

Verschiedene Verfahren – verschiedene Ergebnisse?

Vergleich der Skalierungsverfahren nach Rasch und Mokken sowie der klassischen Testkonstruktion am Beispiel alltagsästhetischer Schemata

Thomas Müller-Schneider

Universität Oldenburg, FB 3, Institut für Soziologie, Ammerländer Heerstr. 114–118, 26129 Oldenburg

Zusammenfassung: Die Theorie alltagsästhetischer Schemata bildet den substanzwissenschaftlichen Hintergrund dieses Aufsatzes. Alltagsästhetische Schemata sind kollektive Dispositionsvariablen zur Beschreibung von kulturspezifischen Erlebnisorientierungen in einer bestimmten Population. In Deutschland (alte Bundesländer) lassen sich gegenwärtig Hochkultur-, Spannungs- und Trivialschema unterscheiden. Zur empirischen Auffindung und Messung der Schemata werden Skalierungsverfahren benötigt, womit sich die Frage erhebt, ob soziologisch relevante Ergebnisse von der Auswahl des Skalierungsverfahrens abhängig sind. Der auf der Basis eines umfangreichen Itemsatzes durchgeführte empirische Vergleich der klassischen Testkonstruktion, der Rasch- und Mokken-Skalierung zeigt, daß dies im wesentlichen nicht der Fall ist. Weiterhin läßt sich auf der methodischen Ebene das Ergebnis festhalten, daß die klassische Testkonstruktion ein brauchbares, die Rasch-Skalierung ein problematisches und die Mokken-Skalierung ein bislang zu wenig beachtetes Verfahren zur Messung kollektiver Dispositionsvariablen ist. Auf der methodologischen Ebene stellen die Ergebnisse des vorgenommenen Vergleichs die Position in Frage, daß unterschiedliche Methoden grundsätzlich unterschiedliche Wirklichkeiten erzeugen.

1. Theoretischer Kontext und Fragestellung

Die folgende methodologische Studie nimmt sich eine Theorie zum Gegenstand, die G. Schulze (1992: 125 ff) entwickelte, um die in Deutschland (alte Bundesländer) entstandenen Formen soziokultureller Kollektivität beschreiben zu können. Ausgangspunkt ist der historische Wandel der normalen Problemperspektive im Alltagsleben, und zwar von der Sicherung des Lebensunterhaltes zur Suche nach dem „schönen Erlebnis“, das den Menschen etwa im Vergnügen und Kunstgenuß, in der intellektuellen Betätigung oder beim Zusammensein mit anderen begegnet. Aus der Perspektive der Erlebnisorientierung erscheint ein Leben um so sinnvoller, je dichter die Abfolge von Handlungsepisoden ist, die als „schön“, „aufregend“, „interessant“ usw. empfunden werden.

Eine entscheidende Rolle bei der Suche nach dem schönen Erlebnis spielt die Alltagsästhetik. Gemeint sind damit bewußte oder beiläufige Entscheidungen, die mit einer Erlebnisabsicht verbunden sind und viele Bereiche des Alltagslebens betreffen: Fernsehen, Musik, Wohnungseinrichtung, Lektüre, Kleidung, Accessoires, Besuch oder Nichtbesuch von Veranstaltungen und Einrichtungen u.a. Trotz dieser vielfältigen Entscheidungsmöglichkeiten ist die Alltagsästhetik kein Wirrwarr regeloser Launen, da die Menschen ihren persönlichen Stil, ein Repertoire relativ stabiler Neigun-

gen, entwickeln, mit dessen Hilfe sie ihr Leben genießen wollen (Genußebene), sich von anderen Menschen abgrenzen (Distinktionsebene) und mit dem sie zentrale Werte verbinden (Ebene der Lebensphilosophie) (Schulze 1992: 105ff).

Persönliche Stile können völlig individuelle Bedeutungen enthalten, aber auch solche, die mit mehreren Menschen geteilt werden. Diese kollektiven Bedeutungsmuster vieler persönlicher Stile führen zum Begriff des alltagsästhetischen Schemas. Dieses ist definiert als eine Gruppe von bedeutungäquivalenten Zeichen mit den Bedeutungsebenen Genuß, Distinktion und Lebensphilosophie; jede dieser drei Ebenen trägt eine Bedeutung, die innerhalb eines bestimmten Kollektivs äquivalent ist. So enthalten beispielsweise die ästhetischen Zeichen „Faust“, „Eine kleine Nachtmusik“, „Die Dreigroschenoper“ und „Feuilleton“ für die meisten Menschen unseres Kulturreiches einen gemeinsamen Bedeutungskomplex: die „Hochkultur“. Das Genußmotiv ist Kontemplation, die Distinktion kann als antibarbarisch bezeichnet werden und auf der lebensphilosophischen Bedeutungsebene ist das Streben nach Perfektion zu erkennen (Schulze 1992: 142 ff).

In der Bundesrepublik Deutschland lassen sich mindestens zwei weitere alltagsästhetische Schemata ausmachen, nämlich das „Trivial-“ und das „Spannungsschema“ (Schulze 1992: 150 ff). Beim Spannungsschema ist „Action“ die Quelle des Ge-

nusses, die Distinktion zielt auf die „Spießbürger“ und die zugrundeliegende Lebensphilosophie ist narzißtisch geprägt. Das Trivialschema zeichnet sich durch das Genußmuster Gemütlichkeit aus, durch die Distinktion von allem Individuellen und Besonderem sowie durch eine harmonieorientierte Lebensphilosophie.

Die Theorie alltagsästhetischer Schemata impliziert die Vorstellung, daß es sich bei den Schemata um kulturspezifische kollektive Dispositionen handelt, die dimensional strukturiert sind. Die dimensionale Struktur ist nicht von vornherein garantiert, sondern muß inhaltlich und empirisch begründet werden. Dimensionen können als wichtige qualitative Bereiche eines inhaltlichen Universums betrachtet werden, für die ein Kontinuum existiert, auf dem Untersuchungseinheiten unterschieden werden können (Atteslander 1991: 258). Das inhaltliche Universum bezieht sich im vorliegenden Fall auf persönliche Stile, wobei einzelne Dimensionen alltagsästhetische Schemata abbilden. Hierbei werden persönliche Stile als Punkte auf einem kollektiven Stilkontinuum angesehen, welches viele alltagsästhetische Entscheidungsmöglichkeiten bündelt. Geschmacksunterschiede treten in diesem dimensionalen Modell nicht als divergierende Stiltypen in Erscheinung, sondern als Differenzen, die sich aus den jeweiligen Ausprägungen der Nähe bzw. Distanz zu einem kollektiven Stilmuster (alltagsästhetischen Schema) ergeben.

Die alltagsästhetischen Schemata spannen einen dreidimensionalen Raum ohne klare Polaritäten auf. Die Nähe zu einem Schema bedeutet nicht notwendigerweise die Distanz zu einem anderen: Action und Kontemplation schließen sich nicht aus, ebenso wenig wie Action und Gemütlichkeit, oder Kontemplation und Gemütlichkeit. Die meisten Menschen basteln heute ihren eigenen Stil ohne zwingende Vorgaben zusammen (Schulze 1992: 157).

Alltagsästhetische Schemata sind kulturspezifische kollektive Dispositionen, die historischen Wandlungen unterliegen können: durch Entstehung und Verdichtung, durch Auflösung oder durch Änderung des mehrdimensionalen Beziehungsgefüchtes. Hochkultur- und Trivialschema haben sich in der Vergangenheit dimensional differenziert, d.h. aus einer bipolaren Dimension sind zwei relativ unabhängige Dimensionen hervorgegangen (Schulze 1992: 158 ff). Weiterhin hat das allgemeine Konsistenzempfinden bezüglich aller drei Schemata seit den 50er Jahren deutlich zugenommen (Müller-Schneider im Druck: 104 ff).

Für die empirische Sozialforschung werden alltagsästhetische Schemata auf der Ebene ihrer Zeichen faßbar, und zwar als konsistente Reaktionen auf übergreifende Zeichengruppen. Wenn eine Person beispielsweise gern klassische Literatur liest, häufig ins Theater geht, Konzerte mit klassischer Musik besucht und eine andere Person keiner dieser Beschäftigungen nachgeht, dann orientieren sich beide Personen am Hochkulturschema, auch wenn sie ihm unterschiedlich nahe stehen (vgl. Schulze 1992: 128). Zur Erkennung und Messung alltagsästhetischer Schemata kann bei geeigneter Datengrundlage somit das methodische Instrumentarium der Skalierungsverfahren eingesetzt werden: diese sind darauf angelegt, viele konsistente Reaktionen zu einer Skala zu bündeln, die dann als alltagsästhetisches Schema interpretiert wird.

Die zentrale Forschungsfrage des folgenden Methodenvergleichs ist, ob die Verwendung verschiedener Skalierungsverfahren zu verschiedenen inhaltlich-theoretischen Ergebnissen führt, oder ob die Theorie alltagsästhetischer Schemata und das darauf aufbauende Modell sozialer Milieus (Schulze 1992: 277 ff) weitgehend unabhängig vom gewählten Instrument ist.

Der Vergleich wird in mehreren Teilschritten dargestellt: Zunächst werden die Datenbasis und die drei verwendeten Skalierungsverfahren vorgestellt (Abschnitt 2), daran anschließend werden der bisherige Stand der Methodenforschung (Abschnitt 3) und die nötigen Kriterien des Vergleichs erläutert (Abschnitt 4). Nach der Darstellung des forschungspraktischen Vorgehens (Abschnitt 5) folgen die empirischen Ergebnisse (Abschnitt 6) und abschließend wird ein Fazit gezogen, das zu weitergehenden Überlegungen führt (Abschnitt 7).

2. Datenmaterial und Verfahren

Das verwendete Datenmaterial wurde 1985 von Schulze durch eine Standardrepräsentativumfrage erhoben. Die Grundgesamtheit bezog sich auf die im Einwohnermeldeamt registrierten Einwohner von Nürnberg (mit eingemeindeten Vororten) im Alter von 18 bis 70 Jahren. Die Nettoausschöpfungsquote lag mit 1014 auswertbaren Interviews bei 60% und somit im Trend sinkender Ausschöpfungsquoten. Aus einer methodischen Begleitstudie ergibt sich, daß die Stichprobenmängel sich aller Wahrscheinlichkeit nach nicht auf die zusammenhangsanalytischen Hauptergebnisse ausgewirkt haben (Schulze 1992: 593).

Mit dem standardisierten Hauptfragebogen wurden u.a. die hier relevanten Informationen über alltagsästhetische Präferenzen erhoben. Die strukturierten Interviews wurden mündlich durchgeführt, wobei die Befragten in einem „Antwortbogen“ mitlesen konnten. Bei den geschlossenen Fragen zur Alltagsästhetik wurde nach dem Interesse an ausgewählten Fernsehsendungen (Frage 13), der Vorliebe für bestimmte Musikarten (Frage 14), den bevorzugten Themengebieten bei der Zeitungslektüre (Frage 17), den sonstigen Lesegewohnheiten (Frage 19), nach Kenntnis und Besuch kultureller Einrichtungen der Stadt Nürnberg (Frage 57) sowie nach der Häufigkeit von ausgewählten Freizeittätigkeiten außer Haus (Frage 59) gefragt (Schulze 1992: 594ff).

Jede dieser Fragen sollte für eine Liste von Items mit mehrstufigen Antwortkategorien beantwortet werden. Bei Frage 13 etwa wurden ungefähr zwanzig Items („Heimatfilme“, „Fernsehshows“, „Quizsendungen“ etc.) mit der Response-Skala „interessiert mich sehr“, „ziemlich“, „mittelmäßig“, „wenig“, „gar nicht“ abgefragt. Dieses Frageschema ist bei allen Fragen vorzufinden, wobei Frage 14, 17 und 19 jeweils die gleiche Response-Skala aufweisen wie Frage 13. Bei Frage 14 wurden Items aus einem breiten Spektrum musikalischer Stile vorgelegt („Jazzmusik“, „klassische Musik“, „Popmusik“, „Blasmusik“ usw.), Frage 17 erfaßte das Interesse der Befragten an geläufigen Themenbereichen von Zeitungen („Politik“, „Wirtschaft“, „Kultur“ etc.) und mit Frage 19 wurde das Interesse an verschiedenen Literaturgattungen und Publikumszeitschriften erhoben („Moderne Literatur“, „Kriminalromane“, „Stern“, „Modezeitschriften“ u. ä.). Bei Frage 57 wurden die Befragten danach gefragt, ob sie bestimmte kulturelle Einrichtungen in Nürnberg (z.B. „Oper“, „Spielzeugmuseum“) kennen und in den letzten Jahren besucht haben. Die Beantwortung erfolgte auf der Antwort-Skala: „kenne ich nicht“; „kenne ich, aber noch nicht besucht“; „einmal besucht“; „mehrmais besucht“. Bei Frage 59 sollten die befragten Personen angeben, wie oft sie ihre Freizeit mit bestimmten Tätigkeiten außer Haus („ins Kino gehen“, „jemand besuchen“ etc.) verbringen. Als Response-Skala wurde die zeitliche Abstufung „sehr oft“, „oft“, „gelegentlich“, „selten“, „nie“ vorgelegt.

Die Befragungsergebnisse bezüglich der mehr als einhundert alltagsästhetischen Entscheidungsmöglichkeiten bildeten die Datengrundlage des angestrebten Methodenvergleichs, für den folgende Skalierungsverfahren ausgewählt wurden: Die

klassische Testkonstruktion, die Rasch- und die Mokken-Skalierung.

Der Grundgedanke der klassischen Testtheorie besteht darin, daß die Zustimmung bzw. Ablehnung von Items auf nur eine Dimension zurückgeführt werden kann, und daß mit steigender Ausprägung dieser Dimension die Wahrscheinlichkeit für die Zustimmung (symptomatische Antwort) zu einem Item zunimmt: Die Itemcharakteristiken einer nach der klassischen Testtheorie konstruierten Skala müssen einen monotonen Verlauf aufweisen (Schnell et al. 1988: 188). Die klassische Testkonstruktion ist das dominante Verfahren in der sozialwissenschaftlichen Forschungspraxis, obwohl die meßtheoretischen Grundlagen der klassischen Testtheorie immer wieder heftig attackiert wurden (Fischer 1974; Wottawa 1977; Hilke 1980; Heidenreich 1984 u. a.). Das Spannungsverhältnis zwischen Forschungspraxis und modelltheoretischer Kritik ist Grund genug, die klassische Testkonstruktion in den Methodenvergleich einzubeziehen. Unter „klassischer Testkonstruktion“ wird im folgenden sowohl die klassische Itemanalyse als auch die Faktorenanalyse verstanden, die als mehrdimensionale Verallgemeinerung der klassischen Testtheorie aufzufassen ist (Hilke 1980: 373 ff.).

Die Rasch-Skalierung wird häufig als Alternative zur klassischen Testkonstruktion gepriesen, weil sie einige modelltheoretische Vorteile bietet (vgl. Schnell et al. 1988: 197 ff.), deren wichtigster das Postulat „spezifisch-objektiver“ Messungen ist, die sowohl den Vergleich von Items unabhängig von Personen, als auch den Vergleich von Personen unabhängig von modellgültigen Items erlauben. Das Rasch-Modell wird durch Item- und Personenparameter definiert, wobei die symptomatische Antwortwahrscheinlichkeit eines Items durch eine logistische Funktion mit diesen Parametern verbunden ist. Ein Vergleich des Rasch-Verfahrens mit der klassischen Testkonstruktion könnte Anhaltspunkte dafür liefern, ob sich die modelltheoretischen Vorteile des Rasch-Modells auf der Ebene der soziologischen Theoriebildung bemerkbar machen und somit den hohen Aufwand bei der Konstruktion einer Rasch-Skala rechtfertigen können.

Die Mokken-Skalierung (einführend: Niemöller/van Schuur 1983) gehört ebenso wie die Rasch-Skalierung zur Klasse der „latent-trait“ Modelle und kann als probabilistische Weiterentwicklung der Skalogrammanalyse nach Guttman aufgefaßt werden. Die Wahrscheinlichkeit für eine sympto-

matische Antwort auf ein Item hängt einerseits vom Personenwert ab, der die Position auf der latenten Dimension repräsentiert, und andererseits von der Schwierigkeit des Items. Im Unterschied zum Rasch-Modell ist die funktionale Form der monotonen Itemcharakteristiken beim Mokken-Modell nicht festgelegt, sie dürfen sich aber nicht überschneiden. Diese Bedingung wird als doppelte Monotonie bezeichnet: Je größer die Schwierigkeit eines homogenen Items, desto geringer muß die symptomatische Antwortwahrscheinlichkeit sein, und je höher der Personenwert, desto höher muß diese Wahrscheinlichkeit sein (Lippert et al. 1978: 256 f.). Die Mokken-Skalierung weist durch ihren mehrdimensionalen Ansatz (Abschnitt 5) ein hohes exploratives Potential auf und legt somit einen Vergleich mit der Faktorenanalyse nahe. Beide Verfahren sind zur Auffindung der Dimensionalität eines Itemsatzes geeignet.

Da sowohl beim Rasch- als auch beim Mokken-Verfahren die dichotomen Modelle eingesetzt wurden, mußte das oben skizzierte Itemmaterial dichotomisiert werden. Um die unterschiedlichen Itemschwierigkeiten nicht zu verwischen, wurde nicht am Median dichotomisiert, sondern an der „inhaltlichen Mitte“ der Antwortvorgaben. Dieser dichotomisierte Datensatz wurde von sämtlichen fehlenden Werten bereinigt und bildete dann zugleich der Alters- und Bildungsvariable sowie der Skalen „Politische Unterordnung“ und „Vertrauen“ (Schulze 1992: 612), die zur zusätzlichen Kontrolle der statistischen Ergebnisse eingesetzt wurden (vgl. Abschnitt 4), mit 855 Fällen die Grundlage aller weiteren statistischen Auswertungen, die im Rahmen des Methodenvergleichs vorgenommen wurden.

Ähnliche Ergebnisse können für die klassische Testkonstruktion und die Mokken-Skalierung erwartet werden, da beide dimensionsanalytischen Modelle die „klassische Itemhomogenität“ (Jansen 1982: 87) interner Konsistenz zugrundelegen; somit sind Interitemkorrelationen die wesentlichen Größen bei beiden Skalierungsprozeduren (vgl. Henning/Six 1977: 192). Itemgruppen mit relativ hohen internen und relativ niedrigen externen Korrelationen schlagen sich sowohl in der Faktorenanalyse als auch beim Mokken-Verfahren in Form von gesonderten Dimensionen nieder, im einen Fall als Faktoren und im anderen als skalierbare Itemsets.

Obwohl sich das Rasch-Modell vom Mokken-Modell letztlich nur dadurch unterscheidet, daß die

funktionale Form der doppelt monotonen Itemcharakteristiken beim Rasch-Modell durch restriktive Annahmen (logistische Funktion) festgelegt ist, lassen sich über die zu erwartenden Ergebnisse kaum Aussagen machen. Das Rasch-Modell zieht nämlich nicht den Homogenitätsbegriff der internen Konsistenz heran, sondern den der lokalen stochastischen Unabhängigkeit. Dies hat zur Folge, daß homogene Rasch-Items nicht unbedingt miteinander korrelieren müssen. Stochastische Unabhängigkeit und perfekte Rasch-Homogenität schließen sich für den Fall nicht aus, daß alle Personen denselben Personenparameter aufweisen (Stelzl 1979: 654). Beim Rasch-Verfahren können somit im Gegensatz zu den beiden anderen Verfahren nicht die Interitemkorrelationen im Mittelpunkt der Skalierungsprozedur stehen, statt dessen wird meist die Bedingung der Stichprobenunabhängigkeit der Personen- und Itemparameter herangezogen, um die Modellgültigkeit zu überprüfen (vgl. Abschnitt 5).

3. Stand der Methodenforschung

Ein systematischer und an substanzwissenschaftlichen Ergebnissen orientierter Vergleich mehrerer Skalierungsverfahren hat in der sozialwissenschaftlichen Methodenforschung für den soziologischen Gegenstandsbereich bislang nicht stattgefunden. In der Forschungsliteratur finden sich jedoch Anhaltspunkte dafür, daß die klassische Testkonstruktion sowie die Rasch- und die Mokken-Skalierung zumindest auf der itemanalytischen Ebene zu ähnlichen Ergebnissen führen.

Beispielsweise nahmen Lippert et al. eine itemanalytische Untersuchung zweier Itempools zur Messung der Einstellungskonzepte Anomie bzw. politische Entfremdung vor und zogen das Fazit (1978: 271), daß die Faktorenanalyse, die Mokken- und die Rasch-Skalierung bei beiden Itempools zu annähernd gleichen Itemsätzen führen. In die gleiche Richtung weisen die von Henning und Six (1977: 196) vorgelegten Ergebnisse unterschiedlicher Verfahrenswege zur Konstruktion einer Machiavellismus-Skala: Klassische Itemanalyse, Faktorenanalyse sowie Mokken- und Rasch-Skalierung führten zu einer Reihe von gemeinsamen Skalenitems, die den unterschiedlichen Kriterien der vier Techniken genügten. Giegler (1985: 27) kam nach der itemanalytischen Untersuchung eines umfangreichen arbeits- und berufsbezogenen Itempools mittels der Faktorenanalyse, der klassischen Itemanalyse und der Rasch-Skalierung

zu der Auffassung, daß die Rasch-Skalierung keineswegs den Gütekriterien der klassischen Testtheorie widersprechen muß und diese sogar verbessern kann.

Diese und ähnliche Studien sind für den beabsichtigten Methodenvergleich jedoch nicht ausreichend: Zum einen handelt es sich um andere inhaltliche Bereiche und zum anderen vernachlässigt die Betrachtung der rein itemanalytischen Ebene weitere Vergleichskriterien, die sich aus der Perspektive des gesamten Theoriekontextes ergeben.

4. Theorierelevante Vergleichskriterien

Um den beabsichtigten Methodenvergleich in forschungspraktisches Handeln umsetzen zu können, müssen relevante Kriterien benannt werden, anhand derer über den Einfluß der Skalierungsverfahren auf substantielle Ergebnisse und somit auf die Theorie alltagsästhetischer Schemata entschieden werden kann. Die folgenden sechs Vergleichskriterien sind jedoch nicht nur für die vorliegende Fallstudie von Bedeutung, sondern betreffen Fragen, die allgemein bei der Messung und Beschreibung kollektiver Dispositionsvariablen sowie deren Weiterverwendung in soziologischen Theoriekontexten auftreten.

Das erste Vergleichskriterium bezieht sich auf die Frage der Dimensionalität des inhaltlichen Universums persönlicher Stile, d. h. auf die Anzahl relevanter qualitativer Bereiche, die begrifflich als alltagsästhetische Schemata gefaßt wurden. Diese Frage kann nicht ausschließlich methodisch entschieden werden, da die gewählte Anzahl relevanter Dimensionen auch von subjektiven Entscheidungen abhängt; dennoch spielt die empirische Begründung bei der Festlegung der Dimensionalität eine wichtige Rolle. Die empirische Begründung des vorliegenden dimensionalen Modells alltagsästhetischer Schemata basiert auf faktorenanalytischen Berechnungen, die von Schulze durchgeführt wurden und auf drei zentrale Dimensionen der gegenwärtigen Alltagsästhetik hinweisen, nämlich auf das Hochkultur-, Spannungs- und Trivialschema (Schulze 1992: 628 f; Abschnitt 5).

Setzt sich die strukturelle Dominanz dieser Dimensionen auch beim Mokken-Verfahren durch, dann kann dies als Hinweis auf die Verfahrensunabhängigkeit der behaupteten dreidimensionalen Struktur kollektiver Stile gewertet werden. Das Vergleichskriterium Dimensionalität kann jedoch nur auf die klassische Testkonstruktion und die

Mokken-Skalierung angewendet werden, da die Rasch-Skalierung für die Erkennung mehrdimensionaler Strukturen grundsätzlich nicht in Frage kommt.

Weiterhin sind die Ergebnisse der drei Skalierungsverfahren nur dann methodenunabhängig, wenn die verfahrensspezifischen Skalen das gleiche messen. Dabei sind zwei Aspekte zu unterscheiden, nämlich die Bedeutungsgleichheit der verfahrensspezifischen Skalenitems und die Äquivalenz der Meßwerte. Diese beiden Aspekte führen zum zweiten und dritten Vergleichskriterium des Methodenvergleichs: der inhaltlichen Entsprechung der skalierten Itemsätze und der Korrelation der Meßwerte.

Wie alle kollektiven Dispositionsvariablen sind alltagsästhetische Schemata latente Konstrukte über deren Existenz auch methodisch zu entscheiden ist. Daher muß untersucht werden, ob die Entscheidung über die Existenz der alltagsästhetischen Schemata unabhängig vom eingesetzten Instrumentarium ist. Die interne Konsistenz der verfahrensspezifischen Skalen kann als Indikator für die Existenz eines alltagsästhetischen Schemas herangezogen werden und bildet somit das vierte Kriterium des Methodenvergleichs.

Das Korrelationsmuster, welches sich aus den Korrelationen zwischen den alltagsästhetischen Schemata und externen Merkmalen ergibt, ist das fünfte Vergleichskriterium. Alter und Bildung sind die wichtigsten externen Merkmale, denn deren statistischer Zusammenhang mit den drei ästhetischen Schemata bildet die empirische Grundlage des von Schulze verwendeten Modells sozialer Milieus. Wenn alle Verfahren Skalen hervorbringen, die ein ähnliches Korrelationsmuster mit den externen Kriterien Alter und Bildung aufweisen, dann hat das untersuchte Instrumentarium keinen relevanten Einfluß auf das Milieumodell. Zur zusätzlichen Kontrolle gingen neben Alter und Bildung die beiden Skalen Vertrauen und politische Unterordnung als externe Merkmale in den Vergleich ein.

Mit dem sechsten Kriterium wird das dimensionale Beziehungsmuster der drei Schemata vergleichend untersucht. Die Beziehung zwischen je zwei alltagsästhetischen Schemata hat drei Extremfälle: Die Schemata sind unvereinbar (vollständiger negativer Zusammenhang), treten nur gemeinsam auf (vollständiger positiver Zusammenhang) oder sie sind willkürlich kombinierbar (statistische Unabhängigkeit). Es ist zu überprüfen, ob sich das von Schulze postulierte Beziehungsmuster ohne

klare Polaritäten (keine bzw. schwache statistische Zusammenhänge) bei allen Verfahren durchsetzt.

Zusammenfassung: Die empirischen Ergebnisse der Skalierungsprozeduren werden im Hinblick auf die Dimensionalität des Itemmaterials (1), die inhaltliche Entsprechung der skalierten Itemsätze (2) und die Korrelationen der Meßwerte untersucht (3); weiterhin wird die innere Konsistenz der Skalen (4), deren Korrelationen mit externen Kriterien (5) sowie das dimensionale Beziehungsmuster der Skalen (6) vergleichend bewertet.

5. Forschungspraktisches Vorgehen

Zur Durchführung der klassischen Skalenkonstruktion wurden im Rahmen des Methodenvergleichs keine Faktorenanalysen gerechnet, da auf bereits vorliegende Ergebnisse in der Untersuchung von Schulze zurückgegriffen werden konnte. Ziel seiner explorativen Faktorenanalysen der alltagsästhetischen Präferenzen war es, Itemgruppen mit hohen gruppeninternen und niedrigen gruppenexternen Korrelationen zu erkennen. Derartige Itemgruppen können als empirisches Substrat einzelner Dimensionen in einem mehrdimensionalen Raum betrachtet werden. Das Hauptkomponentenmodell der Faktorenanalyse ist zur Auffindung solcher Itemstrukturen geeignet. Es wurden mehrere Hauptkomponentenanalysen mit unterschiedlicher Anzahl vorgegebener Dimensionen gerechnet und mit dem Varimax-Kriterium rotiert. Auf diesem Wege konnten drei große Itemfelder gefunden werden, die sich nach dem Prinzip der Einfachstruktur darstellen lassen (Schulze 1992: 628 f.).

Die so gefundenen Itemfelder wurden dann eindimensional mit der klassischen Itemanalyse weiterverarbeitet, um die vielfältigen Probleme der dimensionsanalytischen Verwendung der Faktorenanalyse zu umgehen (vgl. Gillespie et al. 1987: 394). Nach der Eliminierung wenig trennscharfer Items wurden die endgültigen Skalen durch den jeweiligen Trennschärfekoeffizienten der Items und ihre Gesamthomogenität (Cronbachs Alpha) charakterisiert.

Für den Methodenvergleich wurden die Items der drei Skalen unter Verwendung des dichotomisierten Datensatzes nochmals einer Itemanalyse unterzogen, in der Absicht alle Items zu eliminieren, deren korrigierter Trennschärfekoeffizient den Wert von .30 unterschritt.

Die Personenkennwerte ergaben sich nach der üblichen Vorgehensweise aus der Summe der sym-

ptomatisch beantworteten Fragen (Summenscore).

Die Verfahrensweise zur Auffindung der Mokken-Skalen bedarf einer eingehenderen Erläuterung. Im Mittelpunkt der Skalierungsprozedur stehen die Homogenitätskoeffizienten H_{ij} , H_i und H . Der Koeffizient H_{ij} mißt die Homogenität eines Itempaars (i,j) und entspricht dem Korrelationskoeffizienten ϕ/ϕ_{max} in 2x2 Tabellen. Der Koeffizient H_i bringt in Analogie zum Trennschärfekoeffizienten die Homogenität eines Items (i) in bezug auf eine Gesamtskala zum Ausdruck und entspricht dem gewichteten Mittel der Koeffizienten H_{ii} , die das Item (i) enthalten. Die Skalenhomogenität H (Loevingers H) ist ein gewichtetes Mittel aller Koeffizienten H_{ij} (Stokman/van Schur 1980: 23).

Das statistische Programm MSP (Debets/Brouwer 1988) bietet die Möglichkeit der schrittweisen Skalierung eines Itemsatzes: Zuerst wird das homogenste Itempaar gesucht (gemessen an H_{ij}), dem dann das Item mit dem besten Skalierbarkeitskoeffizienten bezüglich des Itempaars (H_i) zugeordnet wird; bei abnehmender Skalierbarkeit wird dieser Skalenkern solange erweitert, bis ein festgelegtes Abbruchskriterium erreicht ist. Auf der Grundlage der verbleibenden Items wird anschließend eine weitere Skalierungsrunde eingeleitet, wobei wiederum das Itempaar gesucht wird, das die größte Homogenität aufweist und somit den Ausgangspunkt einer möglichen Skala bildet. Die Prozedur der Skalenbildung wird solange weitergeführt, bis keine Skalen mehr gebildet werden können. Dieser mehrdimensionale Ansatz der Mokken-Skalierung hat ein hohes exploratives Potential, das bei der Skalierung des vorliegenden Itemmaterials maximal ausgeschöpft wurde.

Die Skalenkonstruktion erfolgte in drei Teilschritten. Ziel des ersten Schrittes war es, auf der Grundlage des gesamten Itemmaterials extrem homogene Kristallisierkerne möglicher Skalen aufzufinden. Die Skalenbildung wurde bereits in einem Stadium abgebrochen, in dem der Skalierbarkeitskoeffizient (Loevingers H) noch sehr hohe Werte aufwies ($H > .50$). Diese Vorgehensweise führte zu einer Reihe von Skalenfragmenten, die mitunter nur vier oder fünf Items enthielten und als Startsets für den zweiten Skalierungsschritt verwendet wurden. Dabei kam es darauf an, die Skalenkerne durch die Herabsetzung des Abbruchskriteriums mit weiteren Items anzureichern, und zwar auch mit solchen, die im ersten Skalierungsschritt einer anderen Skala angehörten. Das Abbruchskriterium wurde so gewählt, daß der jewei-

lige Skalierbarkeitskoeffizient der Skalen deutlich über .40 und somit im Bereich einer „guten“ Skala lag (vgl. Lippert et al. 1978: 259).

Der dritte Skalierungsschritt diente zur Konstruktion der endgültigen Mokken-Skalen. Die drei Itemsätze, die dem Hochkultur-, Spannungs- und Trivialschema entsprachen, wurden im Hinblick auf ihre Holomorphie geprüft, die eine wichtige Komponente des Mokken-Modells ist und mit der sogenannten P-Matrix überprüft werden kann (Lippert et al. 1978: 266 ff.). Einige Items mußten wegen mangelnder Holomorphie eliminiert werden.

Der Meßwert einer Person ist beim Mokken-Verfahren durch das Ordnungstheorem theoretisch begründet und, analog zur klassischen Testkonstruktion, durch die Summe der symptomatischen Antworten definiert (vgl. Lippert et al. 1978: 258).

Bei der Rasch-Skalierung konnte auf die Ergebnisse der klassischen Testkonstruktion und der Mokken-Skalierung nicht gänzlich verzichtet werden, weil das Verfahren zur Erkennung mehrdimensionaler Strukturen nicht geeignet ist. Um dennoch weitestgehend unabhängig zu arbeiten, wurde für jede durch die ersten beiden Verfahren aufgefundenen Dimension ein weit über diese hinausreichender inhaltlich ausgewählter Itempool erstellt, aus dem dann jeweils eine Rasch-Skala gebildet wurde.

In der Forschungspraxis finden vorwiegend Modelltests Anwendung, die sich der Teilgruppenkonstanz der Itemparameter bedienen (Kubinger 1988: 24): bei Modellgültigkeit müssen die Itemparameter in allen möglichen Teilstichproben gleich sein. Diese Bedingung des Rasch-Modells kann mit dem bedingten Likelihood-Ratio-Test von Andersen (1973) überprüft werden. Der Test wurde an jeder der drei vorsortierten Itemgruppen mit Hilfe des statistischen Programms PML (Gustafsson 1981) vorgenommen, und zwar im Hinblick auf Teilgruppenbildungen nach dem Alter, der Bildung und dem Rawscore. Die Tests erbrachten für alle drei Itemgruppen und bei allen Teilgruppenbildungen hochsignifikante Ergebnisse, d. h. die Bedingungen des Rasch-Modells waren in keinem Fall erfüllt.

Um dennoch Rasch-Skalen bilden zu können, mußten im jeweiligen Itemsatz die Items mit der stärksten Modellabweichung entfernt werden. Die Itemselektion erfolgte auf der Grundlage des grafischen Modelltests, wiederum bei allen Teilgruppenkriterien. Insgesamt kam es darauf an, für jedes alltagsästhetische Schema eine möglichst ho-

mogene Skala zu finden, wobei die Differenzen der teilgruppenspezifischen Itemparameter den Betrag von 1.0 nicht überschreiten sollten (vgl. Giegler 1985: 20). Auch nach der Itemselektion war der bedingte Likelihood-Ratio-Test nur bei einer Skala nicht signifikant. Die Signifikanz alleine ist jedoch nicht entscheidend, da die ausschließlich inferenzstatistische Prüfung des Rasch-Modells nicht sinnvoll ist: Unter realen Bedingungen verletzt jede denkbare Itemgruppe die modelltheoretischen Bedingungen, wenn nur der Stichprobenumfang groß genug gewählt wird (Gittler 1986: 388).

Für jede Rasch-Skala wurden zwei verschiedene Meßoperationen vorgenommen, nämlich einerseits durch die Bildung der sogenannten Rawscores (additiver Punktwert) und andererseits durch die Berechnung der modellspezifischen (logarithmierten) Personenparameter. Auf diese beiden Varianten der Rasch-Skalen wird im nächsten Abschnitt ausführlich eingegangen.

6. Ergebnisse des Methodenvergleichs

Im folgenden werden die empirischen Ergebnisse für jedes der sechs Vergleichskriterien vorgelegt und interpretiert.

(1) Dimensionalität des Itemmaterials. Die dreidimensionale Struktur kollektiver Stile trat auch bei der Mokken-Skalierung deutlich zutage, und zwar nach dem zweiten Skalierungsschritt, bei dem die extrem konsistenten Skalenkerne des ersten Schrittes mit weiteren homogenen Items angereichert wurden (vgl. Abschnitt 5). Diese Skalierungsprozedur führte zu drei Itemgruppen, die hohe interne Korrelationen aufwiesen, und die, gemessen an der Itemzahl, wesentlich umfangreicher waren als andere Skalenfragmente. Die Dominanz dieser Itemgruppen wurde als Hinweis auf die dreidimensionale Struktur des Itemmaterials und der dahinterstehenden alltagsästhetischen Schemata bewertet. Damit konnten die faktorenanalytischen Ergebnisse von Schulze repliziert werden, wobei über die inhaltliche Entsprechung der Skalen noch entschieden werden mußte.

(2) Inhaltliche Entsprechung der Skalen. Die empirische Basis zur Durchführung des inhaltlichen Vergleichs der Skalenitems findet sich in den Tabellen 1, 2 und 3. Jedes Item, das bei mindestens einer Skalierungstechnik den festgelegten Selektionskriterien genügte, wurde tabelliert; ein Stern signalisiert die Modellverträglichkeit bei Anwen-

Tabelle 1 Trivialschema.

Item	Klassische Testtheorie r_{it}	Mokken H_i	Rasch d_i
1 Heimatfilme	*	.58	*
2 Volkstheater	*	.64	*
3 Deutsche Schlagermusik	*	.55	*
4 Fernsehshows	*	.53	.
5 Sendungen über Franken	*	.42	.
6 Volkslieder	*	.69	.
7 Goldenes Blatt		*	.05
8 Filmkomödien	*	.40	.
9 Leichte Unterhaltungsmusik	*	.44	.
10 Sonderangebote/Werbung	*	.38	.
11 Natursendungen	*	.38	.
12 Bayerische Volksmusik	*	.67	.
13 Blasmusik	*	.66	.
14 Anzeigenblätter		*	.
15 Heimatromane		*	.
Gütekriterien	Alpha = .85	H = .48	

dung des jeweiligen Verfahrens. Im obersten Rahmen einer Tabelle sind jeweils die Items zu finden, die von allen drei Skalierungsverfahren akzeptiert wurden; in den darunter liegenden Rahmen wird auf die Items Bezug genommen, die bei zwei Verfahren die Modelltests bestanden haben, und in den untersten Rahmen stehen die Items, die sich bei nur einem Skalierungsverfahren als homogen erwiesen. Folgende Itemkennwerte wurden ausgewiesen: der korrigierte Trennschärfeleffizient r_{it} (klassische Testtheorie), der Skalierbarkeitskoef fizient H_i (Mokken-Skalierung) und der Betrag der Itemparameterdifferenzen d_i , bezogen auf die niedrige und die hohe Rawscoregruppe (Rasch Skalierung).

Die Einschätzung der inhaltlichen Entsprechung der Skalen wird bei jedem alltagsästhetischen Schema von zwei Interpretationsschritten abhängig gemacht. Beim ersten Interpretationsschritt kommt es darauf an, die bedeutungsäquivalenten Inhalte der Skalen herauszustellen, und beim zweiten Interpretationsschritt werden die verfahrensspezifischen Inhalte bewertet. Die bedeutungsäquivalenten Inhalte werden durch verfahrensübergreifende Itemgruppen erfaßt. Unter einer solchen Itemgruppe sind Items zu verstehen, die ein gemeinsames ästhetisches Bedeutungsmuster erkennen lassen und von denen zumindest eine Teilmenge bei allen drei Skalierungsverfahren

zu finden ist. Verfahrensübergreifende Itemgruppen können zu einem ästhetischen Gesamtein druck verbunden werden und bilden dann den bedeutungsäquivalenten Kern des alltagsästhetischen Schemas. Beim zweiten Interpretations schritt müssen die verfahrensspezifischen Items – das sind die Items, die keiner übergreifenden Gruppe zugeordnet werden können – skalenweise untersucht werden; wenn sie den ästhetischen Ge samtein druck, der sich beim ersten Interpretationsschritt ergeben hat, bei keiner der drei Skalen verändern, dann ist die inhaltliche Entsprechung der Skalierungsergebnisse gegeben.

Trivialschema (s. Tab. 1): Die Items „Deutsche Schlagermusik“ (3), „Volkslieder“ (6) und „Leichte Unterhaltungsmusik“ (9) bilden die erste verfahrensübergreifende Itemgruppe und weisen das ästhetische Bedeutungsmuster Trivialmusik auf. Diese Musikrichtung wird bei jeder Skalierungstechnik durch mindestens zwei der angesprochenen Items repräsentiert, wobei die deutsche Schlagermusik bei allen Skalierungstechniken zu finden ist. Die zweite verfahrensübergreifende Itemgruppe enthält folgende Items: „Heimatfilme“ (1), „Volkstheater“ (2), „Fernsehshows“ (4) und „Film komödien“ (8). Alle Genres zeichnen sich durch ihre harmonische Grundtendenz aus, sichtbar in heilen Spielfilmwelten, happy ends und netten Moderatoren. Mindestens drei der vier Genres

Tabelle 2 Hochkulturschema.

Item	Klassische Testtheorie	r_{it}	Mokken	H_i	Rasch	d_i
1 Moderne erste Musik (z.B. Bartok)	*	.42	*	.38	*	.24
2 Oper (Fernsehen)	*	.52	*	.48		
3 Klass., modernes Theater (Fernsehen)	*	.48	*	.38		
4 Oper (Musikart)	*	.55	*	.50		
5 Klassische Musik	*	.58	*	.54		
6 Konzerte mit klass. Musik besuchen	*	.40	*	.44		
7 Galerien besuchen	*	.32	*	.35		
8 Germanisches Nationalmuseum			*	.46	*	.32
9 Schauspielhaus			*	.45	*	.28
10 Gedichte lesen	*	.38			*	.01
11 Kulturteil (Zeitung)	*	.40				
12 Klassische Literatur	*	.49				
13 Oper (Besuch)			*	.50		
14 Ein Buch lesen					*	.17
15 Fortbildung					*	.08
16 Politischer Teil (Zeitung)					*	.01
Gütekriterien	Alpha = .79		$H = .46$			

tauchen bei jedem Skalierungsverfahren auf. Trivialmusik und harmonieorientierter Fernsehkonsum erzeugen einen ästhetischen Gesamteindruck, der als Trivialschema bezeichnet werden kann.

Beim zweiten Interpretationsschritt werden die verfahrensspezifischen Items untersucht. Die Itemgruppe Trivialektüre, mit den Items „Goldenes Blatt“ (7) und „Heimatromane“ (15), ist auf die Techniken nach Mokken und Rasch beschränkt; die beiden Items untermauern jedoch den ästhetischen Gesamteindruck des Trivialschemas. Die noch verbleibenden Items „Fernsehsendungen über Franken“ (5), bevorzugte Zeitungsektüre „Sonderangebote/Werbung“ (10), „Natursendungen“ (11) und Interesse für „Anzeigenblätter“ (14) bilden in den einzelnen Skalen keine neuen ästhetischen Muster, welche den bedeutungsäquivalenten Gesamteindruck verändern können.

Hochkulturschema (s. Tab. 2): Die Itemtabelle weist schon nach kurzer Inspektion auf einen ästhetischen Gesamteindruck hin, der üblicherweise

als Hochkultur bezeichnet wird. Das augenfälligste Bedeutungsmuster ist ernste Musik, die gewissermaßen das Markenzeichen gehobener Kultur ist: „Moderne ernste Musik“ (1), „Oper“ (2,4) und „klassische Musik“ (5). Diese Itemgruppe wurde von jedem Verfahren erfaßt, auch wenn in der Rasch-Skala nur eines der genannten Items zu finden ist. Außerhäusige Betätigungen, die der gehobenen Freizeitkultur zuzurechnen sind, bilden die zweite verfahrensübergreifende Itemgruppe. Dazu gehört der Besuch folgender Nürnberger Einrichtungen und Veranstaltungen: Die „Oper“ (13), das „Schauspielhaus“ (9), „Galerien“ (7), das „Germanische Nationalmuseum“ (8) sowie „Konzerte mit klassischer Musik“ (6). Die Mokken-Skala enthält alle fünf Items, die Rasch- und klassisch konstruierte Skala enthalten jeweils zwei dieser Items.

In den drei Skalen finden sich auch verfahrensspezifische Inhalte, die jedoch den ästhetischen Gesamteindruck der drei Skalen nicht verändern. Die gehobene Lektüre ist ein solcher Inhalt, der durch die folgende Itemgruppe vertreten wird: „Gedichte lesen“ (10), Präferenz für „klassische Literatur“

Tabelle 3 Spannungsschema.

Item	Klassische Testtheorie	r_{it}	Mokken	H_i	Rasch	d_i
1 Raggae-Musik	*	.50	*	.43	*	1.00
2 Rockfestival (Veranstaltung)	*	.46	*	.43	*	.63
3 Kino	*	.39	*	.42	*	.27
4 Popmusik (Fernsehen)	*	.61	*	.55		
5 Rockmusik	*	.61	*	.55		
6 Oldies	*	.42	*	.44		
7 Konzerte (Jazz, Liedermacher ...)			*	.45	*	.43
8 In ein Nachtlokal gehen			*	.41	*	.16
9 Flippern/kickern			*	.38	*	.29
10 Soul-Musik	*	.52				
11 Folk-Musik	*	.45				
12 Blues	*	.45				
13 Sich mit anderen in der Stadt treffen	*	.32				
14 Kneipe, Lokal	*	.40				
15 Trempelmarkt			*	.43		
16 Altstadtfest			*	.40		
17 Diskothek			*	.38		
Gütekriterien	Alpha = .83		$H = .47$			

(12) sowie „Kulturteil“ (11) und „politischer Teil“ (16) einer Zeitung. Keines dieser Items wurde von der Mokken-Skalierung erfaßt, was aber insgesamt ohne inhaltliche Folgen bleibt, da die gehobene Lektüre lediglich eine äquivalente Erweiterung des hochkulturellen Bedeutungsmusters ist.

Spannungsschema (s. Tab. 3): In jeder der drei Skalen finden sich Variationen der Pop- und Rockmusik (Items 1,4,5,6,10,11,12); diese Items bilden die erste verfahrensübergreifende Itemgruppe. Die zweite kommt in den außerhäusigen Freizeittätigkeiten zum Vorschein: ins „Kino“ (3), in die „Kneipe“ (14), in die „Diskothek“ (17) und in ein „Nachtlokal“ (8) gehen; „sich mit anderen in der Stadt treffen“ (13), zum „Flippern/Kickern“ (9) gehen und andere Items (2,7,15,16). Der gemeinsame Nenner dieser Freizeittätigkeiten liegt in der Actionorientierung und der Suche nach neuartigen Erlebnissen. Die musikalischen Inhalte und die Freizeittätigkeiten erzeugen einen ästhetischen Gesamteindruck, der als Spannungsschema bezeichnet werden kann; bei diesem Schema steht die körperliche und psychische Spannungssuche sowie deren Entladung im Mittelpunkt der Erleb-

nisorientierung. Bei keiner der drei Skalen findet sich ein verfahrensspezifisches Item.

Für das zweite Vergleichskriterium kann folgendes Gesamtergebnis formuliert werden: Jedes der drei alltagsästhetischen Schemata läßt sich inhaltlich wiedererkennen, und zwar unabhängig davon, ob das Mokken-, das Rasch- oder das klassische Verfahren zur Konstruktion der Skalen eingesetzt wurde. Die verfahrensspezifischen Inhalte verändern den bedeutungsäquivalenten Kern des jeweiligen alltagsästhetischen Schemas nicht; dies weist auf die inhaltliche Entsprechung der Skalen bei allen drei Konstruktionsverfahren hin.

(3) Korrelation der Meßwerte. In Tabelle 4 werden für jedes alltagsästhetische Schema die sechs Produktmomentkorrelationen der vier Meßtechniken ausgewiesen. Folgende vier Ergebnisse sind bemerkenswert: Erstens weichen die Korrelationen teilweise deutlich vom Idealfall des vollständigen Zusammenhangs ab, so daß keine vollkommene Äquivalenz der Meßtechniken gegeben ist; dabei läßt sich zweitens erkennen, daß sich die Meßwerte der Mokken- und der „klassischen“ Skalen in allen drei Fällen sehr ähnlich sind und deren Korre-

Tabelle 4 Intraskalenkorrelationen (Produktmomentkorrelationen)

	Trivialschema				Hochkulturschema				Spannungsschema			
	Klassische Testtheorie	Mokken	Rasch (Raw-score)	Rasch	Klassische Testtheorie	Mokken	Rasch (Raw-score)	Rasch	Klassische Testtheorie	Mokken	Rasch (Raw-score)	Rasch
Klassische Testtheorie												
Mokken	.92				.83				.89			
Rasch (Rawscore)	.91	.87			.58	.68			.76	.81		
Rasch	.85	.80	.99		.53	.62	.99		.54	.70	.99	

lationen mit den zwei Rasch-Varianten beim Hochkultur- und Spannungsschema vergleichsweise niedrig ausfallen. Drittens haben die Rawscores des Rasch-Verfahrens grundsätzlich eine größere Ähnlichkeit mit den Mokken- und „klassischen“ Skalen als die logarithmierten Personenparameter. Dies gilt, obwohl viertens die beiden Rasch-Varianten eine nahezu exakte lineare Abhängigkeit aufweisen. Auf die beiden letzten Teilergebnisse wird im folgenden näher eingegangen.

Die Summenscores einer Rasch-Skala sind nahezu lineare Transformationen der logarithmierten Personenparameter; darauf hat bereits Oesterreich (1978: 344) hingewiesen. Somit erhebt sich die Frage, warum die Korrelationen der Personenparameter mit den Mokken- und klassisch konstruierten Skalen teilweise deutlich unter den Korrelationen der Rawcores mit diesen Skalen liegt. Die unterschiedliche Behandlung der Extremwerte bietet einen Ansatzpunkt zur Beantwortung dieser Frage. Das Rasch-Modell sieht für die beiden Extremwerte des Rawcores keine Schätzung der Personenparameter vor, folglich müssen die beiden Personengruppen mit den extremen Rawscorewerten von allen nachfolgenden statistischen Auswertungen ausgeschlossen werden. Die niedrigeren Korrelationen können somit auf den entstandenen Streuungsverlust bei der Verwendung der Personenparameter zurückgeführt werden. Besonders deutlich ist dies beim Spannungsschema zu sehen, bei dem die untere Extremwertgruppe vergleichsweise stark besetzt ist: Die Korrelation mit der Skala nach der klassischen Testkonstruktion sinkt von .76 (Rawcore) auf .54 (logarithmierte Personenparameter) ab.

Wenn es um die Zusammenhangsanalyse geht, sind die Rawcores der Rasch-Skalen offensichtlich besser zur Messung der alltagsästhetischen Schemata geeignet als die Personenparameter, da sie alle Stufen der Skalendifferenzierung enthalten

und zudem nahezu lineare Transformationen der Personenparameter sind.

Beim dritten Vergleichskriterium bietet sich keine eindeutige Gesamtinterpretation der Ergebnisse an und es bleibt abzuwarten, ob die erreichte Äquivalenz der verfahrensspezifischen Maßwerte konsistente inhaltlich-theoretische Aussagen zuläßt. Diese Frage wird im Zusammenhang mit dem fünften Vergleichskriterium zu klären sein, bei dem die Korrelationen mit externen Merkmalen verglichen werden.

(4) Innere Konsistenz. Für dieses Vergleichskriterium werden folgende Maße herangezogen: Cronbachs Alpha im Rahmen der klassischen Testtheorie sowie Loevingers H für das Mokken-Verfahren; ein vergleichbares Maß ist für die Rasch-Skalierung nicht vorhanden, so daß die Frage nach der Itemhomogenität bei diesem Verfahren deskriptiv nicht befriedigend beantwortet werden kann. Die jeweiligen Skalenwerte von Alpha und H finden sich in den Fußzeilen der Tabellen 1, 2 und 3. Nach den üblichen Kriterien untersucht, haben alle Mokken- und „klassischen“ Skalen eine ausreichende innere Konsistenz, d. h. die Existenz der alltagsästhetischen Schemata konnte methodenunabhängig nachgewiesen werden.

(5) Externe Korrelationen. In Tabelle 5 sind die Korrelationsmuster zu finden, die sich jeweils aus den Korrelationen der verfahrensspezifischen Skalen mit den theorierelevanten Merkmalen Alter und Bildung sowie den Skalen politische Unterordnung und Vertrauen ergeben.

Die drei Korrelationsmuster der Skalen nach der klassischen Testkonstruktion sowie der Mokken- und Rasch-Skalen (Rawcore) sind weitgehend konsistent: Bei allen Verfahren zeigt sich die Altersabhängigkeit des Trivial- und Spannungsschemas sowie die Bildungsabhängigkeit des Trivial- und Hochkulturschemas. Die genannten Zusam-

Tabelle 5 Korrelationen mit theorierelevanten Merkmalen (Produktmomentkorrelationen)

	Klassische Testtheorie			Mokken			Rasch (Rawscore)			Rasch		
	Trivial-schemata	Hoch-kultur-schemata	Span-nungs-schemata	Trivial-schemata	Hoch-kultur-schemata	Span-nungs-schemata	Trivial-schemata	Hoch-kultur-schemata	Span-nungs-schemata	Trivial-schemata	Hoch-kultur-schemata	Span-nungs-schemata
Alter	.42	.10	-.55	.38	.05	-.55	.31	-.11	-.42	.20	-.08	-.19
Bildung	-.41	.30	.19	-.42	.37	.14	-.38	.43	.12	-.23	.39	.06
Politische Unterordnung	-.53	.13	.35	-.54	.17	.33	-.47	.31	.28	-.35	.25	.09
Vertrauen	-.39	.13	.24	-.40	.15	.19	-.38	.29	.19	-.26	.23	.12

Tabelle 6 Interskalenkorrelationen (Produktmomentkorrelationen)

	Klassische Testtheorie			Mokken			Rasch (Rawscore)			Rasch		
	Trivial-schemata	Hoch-kultur-schemata	Span-nungs-schemata	Trivial-schemata	Hoch-kultur-schemata	Span-nungs-schemata	Trivial-schemata	Hoch-kultur-schemata	Span-nungs-schemata	Trivial-schemata	Hoch-kultur-schemata	Span-nungs-schemata
Trivialschema												
Hochkul-turschema	-.09			-.22			-.32			-.19		
Span-nungs-schemata	-.32	.02		-.31	.02		-.22	.14		-.24	.06	

menhänge sind konstitutiv für das erwähnte Modell sozialer Milieus (vgl. Abschnitt 1). Verfahrens-unabhängig sind auch die hohen Korrelationen zwischen dem Trivialschema und den externen Merkmalen Vertrauen und politische Unterordnung.

Das Korrelationsmuster, das sich für die Rasch-Skalen mit den Personenparametern ergibt, weicht deutlich von den anderen Korrelationsmustern ab, alle Korrelationen fallen niedriger aus. Dieser Effekt ist auf den Differenzierungsverlust bei Verwendung der Personenparameter zurückzuführen. An dieser Stelle zeigt sich, daß die (logarithmierten) Personenparameter nicht nur eine unnötige Komplizierung des einfachen additiven Punktwertes (Rawscore) sind, sondern auch zu methodisch bedingten Artefakten führen können.

(6) Dimensionales Beziehungsmuster. Die Korrelationen zwischen den drei alltagsästhetischen Schemata werden für jedes Skalierungsverfahren in Tabelle 6 ausgewiesen. Die Beziehung zwischen dem Spannungs- und Hochkul-turschema wird konsistent wiedergegeben, die diesbezügliche Korrelation liegt bei jedem Verfahren in der Nähe statistischer Zusammenhangslosigkeit. Im Verhältnis von Trivial- und Hochkul-turschema bzw. Trivial-

und Spannungsschema kommt es zu verfahrensspezifischen Abweichungen. Die Unterschiede der statistischen Zusammenhänge könnten zu unterschiedlichen Interpretationen führen, auch wenn bei allen Verfahren ein dreidimensionaler Raum ohne klare Polaritäten (schwache Zusammenhänge) zu erkennen ist.

7. Fazit und Ausblick

Eingangs wurde die Frage aufgeworfen, ob der Einsatz verschiedener Skalierungsverfahren zu unterschiedlichen inhaltlich-theoretischen Ergebnissen führt; diese Frage wurde vor dem theoretischen Hintergrund alltagsästhetischer Schemata gestellt und anhand von sechs Vergleichskriterien für drei Skalierungstechniken empirisch geklärt. Auf der Grundlage der vorgelegten Ergebnisse läßt sich folgendes Fazit ziehen: Beim gegebenen soziologischen Theoriekontext hat der alternative Einsatz der untersuchten Skalierungsverfahren keine relevante substanzwissenschaftliche Bedeutung.

Dieses Fazit stärkt zunächst die empirische Begründung der drei alltagsästhetischen Schemata, die in der Theorie der Erlebnisgesellschaft (Schul-

ze 1992) und deren historische Entfaltung (Müller-Schneider im Druck) an prominenter Stelle stehen. Die im Methodenvergleich gewonnenen Einsichten können nicht ohne weiteres verallgemeinert werden, da eine Fallstudie, wie sie hier vorgelegt wurde, nur begrenzte Aussagekraft besitzt. Man kann jedoch davon ausgehen, daß die Mokken-Skalierung und die klassische Testkonstruktion bei jeder deutlich strukturierten itemkorrelativen Situation zu ähnlichen statistischen Ergebnissen führen würden, da die Interitemkorrelation bei beiden Verfahren im Mittelpunkt der Skalierungsprozedur steht.

Beim Rasch-Verfahren ist die Frage der Verallgemeinerbarkeit sehr viel schwieriger zu beantworten, da dieses Modell nicht den Homogenitätsbegriff der internen Konsistenz, sondern die lokale stochastische Unabhängigkeit heranzieht (Abschnitt 1). Die vorgelegten Ergebnisse sind wohl nur für den Fall verallgemeinerbar, daß das Itemmaterial mit der Faktorenanalyse oder dem Mokken-Verfahren vorstrukturiert wird, da dann aus einem Satz korrelierender Items die jeweils modellverträglichen ausgewählt werden. Für den Fall, daß das Rasch-Verfahren ohne Vorstrukturierung des Itemmaterials eingesetzt wird, muß die Frage der Verallgemeinerbarkeit offen bleiben.

Im Zusammenhang mit dem inhaltlichen Fazit und der Verallgemeinerung der Ergebnisse werden abschließend einige weitergehende method(ologische Überlegungen angestellt:

1. Die methodische Fachdiskussion über Skalierungsverfahren sollte das typisch soziologische Erkenntnisinteresse kollektiver Strukturen in großen Populationen in Rechnung stellen. Die übliche Kritik an der klassischen Testtheorie orientiert sich an typisch psychologischen Erkenntnisinteressen, die sich auf Individualdiagnosen oder individuelle Veränderungsmessung beziehen (vgl. Schulze 1990: 5). Wenn soziologisches Erkenntnisinteresse zugrunde gelegt wird, kann die Brauchbarkeit der klassischen Skalenkonstruktion neu beurteilt werden.

2. Die klassische Skalenkonstruktion ist nämlich ein brauchbares Verfahren zur Auffindung und Messung kollektiver DispositionsvARIABLEN. Mit Brauchbarkeit sind zwei Sachverhalte gemeint: Erstens kann anhand klassischer Gütekriterien (z. B. Cronbachs Alpha) über die Existenz einer kollektiven DispositionsvARIABLEN in einer bestimmten Population entschieden werden und zweitens sind die gemessenen Zusammenhänge zwischen Skalen nach der klassischen Testtheorie und anderen

Merkmalen valide Indikatoren realer Zusammenhänge. Damit werden die modelltheoretischen Mängel der klassischen Testtheorie nicht in Abrede gestellt, aber deren Relevanz für die soziologische Theoriebildung.

3. Die häufig postulierte Überlegenheit der Rasch-Skalierung wird immer fragwürdiger. Die begrenzte Leistungsfähigkeit der Modellkontrollen und die Probleme bei der Itemselektion sind bereits bekannt (Gittler 1986; Gustaffson 1980; Stelzl 1979 u. a.); hinzu kommt die erläuterte Problematik bei der Verwendung der Personenparameter, die das metrische Niveau einer Rasch-Skala garantieren sollen und für die „spezifisch-objektive“ Messung einer DispositionsvARIABLEN unerlässlich sind. Ganz unabhängig von methodologischen Einwänden gegen die Möglichkeit „spezifisch-objektiver“ Messungen in den Sozialwissenschaften (Schulze 1989) droht das Rasch-Modell an der Umsetzung in die Forschungspraxis zu scheitern.

4. Die Mokken-Skalierung fand in den Sozialwissenschaften bislang zu wenig Beachtung. Dieses Verfahren ist insbesondere zur Auffindung mehrdimensionaler Strukturen geeignet und bietet der Faktorenanalyse gegenüber einen Vorteil. Die Faktorenanalyse setzt eine orthogonale oder eine willkürlich gewählte oblique Struktur der Faktoren voraus; das Mokken-Verfahren hingegen findet einzelne Dimensionen unabhängig von einem postulierten Beziehungsmuster auf, d. h. Orthogonalität oder eine bestimmte oblique Struktur der Dimensionen werden nicht vorausgesetzt.

5. Offensichtlich gilt das Diktum „unterschiedliche Methoden erzeugen unterschiedliche Wirklichkeiten“ (Bonß/Hartmann 1985: 20) nicht grundsätzlich, da alle Skalierungsverfahren die gleiche Wirklichkeitsauffassung über Struktur und Inhalt kollektiver Stile sowie deren Zusammenhang mit theorierelevanten Kriterien hervorbrachten. Verschiedene Verfahren können durchaus zu gleichen Ergebnissen führen.

Literatur

- Andersen, E.B., 1973: A Goodness Of Fit Test For The Rasch Model. *Psychometrika* 38: 123–140.
 Atteslander, P., 1991: Methoden der empirischen Sozialforschung. Berlin/New York: de Gruyter.
 Bonß, W./Hartmann, H., 1985: Konstruierte Gesellschaft, rationale Deutung. S. 9–46 in: Dies. (Hrsg.), Entzauberte Wissenschaft. Zur Relativität und Gelassenheit soziologischer Forschung. Soziale Welt: Sonderband 3. Göttingen: Schwartz.

- Debets, P./Brouwer, E., 1988: Program MSP version 1.40. Amsterdam: Disc.
- Fischer, G.H., 1974: Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen. Bern/Stuttgart/Wien: Huber.
- Giegler, H., 1985: Rasch-Skalen zur Messung von „Arbeits- und Berufszufriedenheit“ und „Arbeits- und Berufsbelastung“ auf Seiten der Betroffenen. Zeitschrift für Sozialpsychologie 16: 13–28.
- Gillespie, M./Tenvergert, E.M./Kingma, J., 1987: Using Mokken scale analysis to develop unidimensional scales. Quality and Quantity 21: 393–408.
- Gittler, G., 1986: Inhaltliche Aspekte bei der Itemselektion nach dem Modell von Rasch. Zeitschrift für experimentelle und angewandte Psychologie 33: 386–412.
- Gustafsson, J.E., 1980: Testing and obtaining fit of data to the Rasch model. British Journal of Mathematical and Statistical Psychology 33: 205–233.
- Gustafsson, J.E., 1981: A computer program for conditional estimating and testing in the Rasch model for dichotomous items. Version 3.1. University of Göteborg.
- Heidenreich, K., 1984: Grundbegriffe der Meß- und Testtheorie. S. 352–384 in: E. Roth (Hrsg.), Sozialwissenschaftliche Methoden. München/Wien: Oldenbourg.
- Henning, J.H./Six, B., 1977: Konstruktion einer Machiavellismus-Skala. Zeitschrift für Sozialpsychologie 8: 185–198.
- Hilke, R., 1980: Grundlagen normorientierter und kriteriumsorientierter Tests. Bern/Stuttgart/Wien: Huber.
- Jansen, P.G.W., 1982: Homogenitätsmessung mit Hilfe des Koeffizienten H von Loevinger: Eine kritische Diskussion. Psychologische Beiträge 24: 96–105.
- Kubinger, K.D., 1988: Aktueller Stand und kritische Würdigung der probabilistischen Testtheorie. S. 19–83 in: Ders. (Hrsg.), Moderne Testtheorie. Weinheim/Basel: Beltz.
- Lippert, E./Schneider, P./Wakenhut, R., 1978: Die Verwendung der Skalierungsverfahren von Mokken & Rasch zur Überprüfung und Revision von Einstellungsskalen. Diagnostica 24: 252–274.
- Müller-Schneider, Th., im Druck: Schichten und Erlebnismilieus. Der Wandel der Milieustruktur in der Bundesrepublik Deutschland. Wiesbaden: DUV.
- Niemöller, K./van Schuur, W., 1983: Stochastic Models For Unidimensional Scaling: Mokken and Rasch. S. 120–170 in: D. McKay/N. Schofield/P. Whiteley (Hrsg.), Data Analysis And The Social Sciences, London: Pinter.
- Oesterreich, R., 1978: Welche der sich aus der Rasch-Skalierung ergebenden Personenkennwerte sind für statistische Auswertungen geeignet? Diagnostica 24: 341–349.
- Schnell, R./Hill, P.B./Esser, E., 1988: Methoden der empirischen Sozialforschung. München/Wien: Oldenbourg.
- Schulze, G., 1989: Skalierungsmodelle, Forschungspraxis und theoretische Konzepte in der empirischen Sozialforschung. Bamberg, Manuscript.
- Schulze, G., 1990: Alltagsästhetik, Milieustruktur und Erlebnismarkt. Projektbericht für die DFG. Anhang II. Meßtheoretische Überlegungen und Skalen. Universität Bamberg.
- Schulze, G., 1992: Die Erlebnisgesellschaft. Kultursociologie der Gegenwart. Frankfurt a. M./New York: Campus.
- Stelzl, I., 1979: Ist der Modelltest des Rasch-Modells geeignet, Homogenitätshypothesen zu prüfen? Ein Bericht über Simulationsstudien mit inhomogenen Daten. Zeitschrift für experimentelle und angewandte Psychologie 26: 652–672.
- Stokman, F./van Schuur, W., 1980: Basic Scaling. Quality and Quantity 14: 5–30.
- Wottawa, H., 1977: Psychologische Methodenlehre. München: Juventa.