

## Zur Skalierung des subjektiven Werts von Zensuren bei Schulkindern

Von W. HOMMERS

Der Schwerpunkt empirisch-psychologischer Auseinandersetzung mit der Problematik von Schulzensuren lag bei der Prüfung ihrer Testgütekriterien. Diese Arbeit befaßt sich dagegen mit dem subjektiven Wert von Schulnoten. Vorrangige Aufgabe war bislang die Erstellung eines Meßverfahrens für den subjektiven Wert der Schulnoten. Allgemeines Kennzeichen des hier benutzten Verfahrens ist seine entscheidungstheoretische Grundlegung und die Zielsetzung, die Abstände der Zensurenwerte zu definieren. Unter Verwendung dieser Methode liegen Arbeiten von BECKER und SIEGEL [4], SCHMIDT [6] und SCHMIDT u. a. [7] vor. Den Untersuchungen ist gemeinsam, daß als Vpn mindestens 17 Jahre alte Schüler oder Studenten mitwirkten. Auch in der methodisch anders ansetzenden Arbeit von MÜLLER-FOHRBRODT und DANN [5] zu diesem Problembereich stammten die Vpn aus der umschriebenen Altersgruppe. Übergeordnetes Ziel dieser Arbeit ist deswegen die Prüfung der Durchführbarkeit einer entscheidungstheoretischen Skalierungsmethode bei Schülern jüngeren Alters.

### 1. Problemstellung

Zur ordinalen Bestimmung der Distanzen des subjektiven Werts von Zensuren nach SIEGEL [8] wird die Vp vor die Wahl zwischen 2 Alternativen gestellt. Entweder kann sie mit Sicherheit eine Note  $x$  bekommen oder mit der Wahrscheinlichkeit  $p = .5$  eine zu  $x$  bessere Note  $y$ . Allerdings geht sie bei letzterer Alternative das Risiko ein, eine zu  $x$  schlechtere Note  $z$  je nach Ausgang eines Zufallexperiments zu erlangen.

Theoretische Grundlage der Skalierung auf Grund von 15 derartigen Situationen ist das normative Erwartungsmaximierungsmodell. Vom objektiven Erwartungsmaximierungsmodell wird die Wahl der Alternative empfohlen, bei der die Produktsumme von Wert und Wahrscheinlichkeit maximal ist. Bei der Sonderform des Modells der Maximierung des subjektiv erwarteten Nutzens (SEU-Maximierung) wird die Wahl der Alternative empfohlen, deren Produktsumme von subjektivem Wert (Nutzen) und subjektiver Wahrscheinlichkeit maximal ist. Setzt man voraus, daß Vpn dieses letztere Modell tatsächlich in Entscheidungssituationen der dar-

gestellten Art benutzen, läßt sich der subjektive Wert der Noten bestimmen. Da die Wahrscheinlichkeiten der riskanten Konsequenzen in den vorliegenden Situationen gleich sind und damit auch ihre subjektiven Wahrscheinlichkeiten, kann bei tatsächlicher Verwendung des Modells durch die Vp die Entscheidung nur durch die Größen der subjektiven Wertunterschiede bedingt sein.

Die Voraussetzung der Verwendung des Modells des subjektiv erwarteten Nutzens durch die Vpn bei der Entscheidungsfindung kann bestritten werden. Zunächst kann eine Vp die Voraussetzung der alleinigen Verwendung der SEU-Maximierung nur probabilistisch erfüllen, d. h. sie macht systematische oder unsystematische Fehler. Weiterhin kann sie in Ermangelung dieser Entscheidungsstrategie völlig zufällig wählen oder auch nach anderen Gesichtspunkten. Sie kann etwa zusätzlich zum erwarteten Nutzen andere Aspekte<sup>1</sup> zu Hilfe nehmen oder stattdessen das objektive Erwartungsmaximierungsmodell verwenden oder angesichts des Risikos der einen Alternative lediglich an möglichst großer Sicherheit orientiert sein. Alle diese Strategien kann sie deterministisch oder probabilistisch verfolgen. Die an so entstandenen Daten anschließenden Skalierungen der Abstände der Zensuren müssen dann als Artefakte angesehen werden. Insbesondere wegen des Alters der Vpn sind einige dieser Möglichkeiten in Betracht zu ziehen, bevor die Verwendung der Skalierungstechnik bei dieser Altersstichprobe endgültig bejaht werden kann.

Prüfbar sind diese Punkte innerhalb des SIEGEL-Ansatzes durch Untersuchung der Transitivität und Konsistenz des Verhaltens in den 15 Situationen. Weiterhin sollen aber die Voraussetzungen der Methode von SIEGEL durch heuristische Alternativinterpretationen des Verhaltens kritisch untersucht werden. Auf Grund der vom Risikoaspekt her in Frage kommenden Validität der Portfolio-Theorie erscheint der Aspekt der Eindimensionalität des Verhaltens hier aufschlußreich. Mit einer bikategoriellen Rasch-Analyse (vgl. FISCHER [2]) läßt sich durch den Modelltest diese Frage bearbeiten. Dabei setzt man voraus, daß ein latenter Trait das Verhalten in den 15 Situationen beschreibt. Mit einer Faktorenanalyse ließen sich aus den Iteminterkorrelationen Anhaltspunkte über die Anzahl von orthogonalen Konstrukten zur Erklärung des Verhaltens gewinnen. Bei fehlender Eindimensionalität wird erwartet, daß sich die Ergebnisse im Sinne der Portfolio-Theorie interpretieren lassen.

Im folgenden werden die Ergebnisse zweier Untersuchungen dargestellt, die die Frage der Anwendbarkeit der Skalierungstechnik von SIEGEL bei Schulkindern unter Zuhilfenahme verschiedener Instruktionen überprüfen sollen.

<sup>1</sup> Von der Portfolio-Theorie nach COOMBS wird z. B. die Zuhilfenahme eines weiteren Aspekts bei der Entscheidung in Risikosituationen angenommen. Zusätzlich zum Erwartungswert wird dort der Abstand der Risiken der Alternativen von einem Idealen Risiko berücksichtigt. HOMMERS [4] konnte die Validität der Portfolio-Theorie auch bei Schulkindern stützen, so daß der Einfluß des Risikoaspekts bei der hier verwendeten Aufgabe mit in Betracht zu ziehen ist.

## 2. Methode

12 Gruppen von 309 Schülern der Klassenstufen 4, 6 und 8 aus Grundschulen, Hauptschulen und Gymnasien wurden untersucht. Die Vpn füllten jeweils 64 Entscheidungen bei 2 Instruktionen und 2 Fächern (Deutsch und Mathematik) in 4 Durchgängen an Hand eines Fragebogens. Die Reihenfolge der Fach-Instruktion-Kombinationen war ausbalanciert. Die Untersuchungen erfolgten im Klassenverband.<sup>2</sup>

Die Instruktionen A und B sollten, gemäß den von SCHMIDT [7] benutzten Bedingungen: Kontrollsituation versus Prüfungssituation, die Schüler in 2 Vorstellungszustände versetzen, während die Instruktion C eine schulkindadäquate Form hatte, die keine durch Verwendung von Wahrscheinlichkeiten mögliche kognitive Überforderung oder Provokation von Risikoeinstellungen enthält.

*Instruktion A:* Ihr schreibt doch mehrere Klassenarbeiten im Jahr! Stell Dir bitte vor, Du hättest im Fach *Deutsch* eine Klassenarbeit geschrieben! Bei der Note könntest Du zwischen 2 Möglichkeiten wählen. Du könntest entweder ganz sicher eine 2 bekommen oder, wenn Du lieber etwas riskieren willst, eine 1 oder aber eine 3. Bevor Du eine der beiden Möglichkeiten wählst, mußt Du wissen, wie Du bei der riskanten<sup>3</sup> Möglichkeit zu einer 1 oder einer 3 kommst. Wir machen nämlich ein Glücksspiel: Wir werfen ein Geldstück hoch. Wenn Zahl kommt, gibt es eine 1, und bei Wappen gibt es eine 3. Die Chancen, eine 1 oder eine 3 zu bekommen, sind also gleich groß. Wählst Du die sichere Möglichkeit, dann bekommst Du auf jeden Fall eine 2.

*Instruktion B:* Stell Dir bitte vor, es ginge um Dein *Versetzungszeugnis*! Bei der Note im Fach *Deutsch* könntest Du zwischen 2 Möglichkeiten wählen. Du könntest usw. (wie oben).

*Instruktion C:* Ihr schreibt doch mehrere Klassenarbeiten im Jahr! Stell Dir bitte vor, Du hättest im Fach *Mathematik* 6 Arbeiten geschrieben! Welche Noten hättest Du dann lieber bekommen?

1. Möglichkeit: 3mal eine 1 und 3mal eine 3
2. Möglichkeit: 6mal eine 2.

Die Instruktionenkombination A und B wurden bei 6 Klassenstufen-Schulgruppen mit 147 Schülern, die Kombination A und C bei ebenfalls 6 Klassenstufen-Schulgruppen von insgesamt 162 Schülern verwandt.

In der zweiten und dritten Zeile der Tabelle I sind die 15 hier eingesetzten Entscheidungssituationen in der benutzten Reihenfolge aufgeführt. Im Unterschied zu SCHMIDT [6] wurde die als Pufferreiz benutzte Einführungssituation an 7. Stelle wiederholt und die Reihenfolge der Situationen verändert.

<sup>2</sup> Die Datenerhebung erfolgte im Rahmen einer Diplomarbeit [3].

<sup>3</sup> Im Original wird dazu die Puffersituation dargestellt und deswegen nicht von riskanter und sicherer Möglichkeit, sondern von erster und zweiter Möglichkeit gesprochen.

Tabelle I. Alternativen in der Reihenfolge der Darbietung unter Nichtberücksichtigung der räumlichen Anordnung im Fragebogen und relative Wahrscheinlichkeiten der riskanten Alternative *R* unter verschiedenen Bedingungen

Abfolge	1	2	3	4 <sup>1</sup>	5	6 <sup>1</sup>	7 <sup>1</sup>	8	9	10	11	12	13	14	15 <sup>1</sup>	16 <sup>1</sup>
Sicheres Ereignis <i>S</i>	2	3	4	3	2	4	2	3	4	3	2	3	2	2	3	2
Unsicheres Ereignis <i>R</i>	1 oder 3	2 5	1 5	2 4	1 5	3 5	1 3	1 4	2 4	1 5	1 4	1 5	1 4	1 5	2 5	1 5
Instruktion A (mehrere Arbeiten)		11	49	38	12	44	42	35	53	16	24	29	26	8	33	17
Instruktion B (Versetzung)		17	43	28	15	38	43	34	49	15	22	26	21	10	26	25
SCHMIDT [6] <i>P</i> (Prüfung)		0	29	21	0	12	11	43	30	1	2	13	10	0	3	2
SCHMIDT [6] <i>K</i> (Kontrolle)		17	85	65	1	52	34	68	81	27	5	65	27	13	47	40
Instruktion C (ohne Wahr- scheinlichkeit)		10	66	46	12	34	49	46	53	11	20	39	25	5	26	17

*Erläuterungen:*

<sup>1</sup> Situationen die bei deterministischer SEU-Maximierung transitiv sein müssen und dann einen Skalentyp der Zensurenabstände definieren.

### 3. Auswertungsplan

Wenn deterministische SEU-Maximierung (und damit eine ordered metric scale nach SIEGEL) vorliegt, dürfen bei den Vpn keine intransitiven Wahlen in Situationen 4, 6, 7, 15 und 16 auftreten (5-Item-Transitivität). Die transitiven Wahlen in diesen 5 Situationen legen außerdem jeweils unterschiedlich viele Wahlen in einigen der restlichen 10 Situationen fest. Zur Unterscheidung nennen wir dies 5-Item-Konsistenz. Es läßt sich individuell prüfen, inwieweit diese vorliegt. Dadurch ist nach der Feststellung der 5-Item-Transitivität eine zweite Prüfung der deterministischen Verwendung der SEU-Maximierung möglich. Es interessiert, wie viele Vpn diese beiden Kriterien erfüllen. Die so verbleibenden Vpn können einerseits als Vpn angesehen werden, die für die Anwendung der Skalierung geeignet erscheinen. Andererseits können transitive Wahlen auch auf anderen Verarbeitungsmechanismen beruhen. Hier ist vor allem an zufällig entstandene Transitivitäten und Konsistenzen zu denken, wenn ihr Anteil sich als klein herausstellen sollte.

Im zweiten Analyseschritt soll versucht werden zu klären, ob gewisse probabilisti-

sche Prozesse dem Verhalten der Vpn zugrunde gelegt werden können. Es sei zunächst vorausgesetzt, daß alle Vpn in gleicher Weise reagieren. Würden sie sich überhaupt nicht systematisch verhalten, dann muß eine Statistik des Verhaltens der Vpn zufällig gemäß einer Binomialverteilung mit  $p=.5$  verteilt sein. Dies kann durch Vergleich der empirischen Verteilung für die individuelle Summe der S-Wahlen und der theoretischen Binomialverteilung geschehen. Bei negativem Ergebnis des Vergleichs (also Signifikanz der Abweichung) soll eine Poisson-Verteilung geprüft werden, wobei man eine absolute Sicherheitsorientierung der Vpn mit geringer Fehlerquote annimmt.

Geht man dagegen von einer Variation sowohl der Vpn als auch der Situationen bezüglich Neigung oder Anreiz zum Verfolgen solchen Sicherheitsstrebens aus, ist eine Prüfung dieser Verteilungen wenig aufschlußreich. Durch das latent-trait-Modell nach RASCH ist für diesen Fall eine Analysebasis gegeben. Da die Risiken der R-Alternativen variieren und bei alleiniger Abhängigkeit des Verhaltens vom bereichsspezifischen Sicherheitsstrebens eine monotone Itemcharakteristik anzunehmen ist (je größer das Risiko der R-Alternativ, um so größer die Wahrscheinlichkeit der Wahl der S-Alternative), wäre dann durch den Modelltest des bikategoriellen Rasch-Modells prüfbar, ob eine homogene Itemmenge vorliegt, bzw. ob die Annahme der Existenz einer latenten Dimension gültig ist. Scheitert dieser Modelltest an der Signifikanzschranke, dann bleibt noch die Möglichkeit einen oder mehrere nicht näher definierte Verarbeitungsmodi anzunehmen. Die Existenz solcher sollte sich in der Faktorenanzahl einer Faktorenanalyse, allerdings nur unter der Voraussetzung der Wirkungslosigkeit möglicher Verfälschungen auf Grund Unzulässigkeit ihrer methodischen Annahmen, bemerkbar machen. Die Verwendung dieses Verfahrens hat heuristischen Charakter. Hiervon unabhängig kann, soweit möglich, ein Vergleich der Ergebnisse mit SCHMIDT [6] vorgenommen werden. Falls die Eindimensionalität des Verhaltens sich als gegeben erweist, kann auch die Altersabhängigkeit des Verhaltens näher untersucht werden.

## 4. Ergebnisse

### 4.1. Vergleich mit früherer Untersuchung

Die Instruktionen A und B sind mit den von SCHMIDT benutzten vergleichbar. Die Wahlverteilungen unterscheiden sich aber im Gegensatz zu SCHMIDT nicht von einander (siehe Tab. I). Es zeigte sich<sup>4</sup>:

- Fehlen der Instruktionsabhängigkeit (die Differenz der Häufigkeiten zwischen A und B ist kleiner als bei P und K [ $\alpha < 1\%$ ]).
- Riskanteres Wählen unter Prüfungsbedingungen bei jüngeren Schulkindern (die Häufigkeiten bei B und P sind verschieden [ $\alpha < 1\%$ ]).
- Mehr sicheres Wählen unter Kontrollbedingungen bei jüngeren Schulkindern (die Häufigkeiten unter A und K sind verschieden [ $\alpha < 5\%$ ]).

<sup>4</sup> Vergleich von 15 Wahlhäufigkeitspaaren in den angegebenen Bedingungen mit dem WILCOXON-Test für abhängige Stichproben.

- Übereinstimmung der Bevorzugungsränge (bei allen möglichen 6 Kombinationen korrelieren die Ränge der Häufigkeiten [ $r_{sp} \geq .82$ ,  $\alpha < .10\%$ ]).
- Annäherung an Gleichverteilung bei jüngeren Vpn (die relative Häufigkeit der Wahl der riskanten Alternative ist bis auf einleuchtende Ausnahmen dem Wert 50% angenähert).

SCHMIDT gibt Häufigkeiten von Skalentypen auf Grund von 5 Situationen an. Es sind 16 transitive und 16 intransitive Quintupel des Wählens in 5 Situationen entsprechend der  $2^5$  Kombinationen des Verhaltens möglich. Ein transitives Quintupel des Wählens bildet eine Skala.

Der Vergleich von Prüfungssituation und Versetzungszeugnis bezüglich Anteil und Verteilung transitiven Verhaltens in 5 bestimmten Wahlsituationen ergab:

Von 45 Vpn bei SCHMIDT zeigten 20% intransitives Verhalten, gegenüber 20% von 147 in dieser Stichprobe. Bei SCHMIDT fielen 89% der Skalen auf die drei häufigsten, während hier nur 53% der Vpn diese Skalen zeigten. Gegenüber 8 überhaupt erzeugten Skalen bei SCHMIDT wurden hier alle 16 Skalen bestimmt.

Der Vergleich von Kontrollsituation und Klassenarbeit ergab:

Bei SCHMIDT wählten 9% intransitiv, hier 21%. Auf 4 bestimmte dort angegebene Skalen entfielen bei SCHMIDT 57% etwa gleichverteilt, hier 46%, von denen aber eine bevorzugt (80%) auftrat.

Den Hauptanteil bildeten in dieser Untersuchung, die 4 Skalen, die ihre größte Zensurenwertdifferenz zwischen den Zensuren „4“ und „5“ hatten. Unter Instruktion A waren dies 50% und unter B 61% der individuellen Skalen.

#### 4.2. Konsistenz des Verhaltens

Durch transitives Verhalten in den 5 Situationen (4, 6, 7, 15 und 16) ist festgelegt, wie sich die Abstände aufeinanderfolgender Zensuren ordnen lassen. Es sind durch dieses Verhalten aber auch weitere Wahlen vorhersagbar, wenn die ordered-metric-Skala zuverlässig durch sie definiert ist. Dabei legt jede der 16 möglichen Skalen der Zensurenabstände eine gewisse Anzahl von Wahlen in den verbleibenden 10 Situationen fest (mindestens 4, höchstens 8). Auf Grund dieser Konsistenz-Vorhersagen wurde die Konsistenz des Verhaltens der Vpn geprüft und damit die Berechtigung der Ableitung der gewonnenen ordered-metric-Skala aus den 5 angegebenen Situationen.

Von 1236 ausgefüllten Fragebögen (pro Vpn 4 Instruktionen) waren 287 durch konsistente, deterministische Wahl der S-Alternative, 3 durch konsistente deterministische Wahl der R-Alternative, 272 intransitiv in den angegebenen 5 Situationen, 665 darin transitiv und 9 durch fehlende Angaben ausgefüllt. Von den 665 transitiven Wahlen waren 308 inkonsistent. Von den 357 (nicht-deterministischen) konsistenten Wahlen fielen 205 auf zwei weiter unten angegebene Skalen. Zu einer davon können auch die 287 deterministischen S-Wähler hinzugezählt werden.

Zwei der 16 Skalen definierten das Verhalten ihrer Vpn zu 90%<sup>5</sup> intern zuverlässig. Alle weiteren 14 Skalen führten zu im Durchschnitt 20% konsistenter Wahlen, wobei die Verteilung dieser Prozentwerte mit Gipfel bei 0% schief war. Unter der Instruktion C traten durchschnittlich 46% konsistente Skalen auf, unter den anderen Instruktionen etwa 25%. Beziehungen zu den Vpn-Gruppen konnten nicht gefunden werden.

Das den beiden hochkonsistenten Skalen zugrundeliegende Verhalten unterschied sich in den 5 definierenden Situationen nur in der Entscheidung in Situation 7, sonst wurde von den Vpn immer die sichere Alternative gewählt. Diese beiden Skalentypen machen 65% der 5-Item-transitiven Verhaltensweisen aus, sie hatten als größte Nutzendifferenz die zwischen „4“ und „5“.

#### 4.3. Eignung einer Instruktionsform ohne Verwendung eines Zufallsexperiments

Die Instruktionsform C wurde zusammen mit der Instruktionsform A durchgeführt. Tabelle II enthält die Häufigkeiten von Quintupel- und Skaleneigenschaften auf Grund der 5 besonderen Items.

Tabelle II. Häufigkeiten des Ankreuzverhaltens in der Stichprobe von 162 Schülern

	Instruktion	
	A	C
Deterministische Wahlen in 15 Items	77	51
Intransitive Wahlen in 5 besonderen Items	78	74
Transitive Wahlen in 5 besonderen Items, davon		
Konsistent	79	127
Inkonsistent	90	72

Die Häufigkeiten zeigen, daß unter Instruktion C mehr konsistent transitives und weniger inkonsistentes oder deterministisches Verhalten auftrat. Der Unterschied entsprach einem Abweichungswert von  $\chi^2 = 18,57$ , der bei 3 Freiheitsgraden auf dem 0,1%-Niveau signifikant ist.

Aber auch bei Instruktion C traten vorrangig die Skalen mit der größten Differenz zwischen den Noten „4“ und „5“ auf. Für 69% des Verhaltens aller Vpn dieser Stichprobe traf dies unter Instruktion C zu, und für 60% unter Instruktion A. Damit liegt der Anteil dieser 4 Skalen noch etwas höher als bei der anderen Vpn-Gruppe.

In 8 der 15 Situationen läßt sich bei dieser Instruktion eine Durchschnittsnotenberechnung gemäß der Bildung eines arithmetischen Mittels durchführen. Vpn, die dieses Verfahren für ihre Entscheidungsfindung anwenden wollen, können dies nicht in den 5 besonderen Situationen tun, die zur Bestimmung der Skala dienen. Allerdings kann durch diese Strategie die Konsistenz einer Skala beeinflußt werden.

<sup>5</sup> Diese und die folgenden Prozentangaben wurden unter Ausschluß deterministischen Wahlverhaltens berechnet.

Tatsächlich läßt sich auf Grund der Zweigipfligkeit und/oder schiefen Verteilung der individuellen Wahlhäufigkeiten der durchschnitts-besseren Alternative, mit Modalwerten bei 4 oder 8 solcher Wahlen, annehmen, daß ein Teil der Vpn sich, wo möglich, dieser Entscheidungsstrategie bediente. Dies trat hauptsächlich bei Gymnasial-Schülern der 8. Klasse auf und wurde in dieser Stichprobe unter beiden Instruktionsformen beobachtet. Dagegen ließ sich dieses Verhalten so nicht in der anderen Schülerstichprobe mit den Instruktionen A und B finden.

#### 4.4. Verteilung individueller Summenwerte

Auf Grund des hohen Anteils von deterministischem Wahlverhalten waren die Verteilungen individueller Summenwerte linksgipflig. Sie wichen aber von der bestangepaßten Poisson-Verteilung oder Binomial-Verteilung äußerst signifikant ab (der kleinste  $\chi^2$ -Wert betrug 189,8 bei  $df=5$ ). Bereinigte man die Verteilungen um die Vpn, die durchgängig die sichere Alternative wählten, ergab sich zwar noch eine signifikante Abweichung von der bestangepaßten Poisson-Verteilung, die bestangepaßten Binomial-Verteilungs-Abweichungen waren dann aber nicht mehr auf dem 5%-Niveau signifikant. Dies legt die Zerlegung der Gesamtpopulation in ein deterministisches und ein probabilistisches Teilkollektiv nahe.

#### 4.5. Modelltest des bikategoriellen Rasch-Modells

In der Tabelle III sind die Ergebnisse des Modelltests im Rasch-Modell aufgelistet. Es zeigten sich zur Hälfte signifikante Modelltests. Addierte man die  $\chi^2$ -Werte jeder Stichprobe, erhielt man jeweils signifikante Summenwerte. Dies wird als Fehlen eines klaren Nachweises der Eindimensionalität eines latent-trait-Modells interpretiert. Die Untersuchung der Altersabhängigkeit dieser Summenwerte erbrachte keine Befunde.

Tabelle III. Ergebnisse der Modelltests des Rasch-Modells bei zwei Verhaltenskodierungen von zwei Verhaltensstichproben (+ 5%-Niveau; ++ 1%-Niveau; +++ 0,1%-Niveau)

S-Kodierung mit 15 Situationen

Methode	1. Stichprobe (N=147)		Methode	2. Stichprobe (N=162)	
	$\chi^2$	df		$\chi^2$	df
A Deutsch	43,1	9+++	A Deutsch	24,6	10++
B Deutsch	14,9	9	C Deutsch	18,4	10+
A Mathematik	13,2	9	A Mathematik	19,1	11
B Mathematik	28,7	11++	C Mathematik	11,2	10
Summe	99,9	38	Summe	73,3	41

#### 4.6. Faktorenanalysen der Interkorrelationen des Wahlverhaltens in den einzelnen Situationen

Die Eigenwerte in den Faktorenanalysen sprachen auf Grund ihres Abfalls nur für einen Faktor. Dieser hatte nur positive Ladungen. Bei Benutzung des Abbruchkriteriums „Eigenwerte größer 1“ ergaben sich durchweg vierfaktorielle Lösungen.

Die weiteren Faktoren hatten positive und negative Ladungen verschiedener Ausprägungsgrade. Klare Zuordnung der Ladungen auf den 4 Faktoren zu Itemeigenschaften wurden nicht gefunden, so daß eine Interpretation der Faktoren nicht vorgenommen werden konnte.

Problematisch war die Berechnung der Faktorenanalysen, weil in den Interkorrelationsmatrizen nur zufällig viele signifikante Korrelationen vorkamen und die Bestimmung der Interkorrelationen wegen der allgemein hohen Wahrscheinlichkeit zu sicheren Wahlen kaum sinnvoll erscheinen dürfte. Es wurde deswegen auf eine weitere Verarbeitung der Faktoren-Matrizen verzichtet.

## 5. Diskussion

Die Ergebnisse der durchgeführten Untersuchungen müssen überwiegend als negativ in bezug auf die mögliche Verwendung der ursprünglichen SIEGELSchen Skalierungstechnik bei jüngeren Schulkindern beurteilt werden. Zwar erwies sich die Technik bei über 50% der Fälle als erfolgreich anwendbar, aber etwa ein Viertel der Fälle bestand in deterministischem Wahlverhalten.

Dafür können vor allem zwei Gründe angeführt werden:

- Kognitive Überforderung der Vpn durch die Verwendung des Wahrscheinlichkeitsbegriffs und damit Instruktionsmängel.
- Unangemessenheit der SEU-Maximierungsvoraussetzung. (Da unter Instruktion C mit konkreten Häufigkeitsanteilen zum Teil abweichende Ergebnisse gefunden wurden und keine klaren Hinweise für Zweidimensionalität im Sinne der Portfolio-Theorie auftraten, hat der erste Grund mehr Gewicht.)

Zu weiteren möglichen ad-hoc-Erklärungsalternativen kann, wie folgt, Stellung genommen werden:

- Zerlegung in Teilkollektive

Der Befund über die Auswirkung des Weglassens der deterministischen Wähler aus der Verteilung der individuellen Summenwerte gibt Anlaß zu der Vermutung, daß die Stichprobe aus deterministischen Wählern der S-Alternative (mit geringer Fehlerquote) und binomialprozeßgesteuerten Zufallswählern besteht. Dies paßt gut zu dem Befund über die Konsistenz der transitiven Skalen. Diese erwiesen sich hauptsächlich nur dann als konsistent wenn die intern vorherzusagenden Wahlen in der Wahl der sicheren Alternative bestanden.

Unverständnis und Künstlichkeit der Situation könnten für das Auftreten von Zufallstechniken verantwortlich sein. Ernstgehalt von Zensuren könnte ein hohes Sicherheitsstreben erklären. Außerdem kann sich der Ernstgehalt von Zensuren auch in der häufigen maximalen Notendifferenz bei „4“ und „5“ bemerkbar gemacht haben.

- Validität der mehrdimensionalen Portfolio-Theorie

Während der vorherige Gesichtspunkt das Verhalten eines Teils der Stichprobe als undeterminiert auffaßt, wäre auch der Aspekt der möglichen Mehrdimensionalität des Verhaltens für das Scheitern eines eindimensionalen Skalierungs-

ansatzes als Erklärung in Betracht zu ziehen. Die Rasch- und Faktoren-Analysen ergaben aber kein klares Bild, so daß sich insbesondere die Validität der Portfolio-Theorie nicht als Begründung der Unzulässigkeit der SEU-Maximierungsannahme anbietet. Die unter Instruktion C auftretende Durchschnittsberechnung weist aber auf die Inhomogenität des Verhaltens einiger Vpn. in 15 Situationen hin.

In bezug auf die Frage nach einer geeigneten Durchführungsform der Skalierung des subjektiven Werts von Zensuren erscheint die Durchführungsform der Instruktion C erfolversprechend, zumal die dort schon ohnehin höhere Konsistenz wegen der von älteren Schülern zum Teil ausgeführten Durchschnittsberechnungen noch unterschätzt wurde. Da aber wegen dieser Durchschnittsberechnung keine methodenimmanenten Kontrollen sinnvoll möglich sind, wird eine direkte Einschätzung der Zensurenabstände zum Zweck der Validitätsprüfung der Skalierung unter Instruktion C vorgeschlagen. Dabei würde sowohl die SEU-Maximierungsannahme wie die Notwendigkeit der Verwendung des Wahrscheinlichkeitsbegriffes umgangen, so daß die Instruktion G noch enthaltene Utilitäts-Maximierungsannahme implizit mitgeprüft werden könnte.

### Zusammenfassung

An 2 Schülerstichproben ( $N=309$ ) der 4., 6. und 8. Klasse wurde die Anwendbarkeit einer entscheidungstheoretischen Skalierungsmethode für den subjektiven Wert von Schulzensuren unter 3 Instruktionen (Versetzung, Klassenarbeit mit Benutzung von Wahrscheinlichkeiten, Klassenarbeit ohne Benutzung von Wahrscheinlichkeiten) erprobt. Vergleiche mit Ergebnissen einer ähnlichen Untersuchung lassen große Unterschiede erkennen. Überlegungen und Auswertungen zur Ein- bzw. Zweidimensionalität des Verhaltens erbrachten keine eindeutigen Ergebnisse. Am ehesten scheinen 2 Teilkollektive vorzuliegen: Sicherheits-Wähler und Zufalls-Wähler. Damit erscheint die entscheidungstheoretische Skalierungsmethode bei Schülern der benutzten Altersgruppen nicht anwendbar, wenn Wahrscheinlichkeiten in den Situationen enthalten sind. Ohne Benutzung von Wahrscheinlichkeiten ergaben sich aber z. T. positive Ergebnisse.

### Summary

Two samples ( $N=309$ ) of fourth, sixth, and eighth graders were used to try a decision-theoretical scaling method of determining the subjective value of marks under three instructions, namely, promotion, class exercise with the use of probabilities, and class exercise without the use of probabilities. Comparisons with results of similar studies show considerable differences. Considerations and evaluations of one- and two-dimensionality of behavior gave no unequivocal results. There seem to be two groups, one comprising 'confidence' and the other 'random' selectors. Accordingly, the decision-theoretical scaling method is not applicable to pupils of the age groups used in this test, if the particular situations involve certain probabilities. Some of the tests which were conducted without the use of probabilities yielded positive results.

### Резюме

В двух выборках школьников (№ 309) четвертого, шестого и восьмого классов был проверен на применимость базирующийся на теории решения метод шкалирования для субъективной ценности отметок в трех условиях (перевод, классная работа с использованием вероятностей,

классная работа без использования вероятностей). Результаты настоящего исследования значительно отличаются от результатов подобных исследований. Размышления и обработка данных не дали однозначных результатов относительно одно- или двумерности поведения. По-видимому существуют две подгруппы: группа выбирающих верный вариант и группа выбирающих случайным образом.

Следовательно базирующийся на теории решения метод шкалирования оказывается неприменимым у школьников исследованных возрастных групп в ситуациях, содержащих вероятности. Без использования вероятностей отчасти были получены положительные результаты.

### Literatur

1. BECKER, S. W., und SIEGEL, S.: Utility of grades: Level of aspiration in a decision theory context. *J. exper. Psychol.* 55 (1958) 81–85.
2. FISCHER, G.: Einführung in die Theorie psychologischer Tests. Bern-Stuttgart-Wien: H. Huber 1974.
3. GATER, T.: Eine Untersuchung zur Erfassung des subjektiven Werts von Schulzensuren bei Haupt- und Oberschülern unterschiedlichen Alters. Diplomarbeit, Kiel 1976.
4. HOMMERS, W.: Zur Validität der Portfolio-Theorie im Entscheidungsverhalten von Schulkindern. *Z. Psychol.* 184 (1976) 604–618.
5. MÜLLER-FOHRBRODT, GISELA, und DANN, H.-D.: Zum Problem der Notengebung: Selbstbeurteilung von Zeugnisnoten. *Z. Entwicklungspsychol. u. Pädagog. Psychol.* 3 (1971) 241–252.
6. SCHMIDT, H.-D.: Leistungschance, Erfolgserwartung und Entscheidung. Berlin: VEB Deutscher Verlag der Wissenschaften 1966.
7. SCHMIDT, H.-D., WIESENBERG, M., und SCHOLZ, O. B.: Über Bedingungen der Veränderung von Nutzenskalen der schulischen Leistungsbewertung. *Z. Psychol.* 179 (1971) 281–299.
8. SIEGEL, S.: A method for obtaining an ordered metric scale. *Psychometrika* 21 (1956) 207–216.

Anschrift des Verfassers:

Dr. phil. WILFRIED HOMMERS

Institut für Psychologie der Christian-Albrechts-Universität

D-2300 Kiel, Olshausenstr. 40–60