zustimmung zum Ausdruck bringt. 6)

Zunächst werden wir auf die Zuordnung von Prozentzahlen zu verbalen quantitativen Urteilen eingehen.

6) Anmerkung zum unvollständigen Paarvergleich:  
Die Auswertung des hier vorliegenden Materials und anderer Studien zu dieser Problematik hat gezeigt, daß es auch notwendig ist, alle möglichen Kombinationen der extremen Urteile den Aps. vorzulegen, da diese nicht, wie von uns üblich üblich angenommen, absolute Grade an Zustimmung bzw. Ablehnung darstellen. Die unvollständige Darstellung erfordert Extrapolationen, die mit erheblichen Verzerrungsrisiken verbunden sind, dessen Verringerung den Mehraufwand an Fragebogenlänge und Arbeitszeit rechtfertigt.

Tabelle 19: Zuordnung von Prozentzahlen zu verbalen quantitativen Urteilen in den Antwortmodellen A, B, C

verbalen Urteil		AM A	AM B	AM C
vollkommen	$\bar{H}$	80,85	75,20	80,86
	s	(27,70)	(31,00)	(30,60)
im großen ...	$\bar{H}$	-	-	66,88
	s	-	-	(29,20)
mit gewissen ...	$\bar{H}$	60,70	57,09	57,26
	s	(24,08)	(19,90)	(25,90)
teilweise	$\bar{H}$	-	36,38	-
	s	-	(26,50)	-
ungefähr	$\bar{H}$	-	-	39,57
	s	-	-	(23,70)
kaum	$\bar{H}$	25,07	19,81	19,75
	s	(17,50)	(16,40)	(15,80)
überhaupt nicht	$\bar{H}$	11,78	15,51	13,24
	s	(11,80)	(19,60)	(26,70)

Die mittlere Zustimmungswerte zur Stufe "vollkommen" (vgl. Tabelle 19) beträgt in keinem der hier analysierten Antwortmodelle 100 %. Dieses Ergebnis sollte bei der Interpretation von Umfrageergebnissen beachtet werden. Die bedeutende Differenz zwischen a priori Prozentwert und empirischen Prozentwert ist kein Zufallsergebnis, sondern ein empirischer Befund, der Beachtung verdient.

Gegen die Annahme eines Zufallsergebnisses sprechen folgende Argumente:

1. Die Ergebnisse stimmen in allen Teiltischproben hinreichend überein.
2. Nur insgesamt 52 % der Befragten (n = 369) ordnen der Stufe "vollkommen" den Wert 100 zu.

Bei der Interpretation von Befragungsergebnissen müssen wir nach unserem Erfund davon ausgehen, daß die Stufe "vollkommen" nicht mit absoluter Zustimmung gleichgesetzt werden darf.

Dieses wird an Tabelle 19 deutlich, daß auch die obichmer-ke verzele Urteil "Wachst nicht" kein absolutes Wert ist, sondern im Durchschnitt noch einen gewissen Zustimmungsgewinn ausdrückt.

Die gemeinsamen Zeichenstufen aller Antwortskala (mit gewissen Einschränkungen, "Wachst") erhalten in allen Teilerhebungen im Durchschnitt annähernd gleiche prozentuale Äquivalente zugeordnet. Während die Befragten die Stufe "mit gewissen Einschränkungen" so stützen wie mit etwa 65 % der Aussagekraft und bei der Stufe "Wachst" nur etwa 20 % an.

Für die Antwortskala 3 ergibt die direkte Auswertung von Prozentzahlen, daß die Stufe "teilweise" nicht wie erwartet 30 % an Zustimmung, sondern ca. 10 % ausdrückt. Im Antwortskala 4 bedeutet die Stufe "in großen und kleinen" etwa 67 % Zustimmung. Im gleichem Maß steht hinter der Antwortskala für die Stufe "ungelöst" eine durchschnittliche Zustimmung von ca. 40 Prozent.

Fassen wir die Ergebnisse der bisherigen Analyse zusammen, so ergibt sich:

1. Die analysierten Antwortskala werden durch verbale quantitative Urteile kodifiziert, die den oberstellte Kontinuum nachherl repräsentieren.
2. Zwischen den verbalen quantitativen Urteilen besteht keine Äquivalenz.
3. Die bei der Datenauswertung den verbalen Urteilen zugeordneten numerischen Äquivalente haben das Niveau von Prozentzahlen.

Sow die Befragungsergebnisse werden aber häufig Funktionen herangezogen, so t-Test, Korrelationskoeffizient, Varianz- und Faktorenanalyse, die den statistischen Aussagen im Hinblick auf die Ergebnisse, es ist also empfehlenswert, das statistische Verfahren von 10 zu verwenden, das die Möglichkeiten der statistischen Analyse voll ausgeschöpft werden können.

Dabei gibt es zwei Möglichkeiten:

1. Umformung der Rangdaten in eine Intervallskala durch Stichprobentransformationen,
2. Korrelationskoeffizienten von Intervallskalen zu den verfügbaren verbalen Urteilen,

Wir wollen den zuletzt genannten Ansatz in Anwendung bringen. Als Ausgangspunkt dienen uns die Ergebnisse des Paarvergleichs (s. Tabelle 19), aus dieser Tabelle können wir ersehen, in welchem prozentualen Verhältnis die Befragten bei den einzelnen Alternativen einen höheren Bestimmungsgrad signalisieren. In der letzten Zeile der Tabelle sind analog zu Tabelle 17 die mittleren prozentuale Bestimmung vermerkt,

Tabelle 20: Korrelationsmatrix der verbalen quantitativen Urteile (N = 72-1 bis 72-9, n = 605)

verbalen Urteil	1	2	3	4	5	6	7	8	9	10
1 überhaupt nicht	xx	69	72	75	78	85	89	95	98	99
2 recht schlecht	36	xx	66	69	72	78	84	88	90	98
3 kaum	24	34	xx	63	76	61	63	63	65	67
4 teilweise	5	3	15	xx	53	66	55	60	60	61
5 in etwa	12	12	24	47	xx	53	61	63	63	62
6 nicht ganz	6	12	13	40	47	xx	53	66	66	74
7 ungefähr	7	15	17	45	39	45	xx	58	61	60
8 mit gewissen	3	12	17	28	39	32	42	xx	63	61
9 in gewissem	2	12	15	26	17	34	19	37	xx	63
10 vollkommen	1	2	3	4	3	24	10	13	38	xx
$\bar{x}$	12	22	21	30	31	37	36	41	46	66

Beim vierstufigen Antwortmodell ergeben Paarvergleiche und das direkte Auswahungsverfahren nahezu identische Werte (26:21; 21:21; 25:25; 10:11) <sup>7)</sup>.

Konkret günstig ist die Anpassung der geschätzten Äquivalente im AN B (25:75; 31:37; 50:35; 22:1) und 10:15. Insbesondere differieren die numerischen Äquivalente in der Stufe "Realisier". Während sich beim direkten Verfahren für die Alternative eine Subjektbewertung von 30 ergibt, erhalten wir beim Paarvergleich 50 Prozent. Der auf individueller Wege gewonnene Wert korrespondiert also mit unseren Erwartungen über das numerische Äquivalent der Stufe "Teilweise" weit besser.

Für das sechsstufige Antwortmodell ergeben sich folgende Proportionen: 55:31; 75:57; 61:37; 56:40; 25:22 und 10:13. Die größte Differenz tritt bei der Stufe "Vollständig" auf. Auch für Antwortmodell C ist die Anpassung der numerischen Äquivalente geringer als beim vierstufigen Antwortmodell.

Nach Anwendung des "Flow of Compensating Income" auf die Ururteile der Igu. erhalten wir für die subjektive äquivalente psychologische Einkommenswerte, die die subjektiven Nutzenunterschiede zwischen den Stufen der Igu kennzeichnen:

AN A:

Überhaupt nicht / kaum / mit gewissen Einschränk. / vollkommen

0,00	0,50	1,50	2,33
------	------	------	------

AN B:

Überhaupt nicht / kaum / teilweise / mit gew. B. / vollkommen

0,00	0,30	1,30	1,80	2,75
------	------	------	------	------

AN C:

Überhaupt nicht / kaum / ungefähr / mit gew. B. / im groß. / vollst.

0,00	0,30	1,30	1,70	2,01	2,70
------	------	------	------	------	------

7) Die erste Spalte der jeweiligen Proportien resultiert aus dem Paarvergleich, die zweite aus dem direkten Auswahungsverfahren.

Die Skalenwerte spiegeln die tatsächlichen Unterschiede zwischen den verbalen Stufen der Antwortmodelle wider. Insbesondere korreliert die größte Distanz zwischen den Stufen "traum" und "mit gelassenen Hirschdrückungen" im AM A auch mit dem Skalenwertverhältnis zum Ausdruck. Ebenso ist zu erwarten, daß die Distanz zwischen den beiden positiven bzw. negativen Stufen des AM geringer ist.

Die hier aufgefundenen Skalenwerte lassen sich linear so transformieren, daß sie praktisch besser handhabbar sind. Allerdings lassen sich Befragtenablen auch dann nicht vermeiden, da die Variabilität der Distanzen durch die verbalen Urteile determiniert ist.

Wir haben ferner den Versuch unternommen, die Differenzierungsfähigkeit des sechsstufigen Antwortmodells im unteren Bereich zu erhöhen. Vor allem sollte die Stufe "ungefähr" durch ein anderes verbales Urteil ersetzt werden, das ein wirklich geäußertes quantitatives Äquivalent besitzt als das bisher verwendete verbale Urteil. Aus unserem Analysematerial ist jedoch eine neue Wortwahl nicht zu gewinnen. Die zum Inkussion stehenden Alternativen (5 und 6 in Tabelle 20) unterscheiden sich im quantitativen Äquivalent nur zufällig. Für die Praxis können wir auf Grund der Skalenanalyse sowie der Transparenzbestimmung einzelner Stufen der AM vor allem den Einsatz des vierstufigen Antwortmodells empfehlen, wobei die ermittelten Skalenwerte bei der Lebensauswertung berücksichtigt werden sollten.

#### 4.2. Zum Einfluß der Frage- / Aussageformulierung auf den Befragungsergebnis bei Meinungsindikatoren

##### 4.2.1. Zur Ähnlichkeit des Befragungsergebnisses von zwei Frage- / Aussageformulierungen

Der Einfluß der Frage- oder Aussageformulierung auf das Befragungsergebnis ist, wie schon nachfolgend (vgl. v.) zu

gab, gegenwärtig methodenkritisch noch nicht ausreichend geklärt. Die daraus resultierende Unsicherheit behindert den Einsatz beider Formulierungsarten bei der Konstruktion von Fragebögen.

Die Ergebnisse früherer Untersuchungen unseres Instituts haben nur wenig zur Klärung dieses Problems beigetragen, sondern eher die Unsicherheit bei der Interpretation von Frage- und Aussageergebnissen vergrößert.

Für diese Entwicklung sehen wir vor allem zwei Gründe:

1. Frage- und Aussageitems wurden nicht mit äquivalenten AM eingesetzt.
2. Die Darbietungsfolge von Frage- und Aussageitems wurde nicht balanciert.

Beide Störgrößen wurden in der Untersuchung FB 51 (vgl. 2.2.) eliminiert.

Zur Kontrolle der inneren Validität unserer Versuchsanordnung werden wir anhand von Tabelle 21 der Frage nachgehen, ob die Darbietungsfolge von Frage- und Aussageitems das Untersuchungsergebnis beeinflusst. In Fragebogen FB 70-4 wurden die Aussageitems nach den Frageitems dargestellt, in FB 72-5 wurde die Reihenfolge dagegen umgekehrt. Die Darstellung der relativen Häufigkeitsverteilungen je Item und Darbietungsposition überzeugen, daß die Expositionsfolge der Formulierungsarten keinen Einfluß auf das Befragungsergebnis nimmt. Diese Aussage wurde durch eine Signifikanzprüfung mittels KST erhärtet (Spalte 3 in Tabelle 21).

Nachdem sich unsere Versuchsanordnung als intern valide erwiesen hat, kann geprüft werden, ob sich die Häufigkeitsverteilungen zu den Frageitems signifikant von den Häufigkeitsverteilungen zu den Aussageitems unterscheiden:

Hierzu betrachten wir jeweils die analogen Zeilen 4 + 5 der



Tabelle 21: Relative Häufigkeitsverteilungen zu den Items  
a bis f in FB 72 und FB 72-5

Aussage-Items

Item	FB-72	Stellung des Aussage- Items im Fragebogen	Stufen				
			1	2	3	4	D
a	4	2	1,7	3,3	23,3	71,6	
	5	1	4,1	4,1	19,3	72,9	0,037
	4-5	-	2,9	3,7	21,6	71,8	
b	4	2	1,7	2,5	24,4	71,4	
	5	1	0,8	4,2	28,4	66,6	0,048
	4-5	-	1,2	3,4	26,6	71,8	
c	4	2	3,3	0,8	35,0	60,9	
	5	1	4,1	8,3	37,2	50,4	0,105
	4-5	-	3,7	4,6	36,1	55,6	
d	4	2	4,2	3,4	42,0	50,4	
	5	1	7,4	4,9	37,2	50,4	0,047
	4-5	-	5,8	4,2	39,6	50,4	
e	4	2	5,8	10,8	25,0	58,4	
	5	1	7,5	11,7	20,0	60,9	0,024
	4-5	-	6,7	11,2	22,5	59,6	
f	4	2	7,5	5,0	25,3	61,6	
	5	1	5,0	5,0	19,4	70,6	0,090
	4-5	-	6,3	5,0	22,6	65,1	

Tabelle 21 (Fortsetzung)

Frage-Items

Item	FR-72	Stellung des Frage-Items im Fragebogen	1		2		3	
			1	2	3	4	5	
a	4	1	2,5	2,5	20,0	75,0		
	5	2	7,4	5,8	49,0	67,8	0,086	
	4=5	-	5,0	4,1	43,5	71,4	(0,031)	
b	4	1	0,8	3,4	22,7	73,1		
	5	2	1,7	5,0	25,6	66,6	0,074	
	4=5	-	1,3	4,2	24,6	69,8	(0,020)	
c	4	1	2,5	2,5	30,5	64,2		
	5	2	3,3	11,6	31,5	50,6	0,139	
	4=5	-	2,9	7,1	30,8	57,2	(0,015)	
d	4	1	2,5	5,0	30,0	51,5		
	5	2	9,1	4,9	33,9	52,0	0,066	
	4=5	-	5,8	5,0	33,9	53,3		
e	4	1	5,0	8,3	24,2	62,5		
	5	2	8,3	10,0	11,3	63,4	0,050	
	4=5	-	6,7	19,2	10,7	63,4	(0,058)	
f	4	1	11,7	5,8	12,8	56,7		
	5	2	9,2	5,9	16,0	63,9	0,060	
	4=5	-	10,5	5,8	13,9	67,7	(0,051)	

Tabelle 21<sup>8)</sup>. Schon allein optisch fällt die Konformität der Ergebnisse auf. Die statistische Prüfung bestätigt unseren Eindruck. Auf Grund dieses Ergebnisses können wir folgern:

Die Frage- bzw. Aussageformulierung führt bei gleichem Iteminhalt nicht zu differierenden Befragungsergebnissen. Diese Feststellung gilt unter der Voraussetzung, daß den Items äquivalente Antwortmodelle zugeordnet werden. Weiter werden die vorliegenden Ergebnisse auf Items, die politisch-staatsbürgerliche Meinungen (Einstellungen) erfassen, eingeschränkt. Trotz letzterer Restriktion ist das ermittelte Resultat für die Gestaltung von Fragebogen äußerst interessant.

Dieser Befund erlaubt

- a) innerhalb eines Fragebogens gleichzeitig mit Frage- und Aussageformulierungen zu arbeiten,
- b) in unterschiedlichen Studien den relevanten Sachverhalt einmal mit Frage- und einmal mit Aussageformulierung anzugehen.

Die Gleichheit der Befunde bei Frage- und Aussageformulierung erweitert die methodischen Möglichkeiten für die Gestaltung von Inqu海岸untersuchungen.

#### 4.2.2. Stabilität und Trennschärfe von Frage- und Aussageitems

Die Gleichheit der Häufigkeitsverteilungen ist für die inhaltliche Interpretation der Befragungsergebnisse von primärem Interesse. Aus methodenkritischer Sicht bedarf es jedoch des Nachweises, daß beide Formulierungsarten auch den

8) Fußnotenmerk in Klammern gesetzt. Werte  $\geq 0,121$  sind auf dem 5 %-Niveau signifikant. Werte ohne Klammern sind auf gleichem Niveau erst bei  $0,175$  bedeutsam.

Forderungen nach Trennschärfe und Meßgenauigkeit genügen. Für den Nachweis der Trennschärfe wurden zwei Kriterien gebildet:

1. Die Trennschärfe wird anhand der Summenwerte der sechs Standarditems in der jeweiligen Frage- bzw. Aussageformulierung beurteilt.
2. Die Trennschärfe wird mittels einer Summe aus 10 Frage- bzw. Aussageteams bestimmt, wobei Frageteams mit dem Aussagekriterium und Aussageteams mit dem Fragekriterium korreliert werden.  
Diese kreuzweise Trennschärfebestimmung ist für den gemeinsamen Einsatz von Fragen und Aussagen in ein und demselben Fragebogen anforderlich.

Tabelle 22: Trennschärfeeffizienten zu den Standard-  
Items beim Kriterium 1

Item/Formulierung	FB 72-4		FB 72-5	
	Frage	Aussage	Aussage	Frage
a	0,67	0,63	0,67	0,59
b	0,57	0,67	0,72	0,77
c	0,61	0,72	0,86	0,76
d	0,55	0,72	0,86	0,82
e	0,41	0,22	0,53	0,41
f	0,64	0,64	0,72	0,63
$\bar{x}_{10}$	0,57	0,60	0,72	0,63

Die Berechnung der punktbiserialen Korrelationskoeffizienten führt sowohl für Frage- als auch Aussageteams (vgl. Tabelle 22) zu annähernd gleichen Ergebnissen. Die Unterschiede in den Durchschnittswerten fallen zwar für die Aussageteams numerisch etwas günstiger aus, sind aber statistisch nicht praktisch ohne Belang.

Unter dem Aspekt der Trennschärfe können sowohl Frage- als auch Aussageteams mit gleicher Präzision zur Abbildung des relevanten Fragegegenstandes verwendet werden. Dabei spielt es keine Rolle (vgl. Tabelle 23), wie das Kriterium definiert ist.

Tabelle 22: Trennschärfe-Koeffizienten zu den Standard-Items beim Kriterium 2

Item/Formulierung	FE 72-4		FE 72-5	
	Frage	Aussage	Frage	Aussage
a	0,59	0,55	0,56	0,64
b	0,60	0,65	0,72	0,64
c	0,58	0,53	0,73	0,79
d	0,65	0,67	0,80	0,75
e	0,35	0,20	0,32	0,40
f	0,55	0,50	0,66	0,72
$\bar{r}_{it}$	0,55	0,52	0,64	0,65
Kriterium	Frage	Frage	Aussage	Aussage

Werden Aussageteams mit einem Kriterium, das aus Frageitems gebildet wird, korreliert, so unterscheiden sich die Trennschärfe-Koeffizienten der parallelen Items nur unwesentlich. Der gleiche Effekt tritt auf, wenn die Trennschärfe von Frageitems an einem Kriterium, das aus Aussageteams besteht, ermittelt wird.

Für die Maßgenauigkeit der Standarditems ergab die Varianzanalyse folgende Ergebnisse:

Aussage-Items:  $r_{it} = 0,73$  (FE 72-4) bzw.  $0,80$  (FE 72-5)

Frage-Items:  $r_{it} = 0,70$  (FE 72-4) bzw.  $0,75$  (FE 72-5)

Diese Koeffizienten lassen erkennen, daß sich beide Formulierungsarten auch hinsichtlich der Meßgenauigkeit entsprechen.

Die Kontrolle der Trennschärfe und Meßgenauigkeit bei Frage- und Aussageteilen führte zu folgenden Ergebnissen:

1. Die Formulierung der Items als Frage oder Aussage beeinträchtigt ihre Differenzierungsfähigkeit nicht.
2. Für die Trennschärfenanalyse können Frage- und Aussageteile zusammengefaßt werden, da die Kriteriumsdefinition keinen Einfluß auf das Analyseergebnis nimmt.
3. Die Meßgenauigkeit von Frage- und Aussageteilen ist als gleichwertig zu betrachten.

#### 4.2.3. Zur Transformation der Befragungsergebnisse bei Frage- und Aussageformulierung

Obwohl sich die Häufigkeitsverteilungen zur Frage- und Aussageformulierung nicht unterscheiden ist es für die Interpretation der Ergebnisse zweckmäßig, die Transformationsprozesse beim Übergang in eine andere Formulierungsart zu kennen.

Die Analyse erscheint um so dringlicher, da sich die Antwortmodelle zu Frage- und Aussageteilen in den verbalen Extremstufen stufen unterscheiden. Die postulierte Äquivalenz der Stufen "vollkommen", "ja" bzw. "überhaupt nicht" und "nein" gilt es zu verifizieren.

Einen ersten Überblick über die Parallelität der Messungen von Frage- und Aussageteilen vermittelt die Korrelationsanalyse. Die entsprechenden Koeffizienten wurden in Tabelle 24 zusammengestellt.

Tabelle 24: Parallelitätskoeffizienten zur Frage- /Aussage-  
Formulierung

Item	FB 72-4 Frage/Aussage	FB 72-5 Aussage/Frage
a	0,67	0,71
b	0,84	0,88
c	0,81	0,94
d	0,83	0,90
e	0,89	0,83
f	0,78	0,80
$\bar{r}$	0,80	0,84

Die Höhe der aufgefundenen Parallelitätskoeffizienten bestätigt unsere Vermutung, daß Frage- und Aussageitems komparative Abbildungen des Fragegegenstandes sind. Die Parallelität der Abbildung wird offensichtlich durch den Iteminhalt bestimmt.

Nächstend gilt es zu eruieren, inwieweit die nichtspezifische Varianz der Parallelitätskoeffizienten zu Lasten von Befragungseinflüssen bzw. notwendiger Unterschiede der Antwortmodelle geht (s. Tabelle 25).

Für die Extrepositionen fällt auf, daß der Anteil konstanter Antwortreaktionen beim Übergang von der Frage- zur Aussageformulierung jeweils geringer ist.

9,1 % der Analysepersonen wählen beim Übergang von der Stufe "ja" der Frageitems nicht die Stufe "vollkommen" der Aussageitems, sondern die Stufe "mit gewissen Einschränkungen". Bei der umgekehrten Transformation ist der Anteil mobiler Appl. geringer. Hieraus läßt sich folgern, daß die Indikatorbereiche beider Stufen etwas voneinander abweichen. Offensichtlich ist der Indikatorbereich der Stufe "ja" etwas weiter als der der Stufe "vollkommen". Mit einer ähnlichen Überlegung läßt sich auch die Unterschiede konstanter Wahlen bei den Extrepositionen "nein" und "überhaupt nicht" erklären. Allerdings gibt die geringe Anzahl relevanter Fälle (vgl. Tabelle 25) uns nur wenig Interpretationssicherheit.

Tabelle 25: Veränderungen in den Antwortpositionen beim Übergang von der Frage- zur Aussageformulierung und umgekehrt

Position des AM	Transforma- tion	(bei der anderen Formulierung gewählte Position in Prozent)				
		1	2	3	4	n
(4) ja	F --- A	1,2	0,2	9,1	<u>89,5</u>	474
(4) vollkommen	A --- F	1,4	0,0	3,6	<u>95,0</u>	446
(3) mit gewissen E.	F --- A	1,7	3,3	<u>84,5</u>	10,5	181
(3) mit gew. E.	A --- F	3,1	5,6	<u>80,6</u>	10,7	196
(2) kann	F --- A	12,1	<u>77,6</u>	7,3	3,0	33
(2) kann	A --- F	15,2	<u>80,5</u>	0,0	4,3	46
(1) nein	F --- A	<u>51,9</u>	18,5	18,5	11,1	27
(2) überhaupt n.	A --- F	<u>87,4</u>	6,3	0,0	6,3	32

Anmerkung: Anteil konstanter Reaktionen in Tabelle 25 unterstrichen.

Die Transformationen bei den Stufen "mit gewissen Einschränkungen" der Frage- und Aussageformulierung erhärten unsere Beobachtung (vgl. 4.1.3.), daß die Übergänge von Stufe zu Stufe fließend sind.

Aus den Befunden ist aber gleichzeitig zu schließen, daß die Meinungen zu den erfragten Sachverhalten nur wenig gefestigt sind. Bei gefestigter Meinung müßte die Zahl der mobilen (pn. beim Übergang von der Frage- zur Aussageformulierung zwischen den Stufen 3 und 4 nahe null sein, da der Indikatorbereich der Stufe 4 geringer ist.

Die Stufe "kann" steht sowohl beim Frage- als auch beim Aussagesitem in Konkurrenz mit der jeweiligen Extremposition. Die Differenzierungsfähigkeit dieser Stufe zu den Extremen "beide" bzw. "Überhaupt nicht", "kann" ist damit nur gering zu veran-



schlagen. Bei der Trennschärfenanalyse konnte der Mangel an Diskriminanz für die Stufe der Aussageitems empirisch aufgezeigt werden (vgl. 4.1.5.).

Auch zum Antwortmodell für Frageitems haben wir durch direkte Zuordnung von Prozentzahlen das subjektive quantitative Äquivalent der verbalen Urteile bestimmt. Es ergaben sich nachstehende Mittelwerte:

ja	78,64 %	(31,30) <sup>9)</sup>
mit gewissen Einschränkungen	59,48 %	(27,30)
kaum	21,29 %	(16,40)
nein	12,57 %	(24,30)

Die mittleren Prozentsätze stimmen mit den Ergebnissen zum Aussage-Antwortmodell (vgl. Tabelle 19) hinreichend überein. Auch beim Frage-Antwortmodell bedeutet die Antwort "ja" nur für 52 % der Probanden eine hundertprozentige Zustimmung zum vorliegenden Fragegegenstand.

Ebenso trifft nur für 45 % der Analysepersonen zu, daß die Antwort "nein" mit null Prozent Zustimmung gleichzusetzen ist.

Als Ergebnisse der Transformationsanalyse sind hervorzuheben:

1. Aussage- und Frageitems können als parallele Abbildungen ein und desselben Sachverhalts aufgefaßt werden.
2. Die den Items zugeordneten Antwortmodelle unterscheiden sich im Indikatorbereich der Extrempositionen. Diese Unterschiede sind aber praktisch unbedeutend, da sich bei beiden AM gleiche Häufigkeitsverteilungen ergaben.
3. Im quantitativen Äquivalent stimmen die Antwortmodelle zu Frage- und Aussageitems hinreichend überein.

---

9) Die Klammerwerte geben die Standardwerte an.

### 5. Zusammenfassung und Schlussfolgerungen

Die Methoden-Studie FB 72 beinhaltet zwei für die Konstruktion und Auswertung von Enqueteuntersuchungen wesentliche Problemstellungen:

1. Es wurde der Einfluß verschiedenstufiger Antwortmodelle auf das Befragungsergebnis von Meinungsindikatoren analysiert, wobei wir uns in dieser Studie auf Items mit politisch-staatsbürgerlichen Fragegegenstand beschränkten.
2. Es wurde der Einfluß der Itemformulierung auf das Befragungsergebnis untersucht. Unter diesem methodischen Aspekt der Fragebogengestaltung betrachteten wir die Wirkung der Frage- und Aussageformulierung ebenfalls bei Items mit politisch-staatsbürgerlichen Fragegegenstand.

Zur Absicherung der empirischen Befunde wurde ein spezieller Versuchsplan erarbeitet, der es durch Balancierung der Bedingungen (Antwortmodelle, Formulierungen) und Randomisierung der Analysenstichprobe gestattet, die innere Validität der vorliegenden Ergebnisse zu sichern. Zu diesem Zweck war es auch erforderlich, fünf Varianten des Erhebungsbogens anzufertigen, sowie die Interviewsituation durch Instruktion und Verlaufsprotokolle zu standardisieren.

In folgenden werden die wesentlichsten Ergebnisse der Studie FB 72 im Überblick genannt. Diese Befunde beziehen sich - das sei ausdrücklich vermerkt - auf Meinungsindikatoren mit politisch-staatsbürgerlichem Inhalt. Die eruierten Fakten gelten für die Hauptbefragungsgruppen unserer Forschungen, also Oberschüler, Lehrlinge und Studenten.

#### 5.1. Schlussfolgerungen zur Antwortmodellproblematik

1. Die Analyse ergab, daß sich verschiedenstufige Antwortmodelle gleicher Dimensionalität nicht nachteilig beeinflussen (4.1.1.). Für die Gestaltung von Fragebogen können wir hier

aus folgern, daß ein Wechsel dimensionsgleicher Antwortmodelle innerhalb eines Fragebogens möglich ist. Methodisch ist es durchaus legitim, die Monotonie des Fragebogens bei sehr ähnlichen Umfragebatterien durch einen Wechsel des Antwortmodells zu mindern.

2. Hinsichtlich ihrer allgemeinen Differenzierungsfähigkeit (Trennschärfe des gesamten AI) erwiesen sich vier-, fünf- und sechsstufige Antwortmodelle als gleichwertig. Für die Konstruktion von Fragebogen ist von Belang, daß die Trennschärfekoeffizienten gegenüber der Kriteriumsdefinition (4.1.2.) relativ unempfindlich sind. Bei der Trennschärfeanalyse können deshalb Items mit unterschiedlichen Antwortmodellen zu einer Batterie vereinigt werden.

3. Die Trennschärfebestimmung einzelner Stufen und deren Zusammenfassung (4.1.4.) verdeutlicht, daß eine differenziellste Interpretation der Befragungsergebnisse methodenkritisch nur wenig abgesichert ist.

Die vier- und fünfstufigen Antwortmodelle kann in dieser Weise die Stufe "vollkommen" als Ausdruck positiver Meinung (Einstellung) den anderen Stufen dieses AI entgegengestellt werden.

Für die Stufe "mit gewissen Einschränkungen" ist beim vier- und fünfstufigen Antwortmodell dagegen immer erst ein empirischer Nachweis der Trennschärfe erforderlich, um eine inhaltliche Wertung dieser Stufe zu vollziehen.

Das sechsstufige Meinungs-Antwortmodell besitzt mit "vollkommen" und "in großen und ganzen" zwei durchgängig diskriminationsfähige Stufen. Für alle anderen Stufen ist auch bei diesem Modell die Trennschärfe von Fall zu Fall empirisch nachzuweisen.

Die in der Interpretationsphase geübte Praxis, einzelne Stufen der Antwortmodelle zusammenzufassen, ist aus methodenkritischer Sicht in der Regel nicht berechtigt. Obwohl die empirische Analyse vielfach signifikante Koeffizienten ergab, erwiesen sie sich bei differenzieller Analyse

in der Mehrzahl der Fälle als statistische Artefakte. Von der unkritischen Zusammenfassung sollte deshalb in der Interpretationsphase dringend abgeraten werden.

4. Die Meßgenauigkeit der Antwortmodelle wird von ihrer Stufenzahl praktisch nicht beeinflusst. Vom Aspekt der Meßgenauigkeit sind alle Antwortmodelle zur Abbildung des Fragegegenstandes geeignet.
5. Aus der Transformationsanalyse sind folgende Ergebnisse zu nennen:
  - Die Abbildungen eines Sachverhaltes mit den analysierten vier-, fünf- und sechsstufigen Antwortmodellen können als parallele, jedoch nicht als identische Messungen betrachtet werden (4.1.3.).
  - Die Invarianz der Extremposition "vollkommen" ist gegenüber einem Wechsel der Stufenzahl der Antwortmodelle größer als die der Zwischenstufen. Jedoch ist auch der Indikatorbereich der Stufe "vollkommen" nicht unabhängig von der Stufenzahl. So ist er beim vier- und fünfstufigen AI größer als beim sechsstufigen.
  - Die Entscheidungskonstanz der Analysepersonen für die Stufe "vollkommen" nimmt mit dem Differenzierungsgrad des AI ab.  
Dieses Ergebnis ist als Hinweis darauf zu werten, daß die Indikatorgrenzen in den verschiedenen Antwortmodellen großen Schwankungen unterliegen.
  - Die Übergänge zwischen den Indikatorbereichen aller Stufen sind fließend. Eindeutige Zuordnungen von Stufen verschiedenstufiger Antwortmodelle sind nicht möglich.  
Im allgemeinen entspricht eine Stufe eines beliebigen Antwortmodells mehreren Stufen in einem anderen AI, wobei in der Regel eine Stufe am deutlichsten korrespondiert.

- Sollen Ergebnisse aus verschiedenstufigen AM verglichen werden, so lassen sich aus der Transformationsanalyse einige Empfehlungen ableiten:
1. Höherstufige Antwortmodelle können in niederstufige transformiert werden. Eine Umkehrung dieses Prozesses ist nicht möglich, da nicht eindeutig feststeht, in welche Stufen die Positionen des niederstufigen AM aufzuspalten sind.
  2. Generell lassen sich alle Antwortmodelle auf ein dreistufiges Modell reduzieren. Die Vergleichbarkeit der Befunde ist bei diesem Verfahren am besten gesichert. Im neuen Antwortmodell werden die beiden Extrempositionen beibehalten. Die Zwischenstufen werden zu einer neuen Position "eingeschränkte Zustimmung" vereinigt.
  3. Ziehen wir die Trennschärfenbestimmung der Stufen und ihrer Zusammenfassungen hier mit ein, so läßt sich auch eine methodenkritisch gut gesicherte Dichotomisierung der AM vorschlagen. Beim vier- und fünfstufigen AM besteht die eine Alternative aus der Stufe "vollkommen" die andere Alternative aus den realistischen Stufen der Antwortmodelle. Beim sechsstufigen Antwortmodell werden einerseits die Stufen "vollkommen" und "im großen und ganzen" und andererseits alle übrigen Stufen des AM zu Alternativen zusammengefaßt.
  4. Prinzipiell ist es auch möglich, bei der Transformation die proportionalen Anteile der Fluktuationsraten (4.1.3.) zu berücksichtigen. Allerdings erfordert dieses Vorgehen beträchtlichen methodenkritischen Aufwand.
  5. Der Einfluß der Antwortmodelle auf die Indikatorfunktionen der Items ist gering. Die Faktorenanalysen führten jeweils zu einem spezifischen Faktor, der 5,5 % bis 7,7 % der aufklärten Varianz bindet. Verzerrungen der Befunde
  6. Der Einfluß der Antwortmodelle auf die Indikatorfunktionen der Items ist gering. Die Faktorenanalysen führten jeweils zu einem spezifischen Faktor, der 5,5 % bis 7,7 % der aufklärten Varianz bindet. Verzerrungen der Befunde

gungsergebnisse durch die Antwortmodelle können bei einer Vielzahl von Eingangsvariablen nicht ausgeschlossen werden. Für wesentliche theoretische Arbeiten ist es daher wünschenswert, den gleichen Sachverhalt in parallelen Studien mit anderestufigen Antwortmodellen zu erheben.

## 7. Die Betrachtung der quantitativen Äquivalenz verbaler Urteile verdeutlicht:

- Die analysierten Antwortmodelle werden durch verbale quantitative Urteile konstituiert, die das unterstellte Kontinuum monoton wachsend repräsentieren.
- Zwischen den verbalen Urteilen besteht keine Äquidistanz. Die Skalen besitzen in der gegenwärtig verwendeten Form Ordinalniveaus.
- Zwischen direktem und indirektem Schätzverfahren zur Bestimmung der quantitativen Äquivalenz treten bei den Stufen des fünf- und sechstufigen Antwortmodells zum Teil erhebliche Differenzen auf. Für das vierstufige Antwortmodell ist dagegen die Anpassung der Schätzwerte sehr befriedigend.
- Für alle Antwortmodelle wurden über das "Law of Comparative Judgment" die entsprechenden Intervallskalen errechnet. Die aufgefundenen Skalenwerte spiegeln die subjektiven Distanzen zwischen den verbalen Urteilen wieder.
- In keinem Antwortmodell bedeuten die Stufe "vollkommen" bzw. "überhaupt nicht" absolute Zustimmung oder Ablehnung. Wählen die Befragten z.B. die Stufe "vollkommen", so stimmen sie dem erfragten Sachverhalt im Durchschnitt zu etwa 80 % zu.  
Über die anderen Zustimmungsgrade informiert 4.1.6.

## 5.2. Schlussfolgerungen zur Frage-/Aussageformulierung

1. Frage- und Aussageformulierung ergeben bei äquivalenten Antwortmodellen in jedem Fall identische Schäfergebnisse.

Literatur:

Glaß, G. / H. Lerner:

Grundlagen der Statistik  
Berlin 1967

Esser, G.:

Verbale Antwortmodelle für Fragen  
und Aussagen  
Unveröff. Untersuchungsbericht,  
ZIJ Leipzig, 1969Ettrich, K.W. / E.  
Schreiber:Zur Interpretation von Items an  
Umfragen  
Unveröff. Untersuchungsbericht,  
ZIJ Leipzig, 1972Ettrich, K.W. / E.  
Schreiber:Zum Einfluß verbaler Antwortmodelle  
auf das Befragungsergebnis bei  
Meinungsindikatoren  
Kolloquiumsreferat, Bogensee 1973

Friedrich, A.:

Methoden der marxistisch-lenini-  
stischen Sozialforschung  
Berlin 1970

Lienert, G.A.:

Testaufbau und Testanalyse  
Weinheim 1961

Lienert, G.A.:

Verteilungsfreie Methoden in der  
Biostatistik  
Weinheim 1962

Nelle, E.:

Umfragen in der Massengesellschaft  
Hamburg 1967